The EMBO Journal vol.10 no.2 pp.369-375, 1991

A conserved heptapeptide restrains the activity of the yeast heat shock transcription factor

Bent K.Jakobsen and Hugh R.B.Pelham[1]

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

[1]Corresponding author

Communicated by H.R.B.Pelham

In yeast, expression of heat shock genes is regulated by a factor (HSF) which binds constitutively to DNA, but activates transcription efficiently only after heat shock. We have compared the HSFs from *Saccharomyces cerevisiae* and *Kluyveromyces lactis*. Both factors contain an activation domain whose activity is masked at low temperature, but the amino acid sequences of these activators are unrelated. Masking requires the evolutionarily conserved DNA binding and oligomerization domains, as well as a short conserved element close to the activator. Although this element contains potential phosphorylation sites, they are not required for induction. We suggest that the conserved element binds either to the structural core of the protein or to another polypeptide, holding the activator in an inactive configuration, and that high temperatures disrupt this interaction. Our results emphasize the importance of global protein structure in the regulation of transcription factor activity.

Key words: heat shock/*K.lactis*/*S.cerevisiae*/transcription factors

Introduction

Cells from a wide variety of species respond to elevated temperatures by activating the transcription of a small set of heat shock genes (for reviews see Craig, 1986; Lindquist, 1986; Bienz and Pelham, 1987; Morimoto *et al.*, 1990). This heat shock response can also be induced by other treatments that damage or denature intracellular proteins. The response is protective: there is evidence that heat shock proteins aid the renaturation or destruction of damaged proteins, and can save cells from the otherwise lethal consequences of hyperthermia. Heat shock proteins also appear to regulate their own synthesis by feedback inhibition (Craig, 1990; Stone and Craig, 1990; Yost *et al.*, 1990).

In eukaryotes, the key component of this regulatory system is a specific DNA-binding protein (the heat shock factor, HSF) which binds to sites that are found upstream of all heat shock genes (for reviews see Bienz and Pelham, 1987; Morimoto *et al.*, 1990). HSF has been purified from several species, and the gene encoding the *Saccharomyces cerevisiae* factor has been isolated. *S.cerevisiae* HSF has a trimeric structure (Sorger and Nelson, 1990). It binds to DNA even at low temperatures and is essential for growth; heat shock induces phosphorylation of the factor and increases its ability to stimulate transcription *in vivo* (Sorger *et al.*, 1987; Jakobsen and Pelham, 1988; Wiederrecht *et al.*, 1988; Sorger and Pelham, 1988). In contrast, the HSF of higher eukaryotes such as *Drosophila* and humans binds to DNA only after heat shock (Zimarino and Wu, 1987; Kingston *et al.*, 1987; Sorger *et al.*, 1987). This change in binding affinity can be induced *in vitro*, and may reflect a change in the conformation or oligomerization state of the factor (Larsen *et al.*, 1988; Mosser *et al.*, 1990; Zimarino *et al.*, 1990).

A major unanswered question in the yeast system is how the activity of DNA-bound HSF is regulated. Deletion analysis of the *S.cerevisiae* factor has defined its basic features. These include a DNA-binding domain, a helical trimerization domain, and a C-terminal region which activates transcription strongly when it is fused to a heterologous DNA-binding protein (Wiederrecht *et al.*, 1988; Sorger and Nelson, 1990; Sorger, 1990; Nieto-Sotelo *et al.*, 1990). However, the precise way in which these domains interact to regulate HSF activity remains unclear.

As a complementary approach to this problem, we have isolated the HSF gene from another budding yeast, *Kluyveromyces lactis*, and looked for conserved sequences that might identify functional domains. The *K.lactis* factor is structurally and functionally related to the *S.cerevisiae* one, but whereas the DNA-binding and trimerization domains are closely conserved, the C-terminal activator has a completely divergent sequence. As with *S.cerevisiae* HSF (Sorger, 1990; Nieto-Sotelo *et al.*, 1990), the C-terminal activator is regulated by other parts of the protein. Regulation involves the main structural domains of the factor, and is dependent on a conserved heptapeptide sequence that is distinct from the activator itself. Although this heptapeptide is flanked by multiple serine residues in both *S.cerevisiae* and *K.lactis* HSF, phosphorylation of these serines is not required for activation. We suggest that induction of HSF activity involves the release of the conserved heptapeptide from a binding site elsewhere on HSF, or on another molecule, causing a significant conformational change to occur in the protein.

Results

Isolation of the K.lactis HSF gene

A Southern blot of *K.lactis* DNA was probed at low stringency with a fragment encoding the DNA-binding domain of *S.cerevisiae* HSF (ScHSF). A single cross-hydridizing gene was identified and cloned (see Materials and methods). Sequencing of this *K.lactis* HSF (KlHSF) gene revealed an open reading frame of 677 amino acids.

Four regions of the encoded protein are homologous to ScHSF (Figure 1). The two largest ones correspond to the DNA-binding domain (amino acids 194-298 in the KlHSF sequence) and the trimerization domain (amino acids 320-372), and are the only regions of the protein predicted to have significant α-helix content. The other two are short sequences: a 10/11 match between amino acids 62 and 73

```
K.l.   1  MGHNDSVETMDEISNPNNILLPHDGTGLDATGISGSQEPYGMVDVLNPDSLKDDSNVDEP
S.c.   1  MNNAANTGTTNESNVSDAPRI-EPLPSLNDDDIEKILQP---NDIFTTDR-TDASTTSST

K.l.  61  LIEDIVNPSLDPEGVVSAEPSNEVGTPLLQQPISLDHVITRPASAGGVYSIGNSSTSSAA
S.c.  56  AIEDIINPSLDPQSAASPVPSSSF-FHDSRKPSTSTHLV-RRGTPLGIYQTNLYGHNSRE

K.l. 121  KLSDGDLTNATDPLLNNAHGHGQPSSESQSHSNGYHKQGQSQQPLLSLNKRKLLAKAHVD
S.c. 114  N-TNPNSTLLSSKLLAHPPVPYGQNPDLLQHA-VYRAQPSS------------GTTNAQ

K.l. 181  KHHSKKKLSTTRARPAFVNKLWSMVNDKSNEKFIHWSTSGESIVVPNRERFVQEVLPKYF
S.c. 159  PRQTTRRYQSHKSRPAFVNKLWSMLNDDSNTKLIQWAEDGKSFIVTNREEFVHQILPKYF

K.l. 241  KHSNFASFVRQLNMYGWHKVQDVKSGSMLSNNDSRWEFENENFKRGKEYLLENIVRQKSN
S.c. 219  KHSNFASFVRQLNMYGWHKVQDVKSGSIQSSSDDKWQFENENFIRGREDLLEKIIRQKGS

K.l. 301  TN-----------------------------------------ILGGTT---------
S.c. 279  SNNHNSPSGNGNPANGSNIPLDNAAGSNNSNNNISSSNSFFNNGHLLQGKTLRLMNEANL

K.l. 309  NAEVDIHILLNELETVKYNQLAIAEDLKRITKDNEMLWKENMMARERHQSQQQVLEKLLR
S.c. 339  GDKNDVTAILGELEQIKYNQIAISKDLLRINKDNELLWQENMMARERHRTQQQALEKMFR

K.l. 369  FLSSVFGPNSAKTIGNGFQPDLIHELSDMQVNHMSNNNHNNTGNIN------------PN
S.c. 399  FLTSIVPHLDPKMIMDGLGDPKVNN-EKLNSANNIGLNRDNTGTIDELKSNDSFINDDRN

K.l. 417  AYHNETDDP-----------------------------------------------MANVF
S.c. 458  SFTNATTNARNNMSPNNDDNSIDTASTNTTNRKKNIDENIKNNNDIINDIIFNTNLANNL

K.l. 431  GPLTPTDQGKVPLQDYKLRPRLLLKNRSMSSSSSSNLN----QRQSPQNRIVGQSPPPQQ
S.c. 517  SNYNSNNNAGSPIRPYK--QRYLLKNRANSSTSSENPSLTPFDIESNNDRKISEIPFDDE

K.l. 487  QQQQ--------QQQQGQPQGQQFS---YPIQGGNQMMNQLGSPIGTQVGSPVGSQYGNQ
S.c. 576  EEEETDFRPFTSRDPNNQTSENTFDPNRFTMLSDDDLKKDSHTNDNKHNESDLFWDNVHR

K.l. 536  YGNQ------------------YSNQ-FGNQLQQQTSRPALHHGSNGEIRELT-PSIVSS
S.c. 636  NIDEQDARLQNLENMVHILSPGYPNKSFNNKTSSTNTNSNMESAVNVNSPGFNLQDYLTG

K.l. 576  DSPDPAFFQDL--------------QNNIDKQEESIQEIQD------WIT----KLNPGP
S.c. 696  ESNSPNSVHSVPSNGSGSTPLPMPNDNDTEHASTSVNQGENGSGLTPFLTVDDHTLNDNN

K.l. 612  GEDGNTPIFPEL------------NMPSYFANTGGSGQSEQPSDYGDSQIEELRNSRLHE
S.c. 756  TSEGSTRVSPDIKFSATENTKVSDNLPSFNDHSYSTQADTAPENAKKRFVEEIPEPAIVE

K.l. 660  PDRSFEEKNNGQKRRRAA
S.c. 816  IQDP-TEYNDHRLPKRAKK
```

Fig. 1. Comparison of the amino acid sequences of *K.lactis* and *S.cerevisiae* HSF. Hyphens indicate spaces introduced to maximize homology; asterisks mark identical residues. The four conserved regions are underlined. The DNA sequence of the *K.lactis HSF* gene has been submitted to EMBL (accession number X55149).

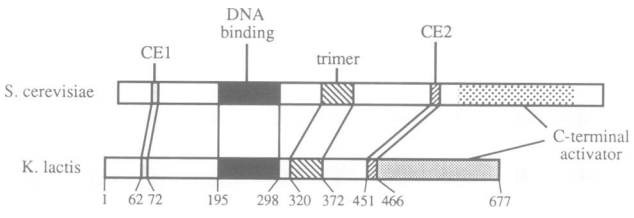

Fig. 2. Schematic overview of the structures of KlHSF and ScHSF. Homologous regions are shaded equivalently: the DNA binding domain, the trimerization domain (trimer), and the two short conserved elements (CE1, CE2). The regions containing the unrelated C-terminal activators are also indicated.

and an 11/16 match between amino acids 451 and 466; we refer to these conserved elements as CE1 and CE2 respectively (see Figures 1 and 2). The sequences flanking the four homologous regions are completely divergent, and their lengths vary considerably between the ScHSF and KlHSF proteins.

Table I. KlHSF function in *S.cerevisiae*

Promoter	*HSF* gene	β-galactosidase activity[a]	
		con	hs
ScHSF	*S.cerevisiae*	16	240
ScHSF	*K.lactis*	4	285
TPI	*K.lactis*	10	60

[a]*HSF* genes were expressed from centromere-containing plasmids using the indicated promoters, the chromosomal *HSF* gene being deleted. β-galactosidase, expressed from an indicator plasmid carrying a synthetic HSF binding site, was assayed in cells grown at 30°C (con) or after heat shock (hs). β-galactosidase activity is expressed in the units defined by Breeden and Nasmyth (1987).

K.lactis HSF can substitute functionally for S.cerevisiae HSF

HSF is essential for growth of *S.cerevisiae*. To see whether the *K.lactis HSF* gene could substitute for the endogenous one, we introduced a plasmid expressing KlHSF into a strain which lacked the chromosomal HSF gene, but carried a copy of ScHSF on a *URA3*-containing plasmid. The strain containing both HSF genes was then treated with 5-fluoro-orotic acid to select against the *URA3* gene. This procedure resulted in a viable strain expressing only *K.lactis* HSF, indicating that the *K.lactis* gene was functional in *S.cerevisiae*.

We then tested the ability of KlHSF to support heat-inducible transcription in the heterologous species. Strains containing only KlHSF were transfected with an indicator plasmid bearing a β-galactosidase gene downstream of a promoter with a synthetic HSF binding site. Table I shows the results obtained with a strain that expressed KlHSF from the ScHSF promoter. Heat shock produced a large increase in β-galactosidase activity in this strain.

Rather surprisingly, we found that increasing the level of KlHSF expression by fusing it to the strong promoter of the triose phosphate isomerase (TPI) gene reduced the amount of β-galactosidase activity produced after heat shock (Table I). One explanation for this result could be that high levels of the active form of HSF compete for some component of the transcription machinery, as has been suggested for other activators. Alternatively, autoregulation of the heat shock response may occur: we have shown that overexpression of HSF leads to enhanced levels of hsp70 at low temperature (Sorger and Pelham, 1988), and high levels of hsp70 are thought to turn off the heat shock response. To avoid this problem in subsequent experiments, KlHSF was always expressed from the ScHSF promoter.

We next confirmed that the *K.lactis* factor, like the *S.cerevisiae* one, remains bound to DNA even under normal growth conditions. Previously, we showed that when overlapping binding sites for HSF and the GAL4 protein were introduced into yeast cells, constitutive binding of HSF was sufficient to prevent binding of GAL4. When these overlapping sites are placed upstream of a β-galactosidase gene little galactose-inducible (GAL4-dependent) enzyme activity is produced at 30°C, but expression is heat-inducible. Base changes that disrupt binding of HSF allow GAL4 to bind, and result in high levels of activity on galactose medium (Jakobsen and Pelham, 1988). Table II shows that the same phenomenon could be observed with *K.lactis* HSF, implying that it too binds to DNA at low temperature.

Table II. Constitutive binding of KlHSF to DNA *in vivo*

	β-galactosidase activity			
	Glucose		Galactose	
	Con	HS	Con	HS
~~HSF~~ ~~GAL4~~	2	4	1	2
HSF GAL4	10	95	16	57
~~HSF~~ GAL4	3	5	705	587

Strains expressed only KlHSF, and contained indicator plasmids with overlapping binding sites for HSF and GAL4, or mutated versions of these. They were grown in glucose or galactose medium at 30°C (Con) or heat shocked (HS).

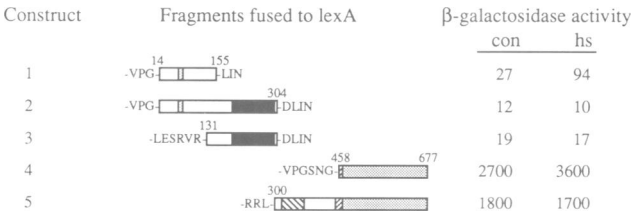

Fig. 3. Assay of lexA fusion proteins. The portions of KlHSF fused to the C terminus of the 87 amino acid lexA DNA binding domain are indicated schematically. Numbers above the boxes show the endpoints of the fragments, whose alignment corresponds to their position in the KlHSF sequence; any unnumbered endpoint is identical to one in a previous construct. Functional regions are shaded as in Figure 2. Amino acids introduced at junctions or at the C terminus are indicated. β-galactosidase activity expressed from a reporter gene bearing a lexA binding site is shown in the same units as in Table I. Note that, to avoid problems of heterotrimerization, construct 5 was assayd in a different yeast strain (see Materials and methods). The levels of activity obtained with this construct are not therefore strictly comparable to those obtained with the others.

K.lactis HSF contains a C-terminal activator whose sequence is not conserved

To search for potential transcriptional activators within the KlHSF sequence, we expressed portions of the protein fused to the DNA binding domain of the prokaryotic repressor lexA. The activity of these fusion proteins was assayed using an indicator plasmid with a lexA binding site. Figure 3 shows that the C-terminal portion of KlHSF activated transcription strongly, whether or not the cells were heat shocked (construct 4). In contrast, fragments from the N-terminal half of the protein had little activity (constructs 1–3). The trimerization domain alone was not tested, but it is known that the equivalent region of ScHSF does not constitute a strong transcriptional activator (Sorger, 1990; Nieto-Sotelo *et al.*, 1990).

We conclude that the C-terminal portion of *K.lactis* HSF contains a transcriptional activator which, when transplanted to another protein, functions constitutively. The C terminus of ScHSF contains an activator with similar properties (Sorger, 1990; Nieto-Sotelo *et al.*, 1990); however, the two activators are completely different in sequence (see Figure 1).

Removal of the C-terminal activator reduces the activity of K.lactis HSF following heat shock

For further analysis of the *K.lactis* HSF molecule, we expressed mutant forms of the protein in a strain that lacked endogenous HSF. The upper part of Figure 4 shows the effect of removing the C-terminal activator. Strains containing the truncated factor produced ~4-fold less

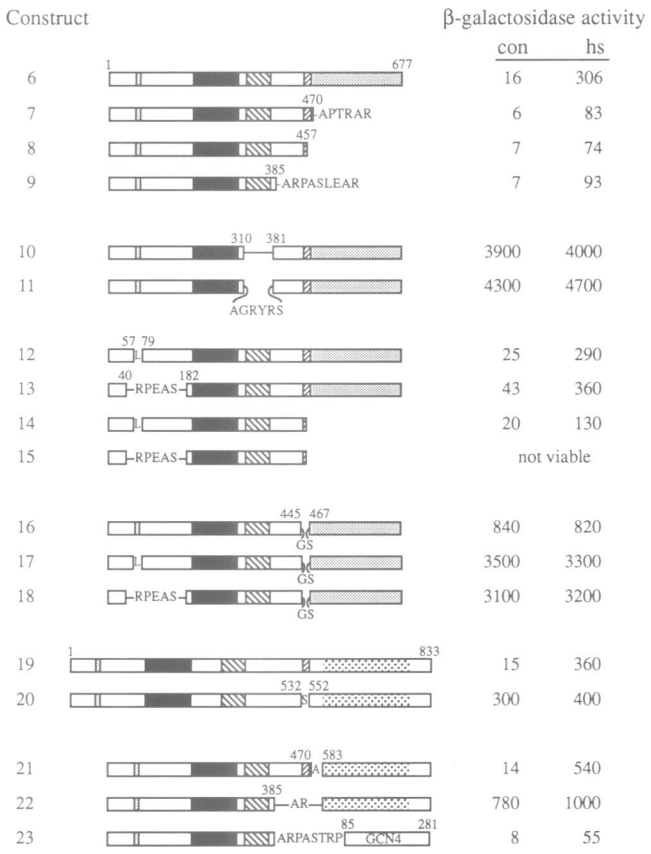

Fig. 4. Assay of HSF mutants. Constructs are drawn with the same conventions as those used in Figure 3. Construct 6 is wild-type KlHSF. Other constructs are mainly derived from this, but construct 19 is wild-type ScHSF, construct 20 is a deletion mutant of ScHSF, constructs 21 and 22 contain the C-terminal activator of ScHSF fused to KlHSF sequences, and construct 23 contains part of the *S.cerevisiae* GCN4 protein, as indicated. All constructs were assayed in strains containing no other HSF protein. β-galactosidase activity was expressed from a reporter plasmid with a synthetic HSF binding site; a control plasmid with a non-functional binding site gave <2 units of activity.

β-galactosidase activity after heat shock than did similar strains expressing the intact protein (compare constructs 7–9 with construct 6). In addition, they were unable to grow at 37°C. In both these regards, the KlHSF mutants behave similarly to equivalent ScHSF mutants.

Although the N-terminal part of HSF can undoubtedly stimulate transcription in a heat-inducible manner, its activity is low. At least 75% of the activity obtained after heat shock is dependent on the presence of the C-terminal activator. Since this activator is constitutive when fused to lexA, its function cannot depend on any biochemical change that occurs to the cell after heat shock. Thus, a key property of the rest of the HSF molecule must be its ability to mask the C-terminal activator, and reveal it only in response to stress.

HSF domains required for regulation of the C-terminal activator

All our assays indicated that the *K.lactis* and *S.cerevisiae* HSFs have a similar overall structure, and function in a broadly similar manner, Strikingly, the portions of the HSF protein known to play a structural role are highly conserved in sequence. Thus, it seems likely that if the masking of the C-terminal activator had a structural basis, the residues involved in this phenomenon would also be conserved. We

therefore investigated which of the four conserved regions were required for correct regulation of the C-terminal activator.

First, we prepared KlHSF mutants that lacked the trimerization domain. Surprisingly, even though HSF is essential for viability and trimerization of ScHSF is important for DNA binding (Sorger and Nelson, 1990), strains containing such mutations grew well. Trimerization of KlHSF is evidently not essential for DNA binding; it may be that some form of oligomerization can be induced by another part of the protein, and that this is sufficient to stabilize the KlHSF–DNA complex. However, the deletion mutants had very high constitutive activity (constructs 10 and 11, Figure 4), comparable to the activity of lexA fusions containing the C-terminal activator (Figure 3). We conclude that repression of the activator requires the normal oligomerization state of the HSF molecule.

To test the involvement of the DNA-binding domain in regulation, we replaced it, and the N-terminal region, with the DNA binding domain of lexA (construct 5, Figure 3). Because this construct retained the trimerization domain, it could potentially form mixed trimers with endogenous HSF, whose activity would be unpredictable (Sorger, 1990). To avoid this problem, we assayed it in a yeast strain containing KlHSF with the trimerization domain deleted. The fusion protein had high constitutive activity, suggesting that trimerization alone is unable to repress the C-terminal activator at low temperature. It seems that regulation requires both of the major conserved structural elements of the protein.

We next tested the activity of KlHSF mutants lacking CE1. These mutants showed a slightly elevated basal level of transcription, but heat-inducibility was not affected (constructs i2 and 13, Figure 4). Similarly, deletion of CE1 from a truncated form of KlHSF that lacked the C-terminal activator had only a slight stimulatory effect on the residual activity of the factor (construct 14). Thus amino acid sequences on the N-terminal side of the DNA-binding domain seem to play only a minor role in the repression of KlHSF activity at low temperature. N-terminal sequences other than CE1 may, however, contribute to the activity of the factor: removal of both the C-terminal activator and amino acids 41–181 (construct 15) resulted in a strain that was non-viable, presumably because the remaining portion of the factor was unable to activate transcription.

Deletion of CE1 from ScHSF also had little effect (Sorger, 1990). A larger N-terminal deletion did cause constitutive activation of the *S.cerevisiae* factor; however, because this effect involves an unconserved part of the molecule, and was not seen with KlHSF, it probably does not identify a fundamental structural feature of the protein.

A conserved motif involved in regulation

One other short region of homology between the *K.lactis* and *S.cerevisiae* factors is found between the trimerization domain and the C-terminal activator (CE2, Figures 1 and 2). Previous studies have shown that this portion of ScHSF is not required for trimerization or DNA binding, and does not form part of the C-terminal activator. However, deletion of CE2 from either KlHSF or ScHSF resulted in constitutive activity at a level comparable to, or higher than, that normally seen after heat shock (compare construct 6 with 16, and 19 with 20, Figure 4). A particularly dramatic effect was apparent when CE2 was removed from constructs lacking CE1—this increased the HSF at 30°C up to 140-fold (compare constructs 12 and 13 with 17 and 18, Figure 4). This result suggests that the CE2 motif is required to maintain the repressed state of HSF, and raises the possibility that activation of HSF following heat shock involves changes in this region of the protein. CE1 also seems to contribute to repression, but it is clearly subsidiary to CE2 in its function.

We next asked whether the sequence of CE2 was crucial, or whether deletions in this part of HSF merely perturbed some larger, precisely-folded structure that involved the C-terminal activator. If such a structure were to exist, it would probably be different for the two species of HSF, since the sequences surrounding CE2 are so different. We therefore constructed a chimeric molecule in which the C-terminal activator from ScHSF was fused to the N-terminal part of KlHSF, the junction being a few amino acids C-terminal to the element; this placed the activator 26 residues closer to CE2 than it is in the ScHSF sequence. Despite this altered spacing and the sequence divergence of the activator, the fusion protein was correctly regulated (construct 21, Figure 4). In contrast, fusion of the ScHSF activator close to the trimerization region, eliminating CE2, resulted in constitutive activity (construct 22); this confirms the importance of the motif.

In an attempt to test the generality of the regulatory mechanism, we also tried replacing the C-terminal activator of KlHSF with the activator of the yeast GCN4 protein. This activator functions well when fused to the lexA DNA-binding domain (not shown; Hope *et al.*, 1988), but showed no more activity when joined to KlHSF sequences than did the corresponding deletion mutant that lacked GCN4 sequences, even when the fusion point was on the N-terminal side of CE2 (construct 23, Figure 4). It appears, therefore, that the GCN4 activator has properties that distinguish it from the HSF activators, and not all activator–HSF fusion constructs are functional.

Phosphorylation of the conserved element is not required for activation

CE2 consists of two adjacent regions: a homology consisting principally of the sequence RXLLKNR (X = L in KlHSF, Y in ScHSF), and a run of serines (with one threonine in the case of ScHSF). Because activation of ScHSF correlates with its phosphorylation on serine residues (Sorger, 1990), it seemed possible that modification of CE2 was an essential part of the induction mechanism. To test this we introduced mutations into intact KlHSF, changing the serine residues of CE2 to other amino acids. Table III shows that changing six of the seven serines to aspartic acid residues, or all seven to alanines, did not affect the inducibility of HSF activity. A more detailed analysis of the response to temperatures between 20°C and 40°C (Figure 5), and to various times of heat shock and recovery (not shown) also failed to show any effect of the mutations. This shows that, despite their conservation, the serines are not crucial for repression, and full induction can be achieved without phosphorylation of the protein in this region.

In contrast, changing the RLLLKNR sequence to ALAAAAA or to LRLNQLQ caused a strong constitutive activation both with intact KlHSF and with the hybrid factor carrying the ScHSF activator (Table III); high activity was retained at temperatures as low as 15°C (not shown). Removal of the C-terminal activator from one of the constitutive mutants abolished activity (Table III, construct

Table III. Effects of mutations in CE2

CE2 sequence	β-galactosidase activity	
	con	hs
In KlHSF:		
RLLLKNRSMSSSSSS	16	310
RLLLKNRAMAAAAAA	15	350
RLLLKNRSMDDDDDD	26	310
ALAAAAASMSSSSSS	430	500
LRLNQLQSMSSSSSS	930	800
In construct 21[a]:		
RLLLKNRSMSSSSSS	14	540
ALAAAAASMSSSSSS	290	660
In construct 7[a]:		
RLLLKNRSMSSSSSS	6	80
ALAAAAASMSSSSSS	3	8

[a]Construct 21 has the C-terminal activator of KlHSF replaced by the equivalent portion of ScHSF; construct 7 is KlHSF without the C-terminal activator (see Figure 4). The first CE2 sequence given for each construct is the wild-type one.

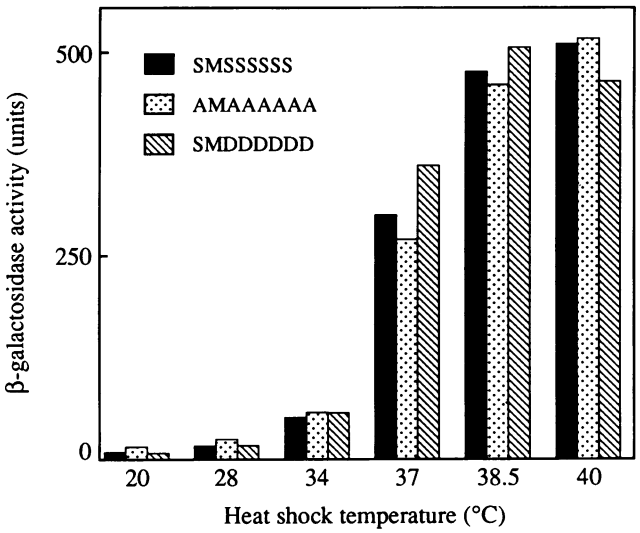

Fig. 5. The serine residues of CE2 are not required for heat induction. Strains containing KlHSF with the wild-type CE2 sequence (filled bars), the seven serines changed to alanines (dotted bar) or the last six serines changed to aspartic acid residues (striped bars) were subjected to heat shocks at the indicated temperatures.

7). We conclude that the conserved sequence RXLLKNR plays an essential role in the regulation of the C-terminal activator at low temperature.

Discussion

The results presented here illustrate the value of inter-species comparisons in the analysis of protein structure. The *S.cerevisiae* and *K.lactis* HSF proteins are functionally interchangeable, yet show a remarkable degree of sequence divergence. The only regions of extensive homology accurately define the two known structural elements: the DNA binding domain and the coiled coil motif involved in trimerization.

In marked contrast, the transcriptional activation domains found at the C terminus of each protein are unrelated at the sequence level. *K.lactis* HSF is glutamine-rich in this region, and contains a long uncharged stretch that includes 11 tandem tetrapeptides with the consensus sequence Q-(hydrophobic)-G-(N,S), followed by a moderately acidic region. Apart from general abundance of acidic and amide residues, the only obvious property in common with the *S.cerevisiae* activator is a predicted lack of α-helical secondary structure: for KlHSF this region is predicted to be only 9% α-helix, and nearly 70% irregular structure or turns. Nevertheless, these activators are functionally interchangeable. The weaker activity that remains after removal of the C terminus has proved harder to localize, but studies of both factors suggest that sequences on the N-terminal side of the DNA binding domain, other than CE1, are involved (this work; Sorger, 1990; Nieto-Sotelo *et al.*, 1990). The amino acid sequence of this region is also unconserved and it has little predicted helical or β-sheet structure. It seems than an activator with a precisely conserved structure is not required either for the stimulation of transcription by HSF or for the regulation of this function in response to heat shock.

A regulatory motif in the HSF molecule

The intrinsically constitutive nature of the C-terminal activator implies that it is repressed under normal growth conditions. Repression is imposed by the rest of the HSF molecule, and we have shown that it requires both the trimerization and DNA-binding domains, the likely structural core of the protein. Of these sequences, the trimerization domain is likely to be most important, because recent studies suggest that at least partial regulation of the ScHSF protein can be achieved without the DNA binding domain (Nieto-Sotelo *et al.*, 1990).

Our most striking observation is that the short conserved sequence RXLLKNR, located between the trimerization domain and the C-terminal activator, is crucial for repression of HSF activity at low temperature. This sequence is not required for trimer formation or for DNA binding (Sorger and Nelson, 1990), and does not form part of the transcriptional activator.

What could be the role of this sequence? Although short, it is certainly large enough to be recognized specifically by some other protein domain. It might, for example, bind to a specific site on one of the conserved, structured regions of HSF—the DNA binding domain or the trimerization domain. This could induce an inactive conformation of the protein by tying the C-terminal activator to a rigid structure in a way that is incompatible with the formation of an active transcription complex (see Figure 6). Alternatively, the conserved element could bind to some other protein, with similar consequences (see below). Deletion or mutation of the element would release the C terminus and allow it to interact with other components of the transcription apparatus.

Induction of HSF activity following heat shock

Studies of the mobility of ScHSF–DNA complexes on native gels have shown that heat shock does not affect either the affinity of HSF for DNA or its basic trimeric structure, but does result in a reduction in gel mobility that can most easily be explained by a conformational change in the protein (Sorger *et al.*, 1987). It seems reasonable to suggest that this change is triggered by the release of the conserved motif from its binding site, and is a major cause of the increased activity of HSF after heat shock.

If this model is correct, the crucial regulatory event is the initial conformational change in HSF. What could cause this?

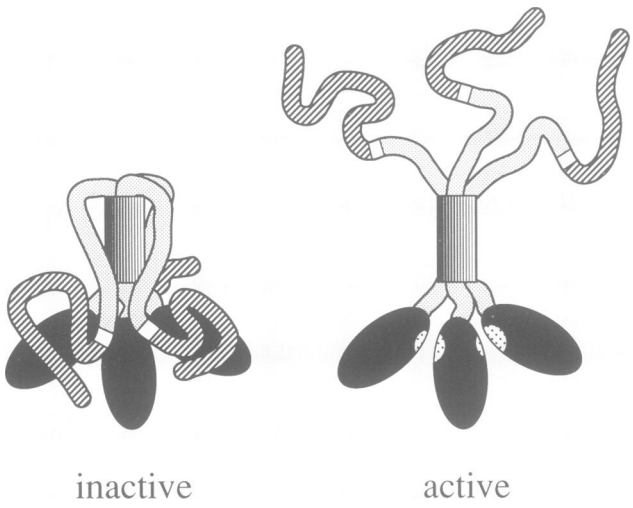

Fig. 6. Schematic model for HSF activation. The figure depicts an HSF trimer, with the DNA binding domains in black and the coiled coil trimerization domain represented as a cylinder. In the inactive state, the C-terminal activator (stripes) is tethered to the structural core of the protein by the conserved heptapeptide in CE2 (white). The binding site for this motif is shown arbitrarily on the DNA binding domain, but could equally be on the coiled coil structure or, as discussed in the text, on a different protein altogether. Upon activation, CE2 is released from its binding site, freeing the C-terminal activator. The portion of the protein on the N-terminal side of the DNA binding domain (much of which can be deleted from KlHSF without loss of regulation) is omitted for clarity.

We have shown that phosphatase treatment of ScHSF from heat-shocked cells increases its gel mobility, implying that phosphate groups are at least partly responsible for maintaining the altered conformation (Sorger *et al.*, 1987). It was thus striking to find a conserved run of serine residues adjacent to the RXLLLKNR sequence. However, changing the serines to alanines or aspartic acid residues did not affect the regulation of KlHSF activity *in vivo*, showing unequivocally that phosphorylation of these serine residues is not an essential part of the induction process. This does not exclude the possibility that phosphorylation at some other site induces the conformational change, or that more subtle regulation of HSF activity is achieved by phosphorylation.

One possible explanation is that phosphorylation does not cause the initial conformational change in HSF; rather, high temperature could directly affect the structure of the protein, resulting in the exposure of serine residues which, as a consequence of their accessibility, become phosphorylated. Phosphorylation might cause further conformational changes, or be required in some other way for transcriptional activation. Recently, it has been shown that the glutamine-rich activation region of the Sp1 transcription factor is multiply phosphorylated, and it has been suggested that this contributes to the activity of the factor, perhaps by inducing a conformational change (Jackson *et al.*, 1990).

Studies of HSF from other species indicate that structural changes can indeed be observed in the absence of phosphorylation. Treatment of extracts of human or *Drosophila* cells at high temperature, or with a variety of denaturing agents, is sufficient to convert HSF from a form that cannot bind DNA to one that can, a transition that does not depend on phosphorylation (Larson *et al.*, 1988; Mosser *et al.*, 1990; Zimarino *et al.*, 1990). Although heat shock does not change the affinity of yeast HSF for DNA, it may be that an ability to respond to denaturing conditions is a universal feature of HSFs from all species.

An alternative model is based on the observation that heat shock proteins, particularly hsp70, autoregulate the heat shock response (Morimoto *et al.*, 1990). Hsp70 is thought to bind to denatured proteins and aid their renaturation (Pelham, 1986), and it is possible that it also interacts with HSF and promotes its inactivation during recovery from heat shock (Bienz and Pelham, 1987). More specifically, hsp70 could bind to and repress HSF at normal temperatures; heat shock would produce other denatured substrates which would compete for hsp70, thus freeing HSF and allowing its phosphorylation. Once the ratio of substrates to hsp70 had been restored to normal, either by production of sufficient hsp70 or by removal or refolding of the substrates, HSF would again be inhibited.

If hsp70 does directly repress HSF, the conserved regulatory element that we have identified might be part of its binding site. The fact that the trimerization domain is required for regulation implies that the formation of a repressed complex requires HSF to be trimeric; thus, this model makes specific structural predictions about the repressed state of HSF. More detailed biochemical analysis of the inactive factor will be required to define its composition and structure, and to identify the precise events that lead to its activation.

Materials and methods

Cloning of the K.lactis HSF gene

K.lactis was kindly provided by John Salmeron (Duke University). Southern blots of *K.lactis* DNA were probed at low stringency (48°C, 2 × SSC) with a fragment corresponding to amino acids 147–259 of ScHSF. A single cross-hybridizing 1.7 kb *Bam*HI–*Pst*1 fragment was identified, and gel-purified fragments of this size were cloned into pUC19. *Escherichia coli* transformants were replica-plated, divided into pools, and screened by low-stringency Southern blotting (this was necessary to avoid the signal arising from weak cross-hybridization with the vector). After two rounds of screening, single colonies were picked and tested to identify the correct clone. This clone was then used to isolate a 3.6 kb *Kpn*I–*Pst*I genomic fragment carrying the entire gene, which was sequenced. Alignment of the KlHSF and ScHSF sequences and structure predictions were performed using the SIP and PIP programs (Staden, 1990).

Expression of K.lactis HSF mutants

Restriction sites, or deletion and point mutations were introduced using synthetic oligonucleotides as described by Kunkel *et al.* (1987).

Although early experiments established that the *K.lactis* HSF promoter was functional in *S.cerevisiae*, all mutants were expressed from the ScHSF promoter, to ensure consistent and optimal levels of protein. They were cloned into a centromere-containing vector carrying the *HIS3* gene as a selectable marker (PRS313; Sikorski and Hieter, 1989), and transformed into a haploid derivative of strain W303 (*ade2 trp1 can1-100 his3 ura3*) which carried the *HSFΔ2::LEU2* chromosomal disruption and the wild-type ScHSF gene on a *URA3*-containing vector (Sorger and Pelham, 1988). The ScHSF gene was then removed by counterselection with 5-fluoro-orotic acid (Boeke *et al.*, 1984), and the resultant *ura3* strain re-transformed with an appropriate indicator gene on a 2 μm *URA3*-containing vector. LexA fusion constructs were expressed from a centromere vector carrying *LEU2* (derived from YCp88-lexA-gcn4Δ19, Hope *et al.*, 1988) in strain Y700 (*MAT* **a** *ade2 trp1 can1-100 leu2 his3 ura3*). Indicator plasmids contained synthetic binding sites for HSF (pHSE-βG; Sorger, 1990) or lexA (YEp21-SC3423, Hope *et al.*, 1988) upstream of the CYC1 TATA box, fused to the *lacZ* gene.

For the experiment shown in Table II, indicator plasmids with overlapping HSF and GAL4 sites were used. These correspond to constructs 1, 2 and 4 described by Jakobsen and Pelham (1988).

The lexA fusion construct that includes the KlHSF trimerization domain (construct 5, Figure 3) was assayed in a strain in which the *LEU2* gene used to disrupt the chromosomal HSF gene had in turn been disrupted with the *TRP1* gene; HSF activity was provided by construct 10 (Figure 4) on

a *HIS3*-containing vector, and the strain also contained an indicator plasmid (*URA3*) and the fusion protein expression plasmid (*LEU2*).

Assay of heat-inducible activity

Strains were grown in medium lacking uracil to select for the indicator plasmids; in the case of the lexA fusions leucine was also omitted, to select for the expression plasmid. Since HSF is essential for viability, it was not necessary to select for a plasmid containing the only *HSF* gene in the strain. Except where otherwise indicated, cells were grown in media containing 2% glucose at 30°C to an OD_{660} of 1–2, heat shocked for 45 min at 40°C, and allowed to recover for 1 h at room temperature before harvesting. For each construct, 4–6 individual yeast colonies from the indicator plasmid transformation were assayed; numbers given in the figures and tables are the average of at least four determinations, and correspond to the units defined by Breeden and Nasmyth (1987).

Acknowledgements

We would especially like to thank Peter Sorger for providing plasmids, strains and unpublished information, and for many helpful discussions, and Ellen Gottlieb and Andrew Travers for comments on the manuscript.

References

Bienz,M. and Pelham,H.R.B. (1987) *Adv. Genet.*, **24**, 31–72.
Boeke,J.D., Lacroute,F. and Fink,G.R. (1984) *Mol. Gen. Genet.*, **197**, 345–346.
Breeden,L. and Nasmyth,K. (1987) *Cell*, **48**, 389–397.
Craig,E.A. (1986) *CRC Crit. Rev. Biochem.*, **18**, 239–280.
Craig,E.A. (1990) In Morimoto,R.I., Tissières,A. and Georgopoulos,C. (eds), *Stress Proteins in Biology and Medicine*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 301–321.
Hope,I.A., Mahadevan,S. and Strühl,K. (1988) *Nature*, **333**, 635–640.
Jackson,S.P., MacDonald,J.J., Lees-Miller,S. and Tjian,R. (1990) *Cell*, **63**, 155–165.
Jakobsen,B.K. and Pelham,H.R.B. (1988) *Mol. Cell. Biol.*, **8**, 5040–5042.
Kingston,R.E., Schuetz,T.J. and Larin,Z. (1987) *Mol. Cell. Biol.*, **7**, 1530–1534.
Kunkel,T.A., Roberts,J.D. and Zakour,R.A. (1987) *Methods Enzymol.*, **154**, 367–382.
Larson,J.S., Schuetz,T.J. and Kingston,R.E. (1988) *Nature*, **335**, 372–375.
Lindquist,S. (1986) *Annu. Rev. Biochem.*, **55**, 1151–1191.
Morimoto,T., Tissières,A. and Georgopoulos,C. (eds) (1990) *Stress Proteins in Biology and Medicine*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Mosser,D.D., Kotzbauer,P.T., Sarge,K.D. and Morimoto,R. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 3748–3752.
Nieto-Sotelo,J., Wiederrecht,G., Okuda,A. and Parker,C.S. (1990) *Cell*, **62**, 807–817.
Pelham,H.R.B. (1986) *Cell*, **46**, 959–961.
Sikorski,R.S. and Hieter,P. (1989) *Genetics*, **122**, 19–27.
Sorger,P.K. and Nelson,H.C.M. (1990) *Cell*, **49**, 807–813.
Sorger,P.K. and Pelham,H.R.B. (1988) *Cell*, **54**, 855–864.
Sorger,P.K., Lewis,M.J. and Pelham,H.R.B. (1987) *Nature*, **329**, 81–84.
Staden,R. (1990) *Comput. Applic. Biosci.*, **6**, 387–393.
Stone,D.E. and Craig,E.A. (1990) *Mol. Cell. Biol.*, **10**, 1622–1632.
Wiederrecht,G., Seto,D. and Parker,C.S. (1988) *Cell*, **54**, 841–853.
Yost,H.J., Petersen,R.B. and Lindquist,S. (1990) In Morimoto,R.I., Tissières,A. and Georgopoulos,C. (eds), *Stress Proteins in Biology and Medicine*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 379–409.
Zimarino,V. and Wu,C. (1987) *Nature*, **327**, 727–730.
Zimarino,V., Wilson,S. and Wu,C. (1990) *Science*, **249**, 546–549.

Received on October 29, 1990