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

I.Török* and F.Karch[+]

Department of Molecular Biology, University of Geneva, 30, quai Ernest-Ansermet, 1211 Geneva 4, Switzerland

ABSTRACT

We present the sequences at the 5' and 3' ends of one hsp 70 gene variant which is derived from the chromosomal locus 87A7. The 5' end of the hsp 70 mRNA has also been determined. 550 bp upstream from the 5' end of the hsp 70 mRNA, there is a very A+T rich region shown by heteroduplex analysis to be also present at the same position in other hsp 70 genes[9]. The 5' end of the hsp 70 mRNA was found 26 bp after a characteristic "Hogness box". The first ATG codon was found 250 bp downstream from the 5' end of the hsp 70 mRNA. We also determined the termination codon at the 3' end of the hsp 70 gene. Comparisons with other genes are discussed.

INTRODUCTION

Heat shock in Drosophila melanogaster induces the activity of about nine specific genes, while most of the other genes expressed at the normal temperature are repressed[1-3].

Genes coding for the major heat induced polypeptide, the 70,000 dalton heat shock protein (hsp 70), have been isolated and their organization studied[4-10]. There are 5 to 9 copies of the hsp 70 gene per haploïd genome[11]. The basic conserved unit for these genes consists of a segment Z_c (Z coding) (2.2 Kb) which is complementary to the hsp 70 mRNA and preceded by a segment Z_{nc} (Z non coding) (0.3 Kb) not represented in the mRNA. The detailed analysis of two hybrid plasmids 56H8 and 132E3 containing non overlapping fragments of the Drosophila melanogaster genome and carrying respectively one and two copies repeated in tandem of the hsp 70 gene, have revealed an additional

region of homology X (0.2 Kb) upstream from Z_{nc} and separated from it by a divergent region Y (0.2 Kb). Although the hsp 70 genes have similar structures in 56H8 and 132E3, the restriction maps reveal some characteristic differences in their organization and their adjacent sequences, probably reflecting their origin on the chromosome. In fact 56H8 is derived from 87A and 132E3 from 87C on the right arm of the third polytene chromosome[12]. As it is likely that these genes are coordinately induced, it is important to determine the nucleotide sequence at their flanking regions in the hope to detect some elements possibly involved in their regulation. Here we report nucleotide sequences at the 5' and 3' ends of the hsp 70 gene in the hybrid plasmid 56H8. These sequences include the X, Y, Z_{nc} elements and the beginning of Z_c. The determination of the 5' end of the hsp 70 mRNA is also reported. These sequences are compared with other eucaryotic genes already sequenced.

MATERIALS AND METHODS

Enzymes and DNA. Plasmid DNA from 56H8 was prepared as described previously[5,9] and further purified by sucrose gradient centrifugation. Restriction enzymes were obtained from New England Biolabs, Inc. (Beverly, Mass.). For all the reactions, we used the buffers recommended by the manufacturer. Bacterial alkaline phosphatase was supplied by Böhringer Mannheim. The enzyme was resuspended in 50 mM Tris-HCl (pH 8) and loaded on a small G75 Sephadex column. After collection of the appropriate fractions, glycerol was added to a final concentration of 50 % with 50 mM Tris-HCl (pH 8), 0.1 mM $ZnCl_2$ and the enzyme was stored at $-20^{\circ}C$. T_4 polynucleotide kinase was purchased from Miles, AMV reverse transcriptase from Böhringer Mannheim, S1 from Seikagu Kogyo Co LTD. Deoxy- and dideoxynucleotide triphosphates were purchased from P.L. Biochemicals, $\alpha-^{32}P$-dATP and $\gamma-^{32}P$-ATP from New England Nuclear. Autoradiography was by contact with X ray no screen films (NS 5T, Kodak).

Preparation of DNA fragments for DNA sequencing. 30 µg of 56H8 DNA were digested with the appropriate enzyme, dephosphorylated in the restriction buffer and ^{32}P labelled at the 5' ends[13].The fragments labelled at both ends were either cleaved by another restriction endonuclease or denatured and their strands separated[13]. After separation on a 7.5 % polyacrylamide gel[14] , the desired fragments were eluted into a dialysis bag by electrophoresis and further purified on a small G50 sephadex column. In some cases we first prepared unlabelled fragments in large scale. 100 to 500 µg of 56H8 DNA was cleaved, the fragments separated on a 1.4% agarose gel and stained with ethidium bromide. Appropriate gel slices containing specific fragments were pressed through a plastic syringe and incubated overnight in 0.1 M NaCl, 10 mM Tris-HCl (pH 7.4) at 4°C in the dark. The mixture was then centrifugated for 30 min at 50,000 rpm in a SW60 rotor and the supernatant was purified and concentrated on a DEAE 52 cellulose column[15].

DNA sequencing was performed according to the method of Maxam and Gilbert[13]. A>G, A>C, G>A, C+T and C reactions were done. At the end of the procedure the samples were resuspended without NaOH in the sample buffer and boiled 1 min at 100°C before loading on the gel. We used 20 % sequencing gels[13] and thin 8% sequencing gels[16]

RNA preparation. Poly(A)$^{+}$ polysomal heat shock mRNA was isolated and purified as described elsewhere[17].

RNA sequencing. The double strand DNA primers (about 0.1 µg) were denatured at 100°C for 3 min in water and hybridized to the heat shock(hs) mRNA (about 2 µg) in 10 mM Tris-HCl (pH7.4), 0.4 M NaCl, 0.5 % SDS, 57 % formamide for 16-20 hours at 42°C[19]. After two ethanol precipitations, the DNA-RNA hybrid was dissolved in 16 µl 50 mM Tris-HCl (pH 8.3), 60 mM KCl, 20 mM DTT, 10 mM MgCl$_2$, 50 µM dCTP, dGTP, dTTP together with 10 µCi of α-^{32}P-dATP (300 Ci/mmol). Four reaction mixtures of 5 µl were then prepared respectively either with 50 µM ddATP, 50 µM ddGTP, 50 µM ddCTP, or 50 µM ddTTP, and were incubated with 1-2 units of AMV reverse trans-

criptase for 10 min at 42°C. 1 µl of chase mix (2.5 mM of the
four deoxynucleotide triphosphates) was then added and incubation
was continued for another 10 min at 42°C[20-22]. The reaction was
stopped by the addition of an equal volume of formamide containing
sample buffer and the samples were loaded on an 8 % thin sequen-
cing gel[16].

S1 nuclease digestion. The DNA-RNA hybridization was done as
mentioned above. The DNA-RNA mixture was digested with 1 unit of
S1 nuclease in 0.03 M Na-Acetate (pH 4.5), 5 mM $ZnSO_4$, 0.25 M NaCl
for 1 hour at 37°C. EDTA was added at a final concentration of
10 mM to stop the reaction.

RESULTS
──────

Sequence analysis of the 5' and 3' ends of the hsp 70 gene in
56H8
    We present here the DNA sequence of the beginning and of
the end of the D. melanogaster hsp 70 gene in the hybrid plasmid
56H8. The restriction map of the Drosophila fragment in this
plasmid and the scheme of sequencing strategy are shown in Fig. 1.
A sequence of about 1 400 bp starts 700 bp upstream from the
coding region ,   includes the entire X, Y and Z_{nc} elements and
spans 700 bp in the Z_c element (Fig. 2a). At the 3' end of the
gene, the sequenced region starts 600 bp upstream from the end of
the coding region and extends 130 bp further downstream (Fig. 2b).
In most cases, 56H8 DNA was digested with SalI, XbaI and XhoI
enzymes, dephosphorylated and labelled at the 5' ends according
to Maxam and Gilbert[13]. The fragments were then cleaved with
other enzymes and separated on 7.5 % polyacrylamide gels in order
to obtain double-stranded fragments labelled at one 5' end only.
In other instances the XbaI fragment was first isolated on a
preparative scale and then digested again by other restriction
enzymes. The resulting fragments, usually 100 to 500 bp long,
were labelled at the 5' ends and the strands separated and
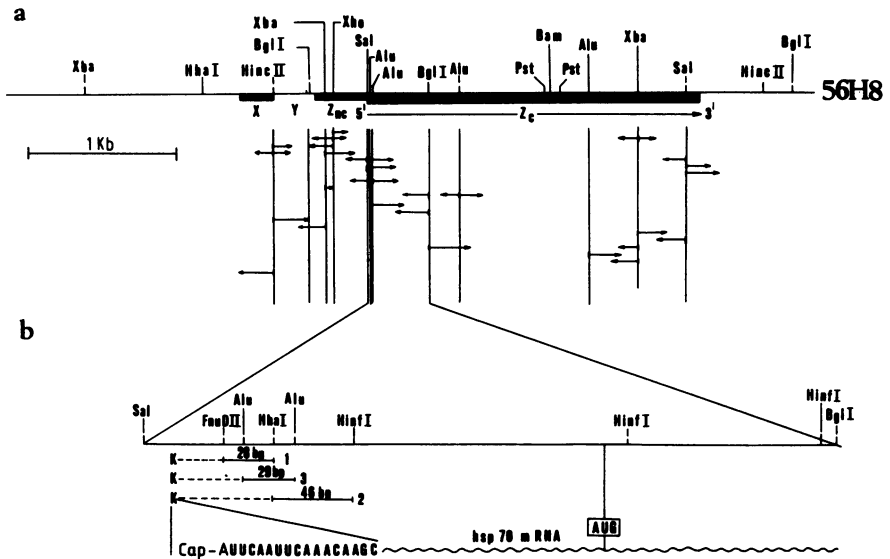isolated by gel electrophoresis. The sequencing was done with

Figure 1. Restriction map from 56H8 and sequencing strateggy

a) The elements X, Y, Z_{nc} and the region complementary to the hsp 70 mRNA (Z_c) are shown as previously[6]. The arrow under Z_c indicate the direction of transcription. The short arrows indicate the restriction site origin, the direction and the extend of the DNA sequence determination. b) Enlargement of the SalI-BglI fragment used for the preparation of primers. The primers 1, 2 and 3 have been hybridized to the hs mRNA and elongated by reverse transcription in the presence of dideoxynucleotides. The dashed lines represent the elongated chains.

overlapping fragments (see Fig. 1) from different restriction sites in both directions according to Maxam and Gilbert[13]. We used both 20 % and 8 % thin gels (0.4 mm)[13,16]. It is possible to determine with 99 % accuracy the first hundred nucleotides on thin sequencing gels. However it is necessary to sequence in both directions between one and two hundred bases, in order to obtain the same precision in this range. In those regions where the determination of the correct reading frame was in question, sequences were checked as mentioned above starting from many restriction sites. In this way it seems unlikely that we could miss or add one base and the most probable mistake might have been an incorrect C instead of T assignment.

```
                     -700                                                -650
   TGTCTCAAAAGTGTGCAAACTTGGAATTTGTTGTTTATTTTTTTGTATATATAATATTTTTAATTTTTTT

                                     -600
   TATTTATAAGTAAAATAAATATTGTTTAATTATATTGAATAAAAAAATTGCGTTTAATTAAGCAAAGAAA

                     -550                                 ▼HincII
   CCTTCATGTTTACCTTTAAAGCCAAAATTCAACTTATTTCACAGTGTAAACAGTNNACAACAACAGTCTT

     -500                                               -450
   GACAACCTGTACGTAATTTTAAATTAAAAACACTAACAATCATCTGCATGCAATTGTCTGTATTAATCTA

                             -400
   ATAAATAAATAGCTTTTTTAAGTTAGTATGTAAATACATTTTGAAGAATATCTTGTCAAAGTTCCATAGG

                 -350                                           -300
   CCTTTCTGGCGGACAACATCCGCTAACAAACCCTTCGATTATCTCTAACATAATTAACTTAAGCAGCCGT

     ▬▬▬▬▬▬▬                               ▬▬▬▬▬▬▬      -250  ▼ Xba
   ATTTATAAAGAAATTTCCAAAATAAAGTATATATATAAATAAAGAATATTCTAGAACCCCCAAAACAAAC
                                                             ▲
                         -200                 ▼ Xho
   CTGGTTGTTGCGCGAGGCATTTGTTTGCCAGAAAGAAAACTCGAGAAATTTCTCTGGCCGTTATTCTCTA
                                             ▲
     -150                                           -100
   TTCGTTTTGTGACTCTCCCTCTCTGTACTATTGCTCTCTCACTCTGTCGCACAGTAAACGGCACGCTATT

                             -50
   CTCGTTGCTTCGAGAGACGCGGCCTGCGAATGTTCGCGAAAAGAGCGCCGGAGTATAAATAGAGGCGCT
                                                           ▬▬▬▬▬▬▬▬
     ▼ Sal        -11                                              50
   TCGTCGACGGAGCGTCAATTCAATTCAAACAAGCAAAGTGAACACATCGCGAAGCGTAAGCTGAGCAACA
     ▲
                                         100
   AACAAGCGCAGCGAACAAGCTAAACAATCTGCAATAAAGTGCAAGTTAAAGTGAATCAATTAAAAGTAAC

                     150
   CAACAACCAAGTAATTAAACTAAAAACTGCAACTACTGAAATCAACCAAGAAGTCATTATTGAAGACAAG

     200                                           250
   AAGAGAACTCTGAATACTTTCAACAAGTCGTTACCGAGGAAGAACTCACACACAATGCCTGCTATTGGAA
                                                         ▬▬

                               300
   TCGATCTGGGCACCACCTACTCCTGCGTGGGTGTCTACCAACATGGCAAGGTGGAGATTATCGCCAACGA

         350                                 BglI ▼        400
   CCAGGCAACCGCACCACGCCGTCCTACGTGGCTTTCACAGATTCGGAACGCCNNNNNGGCGACCAACAAC
                                                  ▲

                                       450
   AACCGCTGGCCATTGAACCCCAGCAACACAGCGTTTCACGCCAAGCGACTCATCGGCCGAAAATACGACG

                     500
   ACCCCAAGATCGCAGAGGACATGAAGCACTGGCCTTTCAAGGCTGGTAAGCGACGGCGGAAAGCCCAGAT

     550                                           600
   CGGGGTGGAGTATAAGGGTGAGTCCAAGAGATTTGCCCCCGAGGAGATCAGCTCGATGGTGCTGACCAAG

                             650
   ATGAAGGAGACGGCGGCGGAGGCATATCTGGGCGAGAGCCCACAGACGCAGTCATCACAGTTCCAGCCTA

   CTTCAACGACT
```

Figure 2. a) <u>Sequences around the beginning of hsp 70 gene in</u>
       <u>56H8</u>

The nucleotide sequence derives from the strand corresponding to the mRNA in the 5' to 3' direction. Some of the relevant restriction sites are indicated along the sequence. The hsp 70 mRNA sequence starts at position 1±2. The TATAAATA motifs and the first ATG codon are shown by thick lines.

```
                                                          50
CGTGCGATGACGAAGGACAACAATGCATTGGGCACCTTCGATCTGTCCGGCATTCCACCTGCACCAAGGG
                         100
GTGTGCCCCAGATAGAAGTTACCTTCGACTTGGACGCCAATGGAATCCTGAACGTCAGCGCCAAGGAGAT
      150                                                 200
GAGCACGGGCAAGGCCAAGAACATCACGATCAAGAACGACAAGGGACGCTCTCGCAGGCCGAGATTGATC
                                  250                     ▼ Xba
GCATGGTGAACGAGGCTGAAAAGTACGCCGACGAGGACGAGAAGCGTCGCCAGCGAGTAACCTCTAGAAA
                                                                   ▲
      300                                                      350
TGCCCTGGAGAGCCACGTCCTCAATGTGAAGCAGGCCGTGGAACAGGCACCTGCTGGCAAACTGGACGAG
                                  400
GCTGACAAGAACTCCGACTTGGACAAGTGCAACGACACTATCCGGTGGCTGGACAGCAACACCACTGCCG
                   450
AGAAGGAGGAGTTCGACCACAAGCTGGAGGAGCTCACCCGCCACTGCTCCCCCATCATGACCAAGATGCA
      500                                                 550
TCAGCAGGGTGCGGGAGCTGGAGCTGGTGGTCCGGGAGCAAACTGCGGCCAGCAGGCGGGAGGATTTGGA
                       ▼ Sal        600
GGCTACTCTGGACGCACGGTCGAGGAGGTCGACTAAGGCCAAAGAGTCTAATTTTTGTTCATCAATGGGT
                                 ▲
      650                                                      700
TATAACATATGGGTTATATTATAAGTTTGTTTTAAGTTTTTGAGACTGATAAGAATGTTTCGATCGAATA
                                  750
TTCCATAGAACAACAATAGTATTACCTAATTACCAAGTCTTAATTTAGCAAAAATGTTATTGCTTATAGA
                       800
AAAAATAAATTATTTATTTGAAATTTAAAGTCAACTTGCA
```

b) Sequences at the end of the hsp 70 gene in 56H8

The sequence is from the same DNA strand as in a). Here, the nummerotation is started arbitrarily next to the AluI cut shown in Fig. la. The translation termination triplet and the AATAAA motif are shown by thick lines as well as the TGC sequence suspected to be the poly.(A) addition site.

Determination of the 5' end of the hsp 70 mRNA

The region complementary to the hsp 70 mRNA was estimated to lie between the two SalI sites in 56H8[6]. In order to find out the position of its 5' end, we used the S1 mapping procedure developed by Berk and Sharp[18]. Starting from the 400 bp SalI-BglI fragment purified on a large scale by electrophoresis on 5 % polyacrylamide gel, we have isolated therefrom the fragment SalI-FnuDII (see Fig. 1b and Fig. 4, band 4). After labelling of both 5' ends, the strands were separated on an 8 % denaturating polyacrylamide gel, taking advantage of the fact that SalI generates staggered ends and FnuDII flush ends so that one strand is four

nucleotides longer than the other (Fig. 3, slot a). Each single strand was separately added to the hsp 70 mRNA under hybridization conditions and subsequently digested by Sl. While the non complementary strand is completely digested, Sl treatment reduced the size of the complementary strand by 8 to 12 nucleotides, as seen on an 8 % 7 M urea sequencing gel (Fig. 3). This indicates that the 5' end of the hsp 70 mRNA is 16 $\pm$ 2 bp downstream from the first SalI site in 56H8.

The Sl procedure does not distinguish between the real extremity of the messenger and the break-point of a possible intron which would put the 5' end of the mRNA in a remote upstream position. Thus we attempted to obtain the sequence of the 5' end of the hsp 70 mRNA indirectly by reverse transcription in the presence of chain terminators, using specific primers located at a short distance downstream from the first SalI site. The 400 bp SalI-BglI fragment mentioned above was cleaved with HhaI or AluI. In order to detect the resulting small frag-
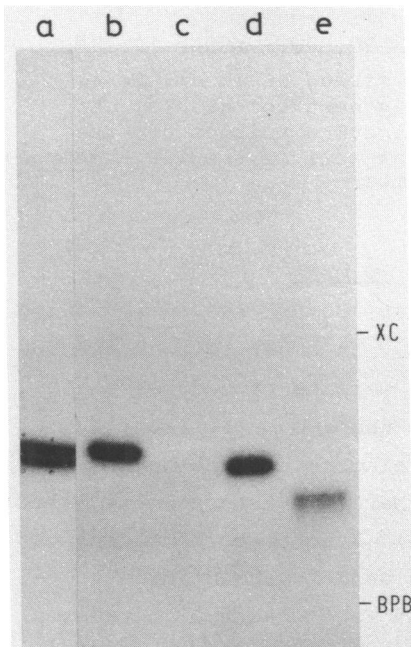

Figure 3. Autoradiogram of Sl resistant DNA fragments after hybridization with hs mRNA

Slot a : separation of the strands from the fragment Sal-FnuDII (band 4 Fig. 4, Fig. 1b). As one strand is four bases longer than the other, the separation is done on a 8 % sequencing gel. Slots b and d : each strand was incubated with hs mRNA for hybridization. Slots c and e : the hybridization mixtures were digested by Sl nuclease. The non complementary strand disappeared (c) while the complementary strand was shortened by about 10 nucleotides (e). We used as molecular weight standard the known size of each strand as well as the position of the xylen cyanol (XC) equivalent to 85 bases and the bromophenol blue (BPB) equivalent to 22 bases in these conditions.

ments, both digests were labelled at their 5' ends. Fig. 4
shows the separation of these labelled digests after cleavage
by FundII and HinfI and the fragments used as primers. After
dephosphorylation to remove the 5' end label, the double stran-
ded primers were hybridized to the poly(A)$^+$ fraction of the
polysomal heat shock mRNA under conditions which favor the
formation of DNA-RNA hybrids[19] . In some cases the 20S fraction
of the poly(A)$^+$ polysomal mRNA was isolated on sucrose gradient,
but this procedure did not show any advantage over the use of
total poly(A)$^+$ polysomal RNA. The DNA sequencing method descri-
bed by Sanger et al.[42] , and recently adapted to RNA sequencing
was used[21,22] .

Fig. 5 shows the separation of the elongated primers on 8 %
thin sequencing gels. A background of unspecific bands appears in
all the slots and with either primers and these ambiguities will
be discussed later. Nevertheless a sequence of 8 nucleotides
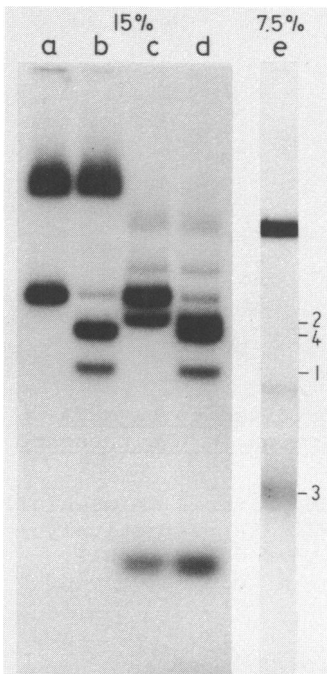5'TGTTCACT3' can be read without ambiguities (Fig. 5a). The com-

Figure 4. **Isolation of DNA primers**

Autoradiograph of a 15 %
polyacrylamide gel. Slot a : HhaI
digest of the SalI-BglI fragment
(Fig. 1b). The digest was ^{32}P-la-
belled at the 5' ends. In slots b,
c and d, the above digest was
further cleaved respectively by
FnudII, HinfI and FnudII plus
HinfI. Slot e : AluI digest of the
SalI-BglI fragment fractionated on
a 7.5 % polyacrylamide gel. The
digest was ^{32}P-labelled at the 5'
ends. The fragments 1 (FundII-HhaI),
2 (HhaI-HinfI), 3 (AluI-AluI) and
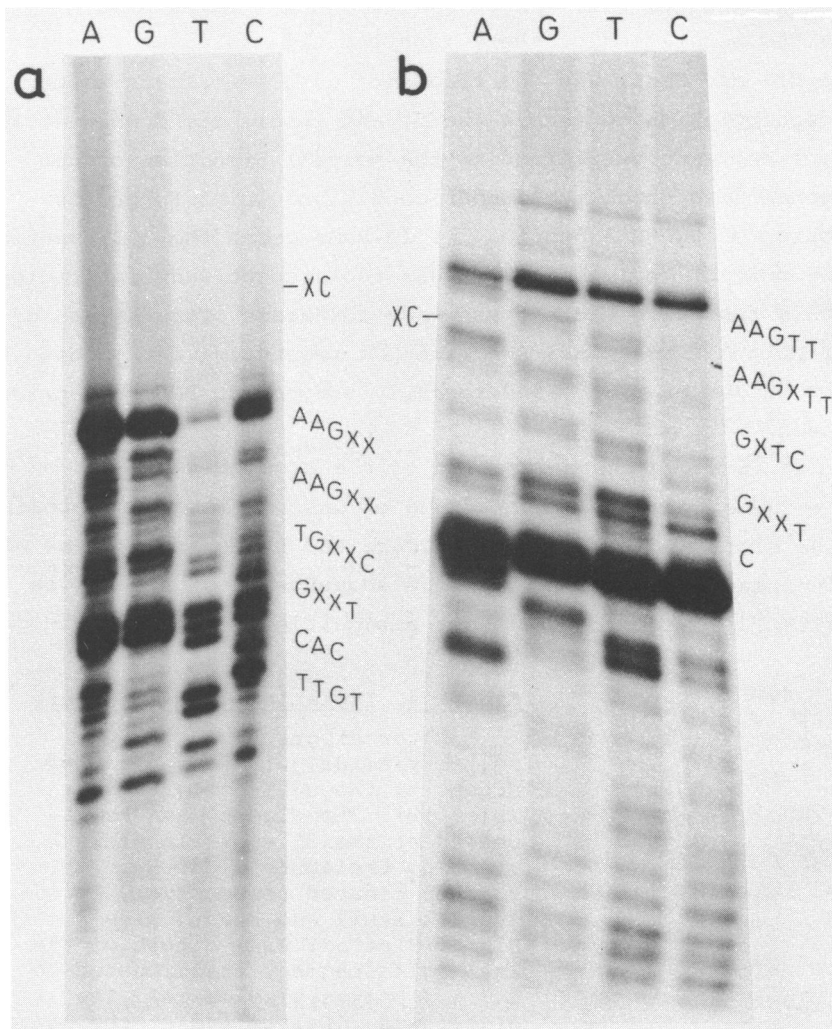4 SalI-FundII) were used as primers.

Figure 5. Autoradiograph of DNA primers hybridized to hs mRNA and elongated by reverse transcription in the presence of dideoxynucleotides

a) with primer 1 (FundII-HhaI), b) with primer 3 (AluI-AluI) (see Fig. 1b and 3). Slots A, G, T and C indicate respectively the reaction mixtures in the presence of either ddATP, ddGTP, ddTTP, or ddCTP. Electrophoresis was in an 8 % thin sequencing gel.

plementary form of this octanucleotide is found only once from
position 20 to 27 in the 1 400 nucleotides determined on both
sides of the first SalI site (Fig. 2a). In Fig. 5a the sequence
is terminated by 5'AAXXGAA3' which could be interpreted as being
5'AAAAGAA3'. This latter heptanucleotide is absent in all the DNA
sequence examined. However in an other experiment the same end is
clearly read 5'AATTGAA3' (Fig. 4b) and the complementary form of
this sequence is found only from position 2 to 8 on the DNA
sequence (Fig. 2a). This is in complete agreement with the
position of the octanucleotide mentioned above. The termination
of the polymerization seems to be the same with both primers and
thus appears to be specific. Moreover the size difference between
the two full reverse transcripts (8 to 11 nucleotides) is in good
agreement with the position of the two primers. As most of the
heat shock induced messengers have a cap structure[23], we shall
assume that the hsp 70 mRNA is capped. In the case of Rous sarco-
ma RNA, Haseltine et al.[43] have shown that the reverse transcrip-
tase stops precisely at the cap site. A strong band appears in
all the slots above the last A in Fig. 5 and we think it likely
that it represents the beginning of the messenger sequence (posi-
tion 1 in Fig. 2a). The unspecific bands of the background
mentioned above could be due to hybridization of a small fraction
of the primers with another region of the template mRNA or with a
possible contaminant. This could explain the presence of bands
above the last position of the messenger sequences in Fig. 5b. On
the other hand, we know that the hsp 70 mRNA is transcribed at
both 87A and 87C loci[6,8] and that there are some slight structural
differences in the genes derived from these two loci (see intro-
duction). In fact, there is roughly 10 % divergence between z_c
sequences from 56H8 and 132E3 (Karch and Török, in preparation).
We believe that the high background of unspecific bands may
reflect the heterogeneity of the hsp 70 mRNA. The reverse trans-
cription sequencing data are in good agreement with the result
obtained by the S1 procedure. However, a possible splicing point
within the first 20 nucleotides of the mRNA cannot be excluded
on the basis of the above data. We tentatively conclude that the

5' end of the hsp 70 mRNA sequence is located at 16 $\pm$ 2 nucleotides downstream from the first SalI site in 56H8.

DISCUSSION

We have sequenced X, Y, Z_{nc} and 650 bp within Z_c at the 5'
end of the hsp 70 gene in 56H8 as well as 750 bp at the 3' end of
the gene (Fig. 2). By S1 mapping we have determined the 5' end of
the hsp 70 mRNA 16 $\pm$ 2 bp downstream from the first SalI site.
The sequencing data obtained by reverse transcription identifies
a characteristic octanucleotide which is found only once on the
DNA sequence at a position fitting exactly with the S1 determina-
tion. The sequence heterogeneity in the hsp 70 mRNA prevents the
reading without ambiguities of the last nucleotides at the 5'
extremity of the messenger. Thus we cannot exclude the possibili-
ty of a splicing within the very first nucleotides of the hsp 70
mRNA. The finding 26 bp upstream from the 5' end of the hsp 70
mRNA, of a characteristic "Hogness box" present at about the
same distance in front of a large number of other genes would be
in agreement with our interpretation (see below).

Figure 6 shows the A+T distribution through X, Y, Z_{nc} and
the beginning of the gene (Z_c). In the region corresponding to
the X element, we find a stretch very rich in A+T (92 %) from
position -700 to -580 (Fig. 2a). This very rich A+T region might
be characteristic of the X element. The attractive possibility
that the X element might play a role in the hsp 70 gene expres-
sion has been discussed and remains to be investigated[6,9,10]. Y
has an A+T content of about 70 %. In the first part of Z_{nc} the
average of the A+T distribution is roughly 70 %, while from posi-
tion -170 to the end of Z_{nc} it is 50 %. The base composition of
Z_c appears to be biphasic. An A+T rich region comprises the first
two hundred nucleotides of Z_c and correspond to the leader of the
mRNA (see below) while the second part which is G+C rich (60 %)
corresponds to the beginning of the coding region. It should be
noted that the first ATG triplet candidate for the initiation
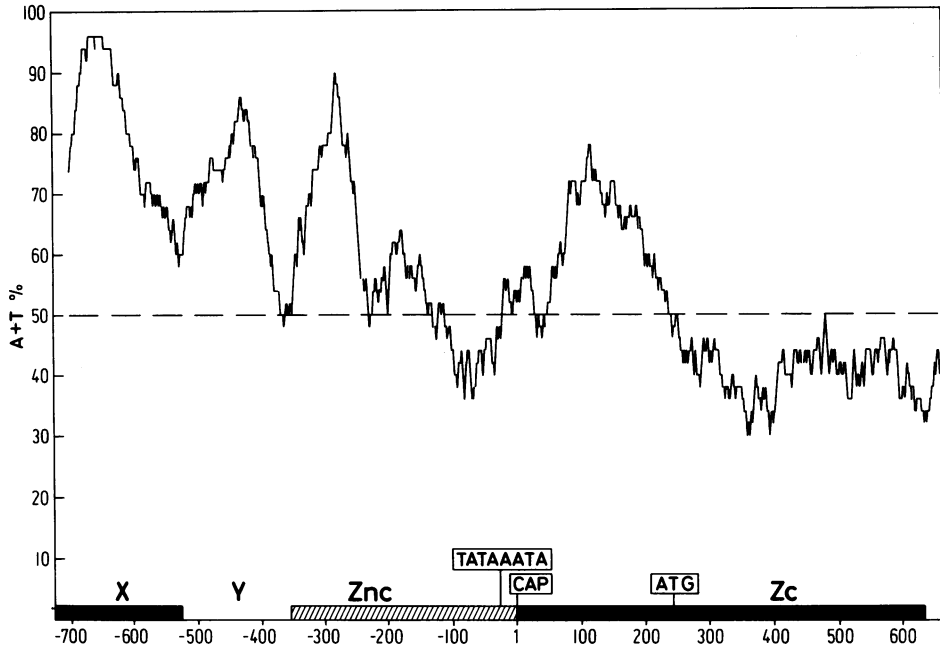codon appears at the beginning of that G+C rich region. A de-

Figure 6. <u>A+T content around the beginning of the hsp 70 gene in 56H8</u>

Each point plotted on the graph indicates the A+T percent in a fragment of 50 bases centered around this point. In this figure, the accurate limits of the sequence elements X, Y and Z_{nc} are shown as determined by sequence comparison between 56H8 and 132E3 (Karch and Török, in preparation).

tailed comparison of the sequence data in the X, Y and Z_{nc} elements of the hsp 70 genes in 56H8 and 132E3 will be reported elsewhere (Karch and Török, in preparation).

<u>5' end of the hsp 70 gene in 56H8</u>

At the 5' end of the hsp gene in 56H8, we have found three TATAAATA motifs, a sequence often present a short distance upstream from the start of eucaryotic messenger RNAs transcribed by RNA polymerase II and referred to as "Hogness box". Two of them appear close to each other in the A+T rich part of Z_{nc} (at position -287 and -257 (Fig. 6 and 2a). The third one also found

in the Z_{nc} element is very close to the SalI site at position
-26 and shares all the characteristics of the "Hogness box". It
is inserted in a G+C rich region, 26 bp upstream from the 5' end
of the hsp 70 mRNA (Fig. 6 and 2a). The "Hogness box" has been
found so far about 25 bp upstream from the cap site in D. melano-
gaster histone genes (Goldberg, personal communication), sea
urchin histone genes[24,25], mouse β-globin[maj] gene[26], mouse
β-globin[min] gene[27], mouse α-globin gene[28], rabbit β-globin gene[29],
adenovirus 2 major late genes[30], mouse light chain immunoglobu-
lin gene [31], chicken ovalbumin gene[32], chicken conalbumin gene[33],
yeast iso-1-cytochrome gene[34] and B mori fibroin gene[35]. Its
similarity to the bacterial "Pribnow box" has been noted[32] and
assuming that the cap site of the mature mRNA and the starting
point of transcription by RNA polymerase II coincide, it is
tempting to speculate that the "Hogness box" might be part of
a promoter signal for transcription. But the fact that the
TATAAATA motif is not found in front of some viral genes indica-
tes however that it is not absolutely necessary to generate the
5' ends of mRNAs. Those viral genes which lack the "Hogness box"
produce mRNAs with 5' end heterogeneity in vivo[36]. Moreover
deletion of a DNA fragment containing the "Hogness box" in front
of the H2A histone gene from sea urchin decreases the rate of
transcription of this gene when injected in frog oocytes, and
also generates mRNAs with 5' end heterogeneity[37].

Sequence homologies appear also in regions just preceding
and overlapping the cap locus. We find a small sequence TTCGTC
at position -13 very similar to a sequence of six bases present
in the same region in adenovirus 2 major late genes[30], mouse
β-globin[maj 26], mouse β-globin[min 27], mouse α-globin gene[28],
rabbit β-globin gene[29] (see Fig. 7). While this sequence is
completely similar to that found in adenovirus 2 major late
genes, its precise position relative to the cap site differs
(Fig. 7). It should be noted that the B. mori fibroin gene, an
other insect gene, does not share any homology in that region
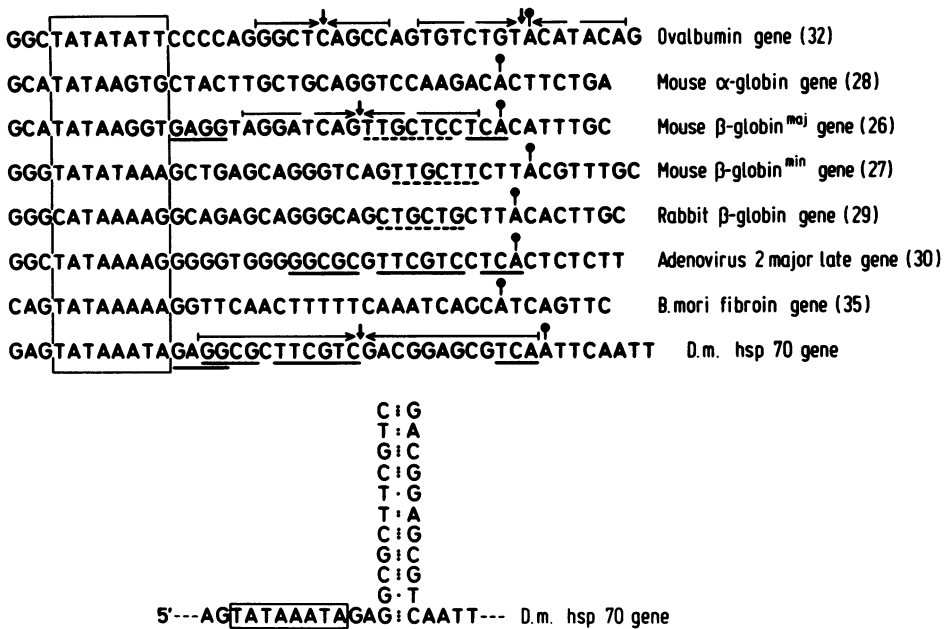with the D. melanogaster hsp 70 gene and with the mammalian genes
mentioned above[35].

```
GGCTATATATTCCCCAGGGCTCAGCCAGTGTCTGTACATACAG  Ovalbumin gene (32)

GCATATAAGTGCTACTTGCTGCAGGTCCAAGACACTTCTGA    Mouse α-globin gene (28)

GCATATAAGGTGAGGTAGGATCAGTTGCTCCTCACATTTGC    Mouse β-globin maj gene (26)

GGGTATATAAAGCTGAGCAGGGTCAGTTGCTTCTTACGTTTGC  Mouse β-globin min gene (27)

GGGCATAAAAGGCAGAGCAGGGCAGCTGCTGCTTACACTTGC   Rabbit β-globin gene (29)

GGCTATAAAAGGGGGTGGGGGCGCGTTCGTCCTCACTCTCTT   Adenovirus 2 major late gene (30)

CAGTATAAAAAGGTTCAACTTTTTCAAATCACCATCAGTTC    B.mori fibroin gene (35)

GAGTATAAATAGAGGCGCTTCGTCGACGGAGCGTCAATTCAATT  D.m. hsp 70 gene
```

```
              C : G
              T : A
              G : C
              C : G
              T · G
              T : A
              C : G
              G : C
              C : G
              G · T
5'---AGTATAAATAGAG : CAATT--- D.m. hsp 70 gene
```

**Figure 7. Comparison between DNA sequences of various genes
upstream from the 5' end of the messenger sequences**

    Each sequence is aligned with respect to the "Hogness box"
which is framed. The strong homology sequences between the hsp
70 gene and the other genes are underlined with continuous lines.
The dashed lines indicate regions of weak homology. The horizon-
tal arrows show the regions of two-fold symmetry and their center
is shown by a vertical arrow. Black spots correspond to the 5'
ends of the mRNAs. A potential hairpin loop within the sequence
upstream from the hsp 70 gene is also drawn.

    A region of two-fold symmetry is found centered between the

third "Hogness box" and the position of the 5' end of the hsp 70

mRNA. The conserved sequence TTCGTC mentioned above can be part

of a potential hairpin loop (Fig. 7). Partial two-fold symmetry

is also found in a similar location in the chicken ovalbumin

gene, rabbit β-globin gene, mouse β-globin[maj] gene and adenovirus

2 late genes[32].

    In 22 out of 23 eucaryotic messengers which have been

studied the AUG codon closest to the 5' end is that used for the

initiation of translation[38]. In the 56H8 hsp 70 gene the first

ATG codon is found 250 bp downstream from the cap site. It seems relevant that it occurs just at the beginning of the G+C rich area (see Fig. 6). We have sequenced 450 bp after this first initiation codon and have not found any termination triplet in the same reading frame, while many of them were found in the other two possible reading frames. Four other ATG triplets have been observed at position 308, 496, 601 and 616. The last three are in the same open reading frame.

So far, there is no evidence for hsp 70 mRNA precursors or intervening sequences within hsp 70 genes[8]. The RNA sequence data by the dideoxynucleotide technique confirms the absence of small intervening sequences between the first ATG codon and the next BglI site in the gene (data not shown). Further RNA sequencing data will be necessary to find out whether the RNA is indeed not spliced within the long leader sequence of 250 bases.

As proposed by Hagenbuchle et al.[39], a sequence rich in purin near the 3' end of the 18S ribosomal RNA can base pair with the 5' end non coding region of mRNA molecules, guiding ribosomal binding for the initiation of translation. The distance between that hypothetical binding and the AUG triplet can vary from 8 to 45 nucleotides in the different messengers studied. 33 bp upstream from the first ATG codon in the 56H8 hsp 70 gene we find a sequence of 8 nucleotides ATACTTTC six of which can base pair with the purin rich region at the 3' end of the 18S rRNA. One of the missmatches is a GU interaction. This postulated interaction may however not be necessary for the eucaryotic messengers since they are monocistronic. Kozak has proposed a scanning model where the ribosome binds at the 5' ends of the messengers and begins translation at the first AUG[38].

3' region of hsp 70 gene

Of the three possible reading frames in the 3' region of the hsp 70 gene, two can be eliminated by the presence of many termination codons. In the third frame, the first terminator codon is found 8 nucleotides downstream from the SalI site. After this point all three reading frames are interspersed by a large

number of termination triplets. The same reading frame is obtained at the 3' end of the second hsp 70 gene in 132E3 (Karch and Török, in preparation).

In the 3' untranslated region of the hsp 70 gene we found the AATAAA sequence which is common in 13 eucaryotic poly(A)$^+$ mRNA 14 to 30 bases upstream from the poly(A) tail[40,41]. A TGC sequence suspected to be the pola(A) addition site is found 30 bases further (Fig. 2b).

ACKNOWLEDGEMENTS

*Present address: Institute of Genetics, Biological Research Center, Hungarian Academy of Sciences, H-6701 Szeged, P.O.B. 521, Hungary.

[+]To whom reprints requests should be addressed.

REFERENCES

1. Ritossa, F.M. (1963) Drosophila Information Service 37, 122-123.
2. Tissières, A., Mitchell, H.K. and Tracy, U.M. (1974) J. Mol. Biol. 84, 389-398.
3. Ashburner, M. and Bonner, J.J. (1979) Cell 17, 241-254.
4. Livak, K.F., Freund, R., Schweber, M., Wensink, P.C. and Meselson, M. (1978) Proc. Nat. Acad. Sci. USA 75, 5613-5617.
5. Schedl, P., Artavanis-Tsakonas, S., Steward, R., Gehring, W. J., Mirault, M.-E., Goldschmidt-Clermont, M., Moran, L. and Tissières, A. (1978) Cell 14, 921-929.
6. Artavanis-Tsakonas, S., Schedl, P., Mirault, M.-E., Moran, L. and Lis, J. (1979) Cell 17, 9-18.
7. Craig, E.A., McCarthy, B.J. and Wadsworth, S.C. (1979) Cell 16, 575-588.
8. Holmgren, R., Livak, K.F., Morimoto, R., Freund, R. and Meselson, M. (1979) Cell 18, 1359-1370.

9. Moran, L., Mirault, M.-E., Tissières, A., Lis, J., Schedl, P., Artavanis-Tsakonas, S. and Gehring, W.J. (1979) Cell 17, 1-18.
10. Goldschmidt-Clermont, M. (1980) Nucl. Acids Res. 8, 235-252.
11. Mirault, M.-E., Goldschmidt-Clermont, M., Artavanis-Tsakonas, S. and Schedl, P. (1979) Proc. Nat. Acad. Sci. USA 76, 5254-5258.
12. Ish-Horowicz, D., Pinchin, S.M., Schedl, P., Artavanis-Tsakonas, S. and Mirault, M.-E. (1979) Cell 18, 1351-1358.
13. Maxam, A.M. and Gilbert, W. (1977) Proc. Nat. Acad. Sci. USA 74, 560-564.
14. Allet, B. (1973) Biochemistry 12, 3972-3977.
15. Smith, H.O. and Birnstiel, M.L. (1976) Nucl. Acids Res. 3, 2387-2389.
16. Sanger, F. and Coulson, A.R. (1978) FEBS Letters 87, 107-110.
17. Henikoff, S. and Meselson, M. (1977) Cell 12, 441-451.
18. Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732.
19. Casey, J. and Davidson, N. (1977) Nucl. Acids Res. 4, 1539-1552.
20. McGeoch, D.J. and Polan, A. (1979) Nucl. Acids Res. 6, 3199-3211.
21. Bina-Stein, M., Thorrens, M., Salzman, N. and Thompson, J.A. (1979) Proc. Nat. Acad. Sci. USA 76, 731-735.
22. Akusjärvi, G. and Petterson, U. (1978) Proc. Nat. Acad. Sci. USA 75, 5822-5826.
23. Levis, R. and Penman, S. (1978) J. Mol. Biol. 120, 487-515.
24. Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H.O. and Birnstiel, M.L. (1978) Cell 14, 655-671.
25. Busslinger, M., Portman, R., Irminger, J.C. and Birnstiel, M.L. (1980) Nucl. Acids Res. 8, 957-977.
26. Konkel, D.A., Tilghan, S.M. and Leder, P. (1978) Cell 15, 1125-1132.
27. Konkel, D.A., Maizel, J.V. and Leder, P. (1979) Cell 18, 865-873.
28. Nishioka, Y. and Leder, P. (1979) Cell 18, 875-882.
29. Van Ooyen, A., Van den Berg, J., Mantei, N. and Weissmann, C. (1979) Science 206, 337-344.
30. Ziff, E.B. and Evans, R.M. (1978) Cell 15, 1463-1475.
31. Tonegawa, S., Maxam, A.M., Tizard, R., Bernard, O. and Gilbert, W. (1978) Proc. Nat. Acad. Sci. USA 75, 1485-1489.
32. Gannon, F., O'Hare, K., Perrin, F., Le Pennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. and Chambon, P. (1979) Nature 278, 428-434.
33. Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F. and Chambon, P. (1979) Nature 282, 567-574.
34. Smith, M., Leung, D.W., Gillan, S., Astell, C.R., Montgomery, D.L. and Hall, B.D. (1979) Cell 16, 753-761.
35. Tsujimoto, Y. and Suzuki, Y. (1979) Cell 18, 591-600.
36. Baker, C.C., Herisse, J., Courtois, G., Galibert, F. and Ziff, E.B. (1979) Cell 18, 569-580.
37. Grosschedl, R. and Birnstiel, M.L. (1980) Proc. Nat. Acad. Sci. USA 77, 1432-1436.
38. Kozak, M. (1978) Cell 15, 1109-1123.

39. Hagenbüchle, O., Santer, M., Argetsinger Steitz, J. and Mans, R.J. (1978) Cell 13, 551-563.
40. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263,211-214.
41. Benoist, C., O'Hare, K.O., Breathnach, R. and Chambon, P. (1979) Nucl. Acids Res. 8, 127-142.
42. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat. Acad. Sci. USA 74, 5463-5467.
43. Haseltine, W.A., Maxam, A.M. and Gilbert, W. (1977) Proc. Nat. Acad. Sci. USA 74, 989-993.