The EMBO Journal Vol.1 No.11 pp.1473–1477, 1982

A synthetic heat-shock promoter element confers heat-inducibility on the herpes simplex virus thymidine kinase gene

Hugh R.B.Pelham* and Mariann Bienz

M.R.C. Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Communicated by A.Travers
Received 14 October 1982

Previous deletion analysis of the *Drosophila hsp70* heat-shock promoter has identified a sequence upstream of the TATA box that is required for heat induction. This region contains homology to other heat-shock promoters, and it was proposed that the common sequence is an important element in the regulation of the heat-shock genes. We have constructed sequences similar to the consensus CT-GAA--TTC-AG from synthetic oligonucleotides and placed them upstream of the TATA box of the herpes virus thymidine kinase gene, in place of the normal upstream promoter element. The resultant genes are heat-inducible both in monkey COS cells and in *Xenopus* oocytes. We conclude that the transcriptional heat-shock response is mediated by some factor that interacts with this sequence.

Key words: COS cells/eucaryotic promoter/heat shock/synthetic oligonucleotides/*Xenopus* oocytes

Introduction

Cells from a wide variety of organisms respond rapidly to thermal stress and certain other environmental stimuli by synthesizing a small number of 'heat-shock' proteins, whose function is thought to be protective (Schlesinger *et al.*, 1982). In most cases this response involves the induction of transcription of the corresponding heat-shock genes. The mechanism of this response must have been highly conserved during evolution, since the cloned *hsp70* heat-shock gene from *Drosophila* is transcribed in a heat-inducible manner when introduced into monkey (Pelham, 1982; Mirault *et al.*, 1982), rat (Burke and Ish-Horowicz, 1982), mouse (Corces *et al.*, 1981), and *Xenopus* cells (Voellmy and Runnger, 1982; Pelham and Bienz, 1982). As a step towards understanding this mechanism, we are attempting to define the DNA sequences involved in the response.

A previous paper described the construction of deletion mutants of the *Drosophila hsp70* heat-shock gene, and their expression on replicating plasmids introduced into monkey COS cells (Pelham, 1982). These studies showed that heat-inducibility is a property of the promoter itself, and identified an essential region upstream of the TATA box that bears sequence homology to other heat-shock promoters. A consensus sequence CTgGAAtnTTCtAGa was derived, and it seemed possible that regulation requires only the presence of this sequence, which in heat-shocked cells apparently constitutes the upstream element of the promoter (Pelham, 1982). We have now tested this hypothesis directly by constructing such a sequence from synthetic oligonucleotides, and show that it alone is sufficient to confer heat-inducibility on the herpes simplex virus thymidine kinase *(tk)* gene.

*To whom reprint requests should be sent.

Results

Construction of plasmids

McKnight *et al.* (1981) described mutants of the *tk* gene that are deleted from the 5′ side, the deletion endpoints being marked by synthetic *Bam*HI linkers. One such mutant (pTKΔ5′-32) was used in this work: it is deleted to within 4 bp of the TATA box, and thus lacks the normal *tk* upstream promoter element which is required for efficient expression of the gene. The *tk* gene was recloned in a pBR322-derived vector which is capable of replicating in COS cells (Gluzman, 1981). These cells produce SV40 T-antigen constitutively; the vector contains an SV40 replication origin, and also lacks procaryotic sequences that are inhibitory to replication in COS cells (Lusky and Botchan, 1981). The *Bam*HI site at the deletion endpoint was joined *via* an *Xba*I linker to the *Sal*I site of the vector; this construction (pTKS0) results in tandem *Sal*I, *Xba*I, and *Bam*HI sites (Figures 1 and 2).

The principal feature of the proposed heat-shock promoter element is the symmetric sequence CT-GAA--TTC-AG. This was generated as shown in Figure 2 by insertion of the adapter CTAGAAGCTT into the *Xba*I site of pTKS0. Insertion of two copies of the adapter generates a larger inverted repeat centred in the same relative position as the inverted repeats present in most *Drosophila* heat-shock genes (see Discussion). Individual plasmids containing one (pTKS1) or two copies of the sequence (pTKS2) were identified by sizing the *Bam*HI-*Sal*I fragments, after end-labelling, on sequencing gels. A related sequence (matching eight of the 10 bases in the proposed element) was produced by inserting two copies of the adapter CTAGAGATCT (pTKS12). As a control, one or two *Hind*III linkers (GCAAGCTTGC) were inserted into the *Xba*I site by blunt-end ligation (pTKS21, pTKS22); this pro-

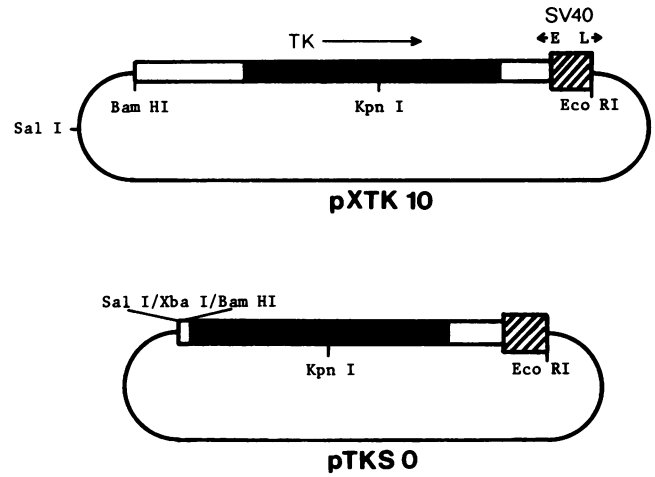

Fig. 1. Basic structures of the plasmids. Thin lines represent procaryotic vector sequences (from pML), and have been lengthened slightly for clarity. The SV40 fragment (shaded box) extends from the *Hind*III site on the early side of the origin of replication to the *Eco*RII site on the late side; it contains only part of one of the 72-bp repeats and lacks enhancer activity (Banerji *et al.*, 1981). The remaining sequences are from herpes simplex virus, and the transcribed portion of the *tk* gene is shown as a black box, transcription being from left to right.

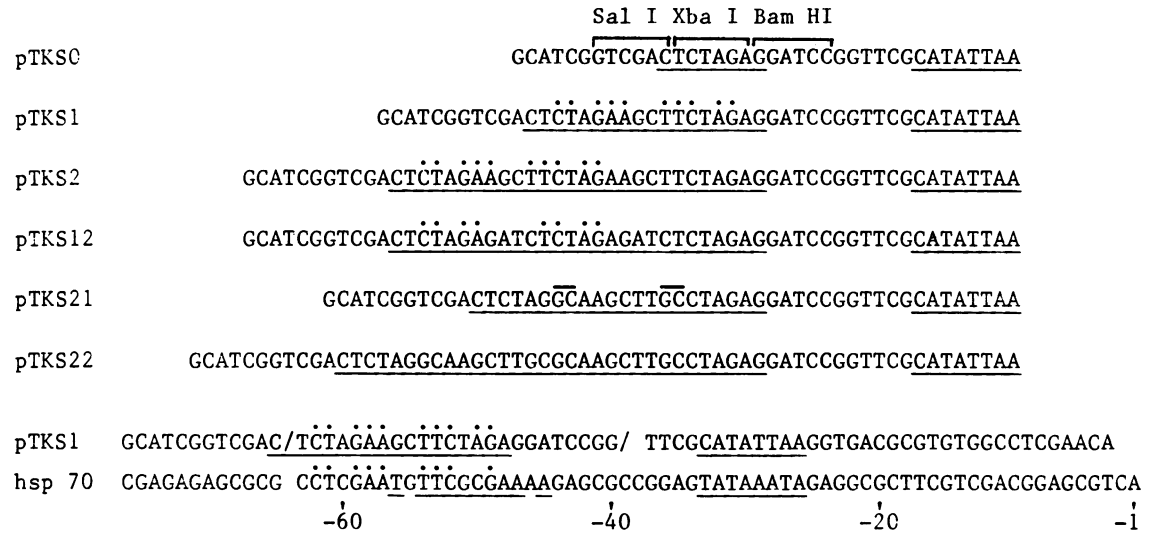

Fig. 2. Sequences of the synthetic promoter elements. The TATA box and inverted repeats upstream from it are underlined. Dots over some bases indicate their identity with the consensus sequence. The overlining in the sequence of pTKS21 indicate the four bases that are inserted relative to pTKS1. Also shown are the sequences up to the transcription start site of pTKS1 and the *Drosophila hsp70* gene; a gap has been introduced into the pTKS1 sequence to allow alignment of the TATA boxes and homologous sequences. The /s in this pTKS1 sequence show the junctions between sequences derived from pBR322 (left), linkers and adapters (centre), and the *tk* gene (right).

duces large inverted repeats with little homology to the proposed functional sequence. Figure 2 shows the various constructions, and also compares the sequence of one of these, pTKS1, with the *Drosophila hsp70* heat-shock gene. The only significant homology between these two genes is in the region of the proposed functional sequence itself, where 10 out of 15 residues are identical.

Assay of plasmids

The *tk* genes were assayed in COS cells as previously described (Pelham, 1982). DNA was introduced into the cells in the presence of DEAE-dextran, and after 30 h duplicate flasks of cells were heat-shocked at 42–43°C or maintained at 37°C for 4 h. RNA was extracted from the cells, and *tk* transcripts quantitated by S1 mapping of their 5′ ends, using as probe an end-labelled DNA fragment which overlaps the normal cap site of the wild-type *tk* gene. Typical results are shown in Figure 3a, and the data is summarized quantitatively in Figure 4. For comparison, an intact *tk* gene with ~700 bases of 5′-flanking sequence (pXTK10) was also assayed in COS cells. As previously described (Pelham, 1982), this yields both transcripts with correct 5′ ends, and ones that initiate at a number of discrete upstream sites (Figure 3a). The level of these transcripts was slightly lower after heat-shock.

pTKS0, which lacks the normal *tk* upstream element, produced ~15-fold fewer correct transcripts, and again these were less abundant after heat-shock (Figure 3a). There are also some relatively abundant transcripts that originate upstream of the gene and thus map to the point of sequence divergence between probe and template; other bands visible in Figure 3 appear to be artefacts of the S1 digestion procedure, but their origin is unclear. Plasmids pTKS1, pTKS2, and pTKS12, which contain sequences similar to the heat-shock promoter element, behaved quite differently. These were at least as inactive as pTKS0 at low temperature, but at high temperature produced 15- to 20-fold more transcripts, a level comparable to that obtained with the normal *tk* gene. In contrast, plasmids pTKS21 and pTKS22 were not heat inducible (Figure 4), even though pTKS21 differs from pTKS1 only by the insertion of four extra base pairs (Figure 2).

To be sure that the inducibility of the synthetic promoters was not in some way depedent on the unusual conditions of the COS cell assay system, we also assayed the plasmids by microinjecting them into the nuclei of *Xenopus* oocytes. After 2 h preincubation, these were heat-shocked at 34°C or maintained at 25°C for 2 h, and the RNA extracted and assayed as before. The results were quantitatively somewhat more variable than with COS cells, but Figures 3b and 4 show that pTKS1 and pTKS2 were again specifically heat-inducible, whereas the control plasmids pTKS0, pTKS21, and pTKS22 were not.

Discussion

Heat inducibility is a property of the upstream promoter element

The *tk* gene is not itself heat inducible, and presumably has never experienced evolutionary pressure to become so. It can be made heat inducible, however, simply by replacing the normal upstream promoter element with a short synthetic sequence that has homology to a heat-shock gene. This clearly demonstrates that the homologous sequence is sufficient to make a promoter regulatable, and suggests that some factor which mediates the stress responses interacts directly with it.

It remains possible that not all promoters are compatible with the heat-shock element. The *tk* gene does have some general properties in common with the *hsp70* heat-shock gene, in that both are expressed efficiently in COS cells and *Xenopus* oocytes, and do not require remote 'enhancer' sequences for activity (Pelham, 1982). Moreover, attempts to make the adenovirus major later promoter heat inducible in the same way have so far been unsuccessful (H.P., unpublished observations).

Identity of the functional sequences

A consensus sequence for the putative heat-shock element can be derived by comparison of the different *Drosophila* heat-shock gene sequences (Pelham, 1982). The statistical significance of this consensus is obviously questionable because of the small number of genes compared, and because

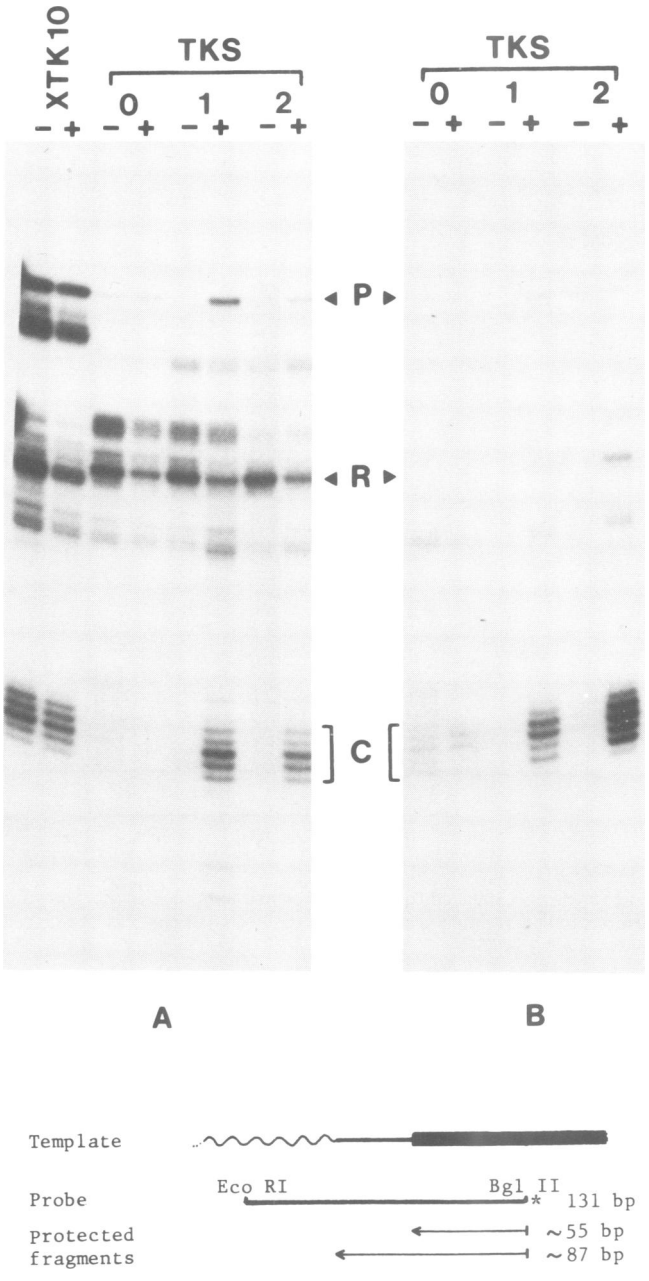

Fig. 3. S1 mapping the 5′ ends of *tk* transcripts. The line diagram outlines the mapping procedure. The transcribed portion of the template is shown as a thickened line, and the wavy line represents sequences in the pTKS series that are not complementary to the probe, which is derived from the wild-type *tk* gene. **Parts A** and **B** show autoradiograms of gels on which the S1-protected fragments were separated. The positions of intact probe (P) and the fragments corresponding to readthrough (R) and correctly-initiated transcripts (C) are indicated. **A**: RNA samples were from COS cells transfected with pXTK10 or with members of the pTKS series, and heat-shocked (+) or not (−). **B**: RNA samples were from *Xenopus* oocytes injected in batches of 15 with plasmids from the pTKS series and heat-shocked (+) or not (−).

there may be some doubt as to the identity of the functional sequence in some of the genes. Nevertheless, a striking feature of the consensus is the symmetric sequence CT-GAA--TTC-AG. The results presented here provide strong evidence for the importance of this sequence. Figure 5 shows the best fit in the seven *Drosophila* heat-shock genes and in the various mutants of the *hsp70* and *tk* genes that have been tested. All of the mutants that retain activity, and most of the wild-type genes fit the consensus at eight or more of the 10 positions. Conversely, mutants that only match it at four or five positions, whichever these are, are not heat-inducible. Mismatches in one or two positions can evidently be tolerated, and no single base has yet been shown to be essential, although pHT6 and 7, which lack only the two adjacent A residues, have very low activity (Pelham, 1982). Certain bases outside these 10 also seem to be preferred *in vivo* – for example, the central two bases tend to be T residues – but these are evidently not essential in our assays.

The data is consistent with the idea that the consensus sequence is a binding site for a protein. Many such binding sites that have been identified in procaryotes have similar dyad axes of symmetry, presumably because they interact with dimeric or tetrameric proteins. If this is the function of the consensus sequence, then proteins with very similar binding properties must be present in *Drosophila, Xenopus,* and monkey cells.

Whatever the role for the regulatory site, its exact distance from the TATA box is not critical for function (Figure 5). In particular, distances of 13 bp (pTKS1) to 19 bp (pHT5) are tolerated in our assays. This implies that the angular separation between these two promoter elements on the DNA helix is also not critical.

Two of the *Drosophila* heat-shock genes appear somewhat anomalous. The *hsp23* gene has the consensus sequence, but it is much further away from the TATA box (99 bp) than in the other genes. Interestingly, the intervening DNA contains stretches of alternating purines and pyrimidines which could in principle form left-handed Z DNA under physiological conditions (Singleton *et al.*, 1982). The *hsp27* gene, on the other hand, shows only weak homology to the consensus sequence (Figure 5), and it is far from clear what features of this gene allow it to function efficiently *in vivo*.

Inverted repeats

Another common feature of the heat-shock promoters is a larger inverted repeat upstream of the TATA box (Holmgren *et al.*, 1981). For five of the seven *Drosophila* promoters this is centered at the 5′ side of the penultimate A residue of the consensus sequence, although the sequence of the inverted repeat itself is not conserved (Pelham, 1982). These are the genes for *hsp70* (10 out of 12 bases form a dyad, Figure 2), *hsp83* (20 out of 24), *hsp22* (12 out of 16), *hsp26* (12 out of 14), and *hsp27* (12 out of 16 or 18 out of 28). The inactivity of pTKS21 and pTKS22 shows, however, that the mere presence of an inverted repeat upstream from the TATA box does not substitute for an upstream promoter element and is not sufficient for heat inducibility. Moreover, in the presence of the consensus sequence there seems to be no correlation between the size or position of the inverted repeat and the efficiency of heat induction. Thus pTKS2 has a longer repeat than pTKS1, with a dyad axis in the appropriate position, yet it is induced no more efficiently than pTKS1. Previous studies have also shown that disruption of the repeat in the *hsp70* gene by replacement of residues −45 to −50 with a 10-base (symmetric) *Hind*III linker does not abolish activity (Pelham, 1982). Taking all the available data into account, we conclude that if there is any requirement for an inverted repeat in COS cells, it need be no longer than 10 bp, and its sequence and exact position are unimportant. The presence of such a feature in the same relative position in most of the *Drosophila* heat-shock genes certainly suggests that it has some function *in vivo*, but this function may be too subtle to be revealed by

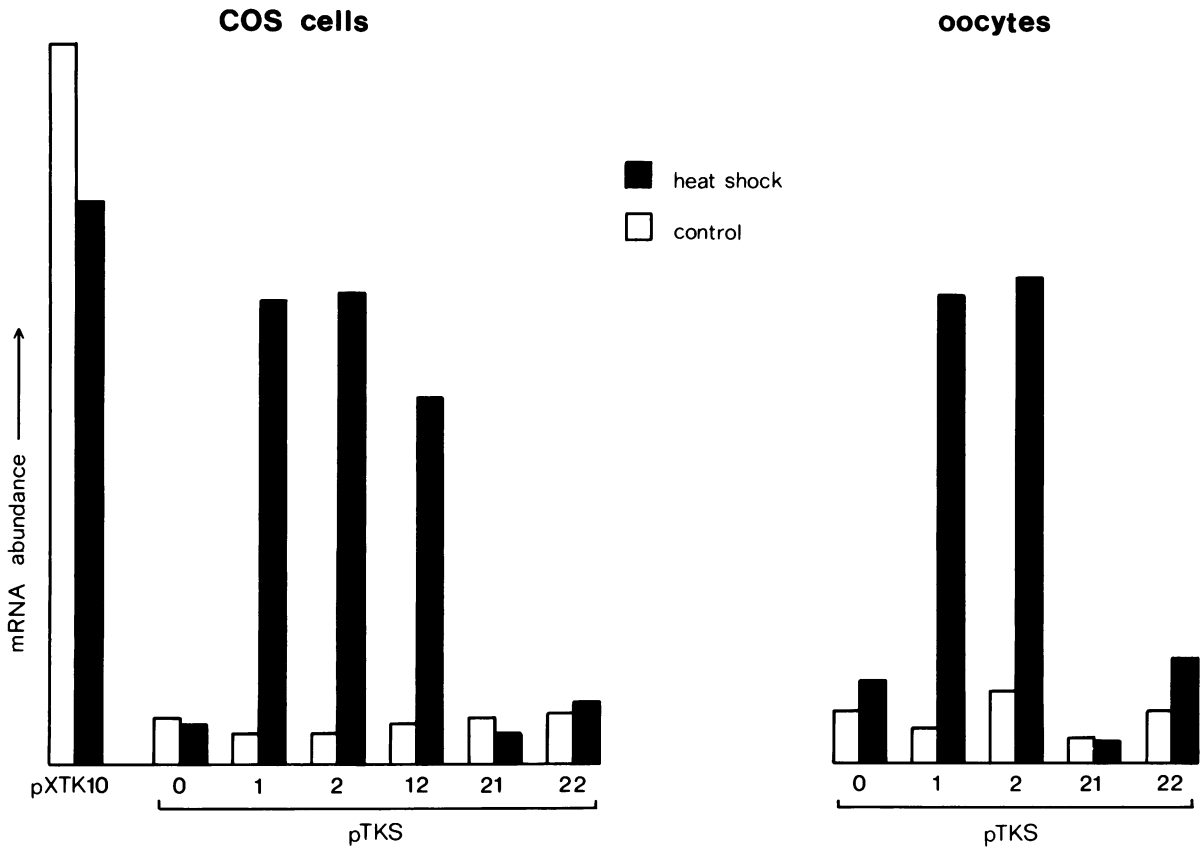

Fig. 4. Summary of results obtained in COS cells and oocytes. The bands corresponding to correctly-initiated transcripts in experiments such as that shown in Figure 3 were quantitated by densitometry. The data in this figure is compiled from five separate experiments, and each of the RNA samples was S1-mapped at least twice. Average values are shown; variability between duplicate experiments can be as great as 40%.

	Consensus CT GAA TTC AG	Bases from TATA	Bases matching consensus	Inducible activity
hsp 83	CTAGAAGTTTCTAG	26	10	(yes)
hsp 70	CTCGAATGTTCGCG	15	9	yes
hsp 26	CCGGACTCTTCTAG	15	8	(yes)
hsp 22	CCGGTATTTTCTAG	28	8	(yes)
hsp 68	CTCGAATTTTCCCC	14	8	(yes)
hsp 23	CGAGAAGTTTCGTG	99	8	(yes)
hsp 27	TTGCCATGCACTAG	21	5	(yes)
pTKS1, pTKS2	CTAGAAGCTTCTAG	13, 13/23	10	yes
pTKS12	CTAGAGATCTCTAG	13/23	8	yes
pHT5	CTCGAATGTTCGGC	19	8	yes
pHT6, pHT7	CTCGTTGGTTCGAG	12, 14	8	weak
pTKS0	TCGGTCGACTCTAG	13	5	no
pTKS22	CAAGCTTGCGCAAG	22	5	no
pHTΔ5'-44	GCTGAGTGGCCAAG	16	5	no
pTKS21	AGGCAAGCTTGCCT	15	4	no

Fig. 5. Comparison of sequences from heat-shock genes and mutants. The upper portion of the figure lists the best fits to the proposed functional sequence in the various *Drosophila* heat-shock genes. The sequences are from Karch *et al.* (1981), Holmgren *et al.* (1981), and Ingolia and Craig (1981). Although all the genes are inducible *in vivo*, only the *hsp70* gene has been assayed in the same way as the mutants shown in the lower part of the figure. Note that pTKS2 and pTKS12 have duplicate copies of the sequence shown. The matches shown for the uninducible mutants are somewhat arbitrary, and several other 4- or 5-base matches can be found at different positions. Data for the *hsp70* mutants pHT5, pHT6, pHT7, and pHTΔ5'-44 are from Pelham (1982).

our assays. Interestingly, the cadmium-inducible mouse metallothionein I gene has a large inverted repeat in about the same position, and there is evidence that an important promoter element maps close to it (Brinster *et al.*, 1982).

There is evidence that sequences outside the consensus can affect the efficiency of heat induction. We have found that deletion mutants of the *hsp70* gene that retain the consensus sequence and only 4 – 6 bp upstream of this frequently (but not always) fail to respond to heat-shock in *Xenopus* oocytes, whereas plasmids with longer flanking sequences respond reproducibly (Bienz and Pelham, 1982). Thus, surrounding features of the DNA sequence may have some function, perhaps in the organization of the local chromatin structure.

Upstream promoter elements

Important elements upstream from the TATA box have been identified for a number of eucaryotic promoters, and mapped in detail for the β-globin gene (Grosveld *et al.*, 1982), the *tk* gene (McKnight and Kingsbury, 1982), and now for the heat-shock genes. It is striking that, despite attempts to find a universal promoter sequence in this region (e.g., Benoist *et al.*, 1980), all of the elements are quite different. There are of course functional differences between these promoters: the β-globin promoter requires, in addition to its upstream element, a remote 'enhancer' sequence for efficient expression (Banerji *et al.*, 1981), and the heat-shock genes are stress-inducible. The diversity of the functional sequences suggests that there may be a variety of factors in cells that each interact only with a specific type of promoter. The upstream elements might thus be an important site of regulation, not only for inducible genes but also, perhaps, for developmentally regulated ones.

In the future it may be possible to identify proteins that interact with these upstream promoter sequences and to study their function. The heat-shock element seems particularly suitable for such studies, being simple and well-defined, and having a clear regulatory role.

Materials and methods

Construction of plasmids

The construction of pXTK10 has been described in detail elsewhere (Pelham, 1982). The *Bam*HI-*Kpn*I fragment from the *tk* gene in pXTK10 (Figure 1) was replaced with the corresponding fragment of the deletion mutant pTKΔ5′-32 (McKnight *et al.*, 1981). The small *Bam*HI-*Sal*I fragment was then replaced with an *Xba*I linker (CTCTAGAG) after filling in the sticky ends with DNA polymerase, to form pTKS0 (Figures 1 and 2). The adapter CTAGAAGCTT was synthesized by the solid-phase phosphotriester method (Gait *et al.*, 1982), and was kindly provided by M.Singh. An adapter GATCTCTAGA had been previously synthesized, and was permuted to CTAGAGATCT by ligation and digestion with *Xba*I. These adapters were ligated into the *Xba*I site of pTKS0 in 5- to 50-fold molar excess, and individual clones characterized by digestion with *Bam*HI, labelling with DNA polymerase, digestion with *Sal*I, and sizing of the insert on a sequencing gel. Plasmids with large numbers of adapters inserted with unstable in *Escherichia coli* HB101, deleting out some of the adapters cleanly. Stable plasmids with 1 – 5 copies of CTAGAAGCTT were obtained, but of eight plasmids with inserts of CTAGAGATCT, all retained just two copies.

Assay of plasmids

Expression of the plasmids in COS cells and *Xenopus* oocytes and analysis of the transcripts was performed as outlined in the text and figures, and as described elsewhere (Pelham, 1982; Bienz and Pelham, 1982).

Acknowledgements

We thank Dr.S.McKnight for supplying the thymidine kinase deletion plasmid, and members of the oligonucleotide synthesis group at the M.R.C. Laboratory of Molecular Biology for the synthetic adapters. We also thank Dr.A.Travers for helpful comments on the manuscript. M.B. is the recipient of an EMBO fellowship.

References

Banerji,J., Rusconi,S., and Schaffner,W. (1981) *Cell,* **27**, 299-308.
Benoist,C., O'Hare,K., Breathnach,R., and Chambon,P. (1980) *Nucleic Acids Res.,* **8**, 127-142.
Bienz,M., and Pelham,H.R.B. (1982) *EMBO J.,* **1**, in press.
Brinster,R.L., Chen,H.J., Warren,R., Sathy,A., and Palmiter,R.D. (1982) *Nature,* **296**, 39-42.
Burke,J.F., and Ish-Horowicz,D. (1982) *Nucleic Acids Res.,* **10**, 3821-3830.
Corces,V., Pellicer,A., Axel,R., and Meselson,M. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 7038-7042.
Gait,M.J., Matthes,H.W.D., Singh,M., Sproat,B.S., and Titmas,R.C. (1982) *Nucleic Acids Res.,* **10**, 6243-6254.
Gluzman,Y. (1981) *Cell,* **23**, 175-182.
Grosveld,G.C., Rosenthal,A., and Flavell,R.A. (1982) *Nucleic Acids Res.,* **10**, 4951-4971.
Holmgren,R., Corces,V., Morimoto,R., Blackman,R., and Meselson,M. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 3775-3778.
Ingolia,T.D., and Craig,E.A. (1981) *Nucleic Acids Res.,* **9**, 1627-1642.
Karch,F., Torok,I., and Tissières,A. (1981) *J. Mol. Biol.,* **148**, 219-230.
Lusky,M., and Botchan,M. (1981) *Nature,* **293**, 79-81.
McKnight,S.L., Gavis,E.R., Kingsbury,R., and Axel,R. (1981) *Cell,* **25**, 385-398.
McKnight,S.L., and Kingsbury,R. (1982) *Science (Wash.),* **217**, 316-324.
Mirault,M.-E., Delwart,E., and Southgate,R. (1982) in Schlesinger,M., Ashburner,M., and Tissières,A. (eds.), *Heat Shock from Bacteria to Man,* Cold Spring Harbor Laboratory Press, NY, pp. 244-248.
Pelham,H.R.B. (1982) *Cell,* **30**, 517-528.
Pelham,H., and Bienz,M. (1982) in Schlesinger,M., Ashburner,M., and Tissières,A. (eds.), *Heat Shock from Bacteria to Man,* Cold Spring Harbor Laboratory Press, NY, pp. 9-12.
Schlesinger,M., Ashburner,M., and Tissières,A. (eds.) (1982) *Heat Shock from Bacteria to Man,* published by Cold Spring Harbor Labortory Press, NY.
Singleton,C.K., Klysik,J., Stirdivant,S.M., and Wells,R.D. (1982) *Nature,* **299**, 312-316.
Voellmy,R., and Rungger,D. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 1776-1780.