Two genes for the major heat-shock protein of Drosophila melanogaster arranged as an inverted repeat

Michel Goldschmidt-Clermont

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

Received 4 December 1979

ABSTRACT

The physical maps of three cloned D. melanogaster DNA fragments with genes for the 70,000 dalton heat-shock protein (hsp 70) are presented. Fragment 122 contains two genes in diverging orientation, forming an inverted repeat around a central spacer. The other two fragments, which are found as polymorphic variants in the fly population, have related structures; they differ by the deletion or the insertion of large DNA segments. The sequence homologies between 122 and a plasmid with two hsp 70 genes in tandem repetition was investigated by heteroduplex analysis. In addition to the basic gene units, the segments share other homologous sequence elements which are found in different combinations near the beginning of the genes.

INTRODUCTION

Heat shock treatment of Drosophila melanogaster results in an extensive change in the pattern of transcription and translation (see review by Ashburner and Bonner). A set of nine new puffs appears in the polytene chromosomes while those previously present regress. Heat shock specific RNA and proteins are actively produced in a variety of tissues and culture cells, while the expression of the genes normally active at 25°C is repressed. This provides a convenient experimental system for the study of gene expression, where several specific genes can be induced simultaneously.

The isolation of genes for the main polypeptide product, the 70 kd heat shock protein (hsp 70) has allowed the study of their structure and organization. The basic unit (Z) consists of a segment Zc (2.2 kb) which is complementary to the...
Nucleic Acids Research

hsp 70 messenger RNA, flanked at its beginning by a sequence Znc (Z non-complementary), (0.3 kb) which is also conserved in the different units, but is not represented in the mRNA. In the detailed comparison of the genes contained in two plasmids, 56H8 and 132E3, an additional short region of homology X (0.2 kb) has been found upstream from Znc

Two sites of heat shock puffs on the chromosomes, at 87A7 and 87C1, encode very similar variants of the polypeptide, hsp 70. The hsp 70 genes are repeated 5 to 9 times per haploid genome, and exist in two variant forms which differ by some restriction sites. It was recently shown that at each of the two chromosomal loci, Z units of only one variant type are found.

Some of the genes are organized in tandem repetitions of two or three. At 87C1, the hsp 70 genes are associated with other tandemly repeated heat-shock genes of the type described by Lis et al. Variations in the number of the tandem units account for part of the polymorphism displayed by the Oregon R population.

Here I describe a cloned DNA segment (122) carrying two hsp 70 genes which, in contrast to those previously isolated, are arranged as an inverted repeat. The comparison of 122 with two other related fragments indicates that this arrangement is subject to a different type of polymorphism: the fragments apparently derive from one another by the deletion or the insertion of large DNA sequences. A heteroduplex analysis of 122 and of the previously isolated plasmid 132E3 shows the presence of homologous sequence elements in various combinations near the beginning (5' end of the mRNA) of the hsp 70 genes.

MATERIALS AND METHODS

Enzymes and DNA. Restriction enzymes and T4 ligase were obtained from New England Biolabs, Inc. (Beverly, Mass.) except EcoRI, which came from the Microbiological Research Establishment (Salisbury, U.K.). The buffers recommended by the manufacturerer were used for all the reactions. Drosophila melanogaster Oregon R and plasmid DNA were prepared and labelled as described previous-
ly, phage DNA was obtained by phenol extraction of phage purified by CsCl equilibrium centrifugation.

Construction and identification of hybrid phages and plasmids. Pools of λ Charon 9 hybrids containing Drosophila EcoRI fragments were prepared as follows. The vector DNA was ligated at the cohesive ends and digested with EcoRI. The vector arms were partially purified from the internal fragments by sedimentation in sucrose gradients for 16 hours at 35,000 rpm in a Beckman SW40 rotor (20-50% sucrose, 0.25 M NaCl, 0.01 M Tris-HCl pH 7.4, 0.001 M EDTA). Vector arms and Drosophila EcoRI fragments (at 300-400 μg/ml each) were ligated at 12°C for 24-36 hours and packaged in vitro into phage particles, yielding 10⁵-10⁶ plaque forming units per μg of input vector DNA. The phages were plated on E. coli SF8 and screened by the method of Benton and Davis. The hybridization probe was a subclone of 132E3 containing a BamHI fragment with hsp 70 gene sequences and the spacer, labeled with ³²P by nicktranslation.

Restriction fragments were transferred to plasmid pBR322 by ligation of restriction digests of phage and plasmid DNA followed by transformation of E.coli HB101. The subclones were identified by colony hybridization and restriction mapping.

These experiments were carried out in a P2 laboratory according to the NIH guidelines issued in June 1976, as adopted by the Swiss Federation of Experimental Biologists in May 1977.

Restriction maps. The maps were derived by analyzing the electrophoretic migration of fragments produced by single or combined restriction enzymes. The size of the fragments was determined by calibrating the agarose gels with an EcoRI and HindIII digest of phage λ as a standard. Because of the surrounding symmetry, the two sites in the central region of 122 (BamHI and BglII) were positioned with respect to the external segments by analyzing a partial BamHI digest. For this analysis, fragments 122 and 123 were transferred from λ Charon 9 to a smaller vector, plasmid pBR322.
Electrophoresis of DNA fragments, transfer to nitrocellulose and hybridization. DNA fragments were separated by agarose gel electrophoresis and transferred to nitrocellulose (Schleicher and Schüll BA85) with 10x SSCP as transfer solvent. (1x SSCP is 0.12 M NaCl, 0.15 M sodium citrate, 0.013 M KH₂PO₄, 0.001 M EDTA, pH 7.2). After baking, the filters were pre-treated for a few hours in hybridization buffer (5x SSCP, 50% formamide, 0.1% SDS) containing 0.02% each of bovine serum albumin, Ficoll and polivinylpyrrolidone. The filters were then hybridized with labelled probe (10⁵ cpm/cm²) at 42°C for 24-36 hours, then washed once in hybridization buffer (½ hour at 42°C), 6 times in 2x SSCP (over several hours at room temperature) and finally twice in 3 mM Tris base (for 10 minutes). After drying they were exposed with Kodak XR5 films and enhancer screens (Ilford Fast Tungstate) for 2-4 weeks.

Electron microscopy. The plasmid DNAs were linearized by digestion with HindIII except 56H8 which was cut with EcoRI. After phenol and ether extractions, the heteroduplexes were prepared and mounted essentially according to Davis et al., but the hybridization was at 37°C for 1 hour and after dialysis the samples were spread in 30% formamide, 0.01% cytochrome C, 0.1 M Tris pH 8.3, 0.01 M EDTA over a hypophase of water.

Lengths were estimated by comparison with internal standards (ØX174 circular single strands (5.375 kb) or pBR322 circular double strands (4.362 kb).

RESULTS

Two hsp 70 genes in diverging orientation. The restriction maps of three Drosophila melanogaster DNA segments, 122, 123 and 168, are shown in Fig. 1. These EcoRI fragments were isolated in λ Charon 9 by screening a random collection of recombinant phages for the presence of hsp 70 gene sequences.

Segment 168 (Fig. 1) has the same restriction map as the EcoRI fragment contained in plasmid 56H8, which we previously described in detail. In the electron microscope, the hetero-
Restriction maps of the Drosophila EcoRI fragments 168, 123 and 122

Full blocks under the lines indicate the position of the sequence Zc which is complementary to the mRNA, hatched blocks indicate the position of Znc. Below the map of 122, thick lines show the extent of the inverted repeat segments as deduced from measurements in the electron microscope (E.M.) and by nuclease SI digestion (SI).

Thin lines at the bottom of the figure locate the two fragments of 122 subcloned in 122X14 and 122B4. The names of some restriction enzymes have been abbreviated (Bam: BamHI, Pst: PstI, RI:EcoRI, Sal:SalI, Xba:XbaI, Xho:XhoI).

duplexes of 168 and 56H8 show complete pairing, thus indicating that we have independently reisolated the same segment (data not shown). The map shows the position of the two components, Znc and Zc, of the basic hsp 70 gene unit Z.

In the map of fragment 122, there is a symmetrical arrangement of a series of sites, forming a large inverted repeat around a central, non-symmetrical region. The comparison with 168
is striking: all the restriction sites which are involved in the symmetry of 122 are also those which span the Znc and Zc segments. This leads to the conclusion that 122 contains two hsp 70 genes arranged in diverging orientation. Further evidence is provided by the heteroduplex analysis reported in a following section: the Z elements of 122 form duplexes with those of 56H8 and 132E3.

After denaturation, the inverted repeat renatures as an intramolecular "snap-back" structure which is readily observed by electron microscopy (Fig. 2). The heteroduplex of the vector Charon 9 and of phage 122 forms a single strand substitution bubble where the Drosophila segment replaces internal vector fragments. Within one of the single strands, the expected stem

Figure 2

Visualization of the inverted repeat in 122 by electron microscopy

Heteroduplexes between the \(\lambda\) Charon 9 vector and the hybrid derivative containing fragment 122 were prepared as described in Materials and Methods. The inverted repeat sequences in 122 form a stem and loop structure which is apparent in the upper right of the micrograph. The two ends of the stem are indicated by arrows.
(arrows) and loop structure is found. The symmetrical segments are $2.74 \pm 0.09$ kb long and are separated by $1.07 \pm 0.10$ kb ($n = 29$). The position and the extent of the inverse repetition were also determined by digesting the snap-back structure with the single-strand specific nuclease SI. The resistant, double stranded fragment, analysed by gel electrophoresis, is 2.7 kb long (data not shown). The position of this fragment on the restriction map is deduced from a further digestion with BamH1 which gives two pieces, 1.55 and 1.15 kb long (Fig. 1). The symmetrical region appears to extend slightly beyond the end of $Z$, but this difference may not be significant. Thus the sequences of the two $Z$ elements account for most, and perhaps all of the inverted repeat in 122.

**Polymorphism in the arrangement of the hsp 70 genes.** The SalI and BglI sites near the beginning of $Zc$ in 122 are characteristic of the hsp 70 gene variants found at the 87A7 locus. In the DNA of Drosophila Kc tissue culture cells and of some Drosophila strains isogenic for the locus, there are two genes of the 87A7 type; they are clustered in a single EcoRI fragment. This fragment with the inverted repeat arrangement represents a major component of Oregon R DNA; it is however not the only arrangement found at 87A7 in the fly population. This polymorphism is reflected by the isolation of other fragments which contain variants of the 87A7 type: 168 (or 56H8) and 123.

Segments 168 and 122 share a common region to the right of the genes, which extends to the RI site at the end of the fragments (Fig. 1). Heteroduplex analysis (see below) shows that the homology of 168 and 122 also extends for 0.6 kb to the left of $Z$. Since only the $Z$ units of the hsp 70 genes are closely conserved while the surrounding sequences differ, the homology found outside of the genes in 168 and 122 suggests that one of these segments may have derived from the other. This could result either from the insertion of a DNA segment in the spacer between the genes or from a deletion.

The restriction maps of 123 and 122 (Fig. 1) are very similar but 123 lacks part of the left gene and most of the
region between the genes, from near the BamHI site in Zc on the left to a position close to the beginning of the gene on the right. Thus 123 is an apparent deletion derivative of 122.

Since the deletion could have arisen during the cloning procedure, I have looked for this derivative in the original DNA (Fig. 3). Restriction fragments of 123 and of total Drosophila DNA digested with SalI and XbaI were separated on agarose gels, transferred to nitrocellulose filters and probed with a radioactive subclone of 56H8 which contains the Zc SalI fragment. The normal Z units have either two or a single SalI site, yielding SalI fragments 2.15 kb or larger (Fig. 3, filled triangles). They all have two XbaI sites, producing 2.1 kb XbaI fragments (Fig. 3, filled triangles). Because the SalI fragment used as a probe extends beyond the XbaI site near the end of Zc, the distal XbaI fragments are detected as faint bands in this experiment. Digestion of 123 by SalI or XbaI generates the usual fragments, but

Figure 3
Detection in Drosophila DNA of the phage 123 restriction fragments containing hsp 70 gene sequences

Drosophila melanogaster (Dm) and phage 123 DNA fragments produced by digestion with XbaI (right) and SalI (left) were separated by agarose gel electrophoresis, transferred to nitrocellulose and hybridized with 32P labelled DNA from a plasmid containing the SalI fragment from Zc of 56H8.

The triangles indicate the position of the fragments derived from the intact Z element (►) and the partially deleted one (▼).
also fragments of smaller size (0.85 and 1.3 kb respectively), which are detected as minor components in the total Drosophila DNA (Fig. 3, open triangles). The truncated gene found in 123 appears to be present in a small fraction of the chromosomes in our Oregon R population. In similar experiments, we found that the EcoRI fragment of 168 and 56H8 is a relatively abundant component in DNA of the same stock.

**Different X elements in front of the genes in 122.** The detailed analysis of the hsp 70 genes of 132E3 and 56H8 revealed that these genes shared a small homologous sequence element named "X" upstream from the beginning of Znc, separated from it by a non-homologous "Y" region.

This additional homology is apparently not shared by the two genes of 122, since no reproducible secondary structure was observed in the loop of the "snap-back" molecule (Fig. 2). I decided however to test the homology of the regions at the beginning of these genes with those of plasmids 132E3 and 56H8. Because it was not possible to observe heteroduplexes between the loop of the 122 "snap-back" and the X elements of the plasmids directly, I have constructed two subclones of 122, each of which contains a single gene. This allows the formation of heteroduplexes of the Z elements by avoiding the intramolecular pairing of the original inverted repeat. These Z duplexes should then facilitate and stabilize the pairing of putative X elements. Subclone 122 X14 (Fig. 1) contains a fragment from a partial XhoI digest of 122 and spans the left gene and the central region. Subclone 122 B4 carries a BamH1 fragment derived from the central region and the first half of the right gene (Fig. 1).

Heteroduplexes of 122 X14 with 132E3 fall into several classes (Table I). The first two types are analogous and show the presence of a common element $X_a$ upstream from the genes in 122 X14 and 132E3. In type I (fig. 5a, table I), the $Y_1$ and $Y_2$ segments have distinctly different sizes. Two arguments assign the larger one, $Y_2$, to 132E3. First, the measure of $k$, from the HindIII site in 132E3 to the beginning of $X_a$, is 1.3 kb. Knowing the distance from the HindIII site to the beginning of Znc (2.5
The measurements of these segments from molecules of type I and II do not differ significantly (less than one standard deviation) and are pooled.

(2) The segments k and m are part of 132E3, they extend from the HindIII cleavage site to the beginning of duplexes Z or D respectively.

(3) Segments m fall in two size classes, depending on whether the duplex D forms with the first or the second repeat of 132E3.

Table 1
Heteroduplex analysis: size measurements

The schematic diagrams (not drawn to scale) indicate the topological structure of the different types of relevant heteroduplexes. For each measurement, the numbers represent the mean length and standard deviation, expressed in kilobase pairs (kb). The number of molecules measured is indicated in the brackets.
kb) places $X_a$ approximately 0.9 kb from Znc: this is the length measured for $Y_2$ (table I). The second argument comes from comparing heteroduplexes of types I and II (Fig. 4). The $Y_1$ region from 122 X14 remains unchanged in both types (Table I), while the $Y_2$ region from 132E3 is included in the V loop of type II.

We can therefore place a region of homology $X_a$ (0.15 kb long) approximately 0.9 kb from the beginning of Znc in 132E3 and 0.15 kb from Znc in front of the left gene of 122 (Fig. 6).

Heteroduplexes of type III have a paired region D which has the length of the Z element. Their occurrence indicates that the hybrids between the $X_a$ elements are not very stable. They are usually found on the second repeat of 132E3 (Table I, $m = 5.30 \pm 0.40$ kb) and thus no $X_a$ element is detected before

![Figure 4](image)

**Heteroduplexes between 122 X14 and 132E3: sequences paired in types I and II**

The dotted lines in these diagrams show the sequences which are paired in heteroduplexes of types I and II (cf. table I). These heteroduplexes form with the pairing of the basic hsp 70 gene unit Z and an additional small region of homology $X_a$, separated from Z by non-homologous sequences Y. Plasmid 132E3 contains two genes in tandem repetition. In type I, the $X_a$ and Z elements of the first 132E3 repeat are paired to the corresponding sequences in 122X14. In type II, $X_a$ from the first 132E3 repeat but Z from the second are involved in the duplex. The V segment belongs to the 132E3 strand and is composed of the $Y_2$ region, of the Z unit from the first repeat and of the spacer between the genes.
the second 132E3 repeat (Fig. 5a). Type III heteroduplexes also sometimes form with the first 132E3 repeat (Table 1, $m = 2.22 \pm 0.29$ kb).

Finally, molecules of type IV are also observed. They have $X_a$, $Y_1$ and $Y_2$ elements identical to those of type I molecules, but the Znc element is interrupted: $R$ is formed with part of Znc from the first 132E3 repeat, and $S$ is formed with $Zc$ from the second. These structures are probably due to a region of decreased homology within Znc in the left unit of 122.

The heteroduplexes between 122B4 and 132E3 belong to types I and II. They show the presence of a homologous element $X_b$ in front of the right gene of 122 and of the first repeat of 132E3 (Fig. 5, Table 1). Comparison of the lengths of the Y elements formed in the two types of heteroduplexes (Table 1) assigns the shorter $Y_3$ element to 122B4 and the longer $Y_4$ to 132E3. Molecules of type III are not observed: the $X_b$ elements form a more stable duplex than the $X_a$ elements.

The heteroduplexes between 122X14 and 56H8 are all of type III and the length of $D$ (Table 1) indicates that the homology is restricted to the Z element.

The heteroduplexes of 122B4 with 56H8 are also of type III, but in this case the pairing extends approximately 0.6 kb further upstream from the left of Znc: $D$ is 2.15 kb (Table 1) while the length of the $Z'$ element to the BamHI site where it is truncated in 122B4 is 1.55 kb. This defines the left end of the homology between 56H8 (or 168) and 122. As described in the previous section, this homology extends to the right beyond $Z$ up to the R1 sites at the end of the fragments.

From the homology of 122B4 and 56H8, and from the position of $X_b$ in 132E3, it is likely that the X element shared by 56H8 and 132E3$^1$ is the same as $X_b$ described here. However we do not detect the presence of an $X_b$ element in front of the second repeat of 132E3. The interaction with 122 may have been too unstable for detection, as it is reported that the X sequences of the second repeat of 132E3 have diverged from those of the first$^1$. 

246
a) 122X14 x 132E3

An additional 122X14 molecule has hybridized to the second repeat.

b) 122B4 x 132E3

The Z' segment is short because the BamH1 fragment subcloned in 122B4 contains only the left part of the Z unit (cf. Fig. 1).

Figure 5

Heteroduplexes of 122X14 and 122B4 with 132E3

The plasmids were linearized by digestion with HindIII, and the heteroduplexes prepared as described in Materials and Methods. Heteroduplexes of type I are shown together with corresponding diagrams. (Thin lines: single stranded DNA, thick lines: DNA duplexes).
Several lines of evidence show that $X_a$ and $X_b$ are different sequences. First, the two elements do not pair in the snap-back formed by 122. Second, $X_b$ is included in the duplex formed with 56H8, while $X_a$ does not interact with 56H8. Finally, their positions relative to $Z_{nc}$ in 132E3 are clearly different, as can be seen from the different lengths of $Y_2$ and $Y_4$ (Fig. 5, a and b; Table 1).

The main results of the heteroduplex analysis are summarized in Fig. 6: the $X_a$ and $X_b$ elements are each present in front of a different gene in 122, they are both found in front of the first gene in 132E3.

**DISCUSSION**

The two hsp 70 gene units in fragment 122 are arranged head to head, in diverging orientation. This configuration of hsp 70 genes of the 87A7 variant type contrasts with the tandem repeats of two or three units which are also found in the Drosophila melanogaster genome at the 87C1 locus \(^1,6,8,9\). The conservation of the two types of arrangements at two distinct

---

**Figure 6**

**Arrangement of the homologous sequences in 122 and 132E3**

The hsp 70 gene units share a common element, $Z$, which was shown previously \(^2\) to consist of a sequence $Z_c$ (complementary to the mRNA) and an additional segment $Z_{nc}$ (non-complementary). The direction of the arrows on $X_a$ and $X_b$ (chosen arbitrarily) shows the relative orientation of these elements in 122 and 132E3.
loci suggests that they may have different selective advantages, perhaps under the diverse conditions which trigger the heat-shock response, though the nature of these advantages remains speculative. It has been estimated that there are 2000-4000 inverted repeat pairs in the *D. melanogaster* genome but it is not known whether the repeats usually comprise coding sequences, as found here for the hsp 70 genes.

In several *D. melanogaster* strains and in Kc cells, the two hsp 70 genes, arranged as in 122, are the only ones present at the 87A7 locus. I have however also isolated from an Oregon R fly population two other DNA segments which carry hsp 70 gene units of the 87A7 variant type. Fragment 123 is an apparent deletion derivative of 122, lacking part of the left gene. Fragment 168 (a new isolate of an EcoRI fragment contained in plasmid 56H8) also shares sequence homology with 122, both upstream and downstream from the gene. It is not clear, however, whether 168 (or 56H8) and 122 would have derived from one another by the deletion or by the insertion of a large sequence. These derivatives are present in a substantial fraction of the chromosomes in our Oregon R fly population and account for some of the polymorphism in the organization of the hsp 70 genes. Two factors could influence the proportion of the variants in the fly population: one is the selective pressure to which they are subjected, which may be low because the heat shock genes occur in multiple copies and because the flies are maintained in the controlled environment of the laboratory. The other factor is the frequency with which variants are produced. In this respect, it may be relevant that on the left of the gene in 56H8 (and 168) there is a dispersed, moderately repetitive sequence which is apparently mobile in the genome (Martin, G.; Mirault, M.-E.; Artavanis-Tsakonas, S. and Schedl, P.; personal communication). This sequence could be involved in the generation of derivatives like those which were isolated. One might consider a simple model where 168 (or 56H8) derives from 122 by the insertion of a mobile sequence in the central region between the genes; imprecise excision could then generate the deletion of 123. The
situation could however be more complex, and involve additional intermediates which were not isolated. A different type of polymorphism occurs at the 87C1 heat-shock locus, where the number of units in the tandem repeats varies, possibly by a mechanism involving unequal crossing-over.

Because large amounts of hsp 70 mRNA are produced at the 87A7 puff site after heat-shock, it may be reasonable to assume that both genes forming the inverted repeat are expressed. If this is the case, signals for the initiation and regulation of their transcription are likely to be present in the central spacer region. In the absence of knowledge on the primary transcripts of these units, we can place the outer limits of the spacer region at the positions corresponding to the 5' ends of the mRNA. Thus the relevant signals would be found within the central segment which includes the symmetrical Znc elements and the non-symmetrical region with the sequence elements X_a and X_b (Fig. 6). Whereas the Znc elements are part of the basic unit conserved with all the hsp 70 genes, X_a and X_b show a more complex pattern. These elements are observed in heteroduplexes of different hsp 70 genes as small regions of homology separated from Znc by non homologous sequences. The element X_b is shared by 132E3 and the gene on the right in 122. It probably corresponds to the X element common to 132E3 and 56H8 described by Moran et al. The element X_a is found further upstream from the first gene in 132E3 and associated with the gene on the left in 122. Thus different hsp 70 genes have different combinations of these elements. It has been suggested that X and Znc sequences at the beginning of the heat shock genes might play a role in mediating their coordinate expression. If this is the case, the different combinations of X_a and X_b elements observed here could allow the differential response of the genes to various signals.

ACKNOWLEDGEMENTS

This work was supported by grant from the Swiss National
Science Foundation no 3.178.77 to A. Tissières. I thank M. Stempfel for her valuable technical assistance, O. Jenni and Y. Epprecht for drawings and plates, E. Boy de la Tour and E. Gallay for advice in electron microscopy, D. Ish-Horowicz, G. Martín, M. E. Mirault, P. Schedl, S. Artavanis-Tsakonas and L. Moran for communicating unpublished results and J.-D. Rochaix for helpful discussions. I am particularly indebted to A. Tissières and M.E. Mirault for their advice and encouragement throughout this work.

REFERENCES