Volume 15 Number 9 1987 Nucleic Acids Research

Expression of intron-containing *C. elegans* heat shock genes in mouse cells demonstrates divergence of 3' splice site recognition sequences between nematodes and vertebrates, and an inhibitory effect of heat shock on the mammalian splicing apparatus

Robert J.Kay, Roland H.Russnak, Donald Jones, Cathleen Mathias and E.Peter M.Candido

Department of Biochemistry, Faculty of Medicine, University of British Columbia, Vancouver, BC V6T 1W5, Canada

Received September 23, 1986; Revised and Accepted April 14, 1987

ABSTRACT

Splicing of a pair of intron-containing heat shock genes from Caenorhabditis elegans has been studied in transfected mouse cells. The hsp16-1 and hsp16-48 genes of C. elegans encode 16,000 Da heat shock polypeptides. Each gene contains a short intron of 52 (hsp16-1) or 55 (hsp16-48) base pairs. When these genes were introduced into mouse cells, they were efficiently induced following heat shock, but splicing of the introns was abnormal. In mouse cells, cleavage of the hsp16 transcripts occurred at the correct 5' splice sites, but the 3' splice sites were located at AG dinucleotides downstream of the correct sites. This aberrant splicing was not solely due to the small size of the C. elegans introns, since a hsp16-1 gene containing an intron enlarged by tandem duplication showed exactly the same splicing pattern. The mouse cells thus seem to be unable to recognize the natural 3' splice sites of the C. elegans transcripts. The efficiency of splicing was greatly reduced under heat shock conditions, and unspliced transcripts accumulated in the nucleus. During a subsequent recovery period at 37°C, these transcripts were spliced and transported to the cytoplasm.

INTRODUCTION

Transcript splicing is used by all eukaryotes to assemble discontinuous protein-coding sequences into a translatable message. While introns are found in only a minority of genes in some unicellular organisms such as S. cerevisiae, they are almost ubiquitous in the genes of more complex organisms. The heat shock genes, which are induced by high temperatures and other forms of stress, are exceptional in that they are predominantly free of introns, even in multicellular organisms (1,2). Since the products of heat shock genes are thought to play a role in the protection of an organism from the adverse effects of the inducing stresses (1,2), one possible rationale for the lack of introns in heat shock genes is that this allows the proteins to be made as rapidly as possible. Another possibility is that the splicing apparatus which removes introns from pre-mRNA's is inactivated by the conditions which induce heat shock genes, and that the lack of introns in the latter allows them to bypass a block in splicing under stress conditions.

© IRL Press Limited, Oxford, England.

We have isolated four genes from the nematode Caenorhabditis elegans which encode 16,000 Da heat shock proteins (3,4). These hsp16 genes are transcriptionally inactive under normal growth conditions, and are efficiently transcribed and translated in heat shocked nematodes. Each hsp16 gene has a single intron located at the boundary between the sequences coding for variable and conserved protein domains. The hsp16 introns range in length from 46 to 58 base pairs, which are typical of intron lengths in C. elegans genes and in the luciferase gene of Photinus (46) but are shorter than the smallest intron that has been found in any vertebrate. Thus, the introns of the C. elegans hsp16 genes are doubly unusual because they are found in heat shock genes and because they are extremely short.

The promoter sequences which determine the heat inducibility of heat shock genes (and presumably the factors which recognize them) have been highly conserved throughout the evolution of eukaryotes. Thus, Drosophila heat shock protein (HSP) genes have been shown to be induced by heat in mammalian cells (5-9) as have Dictyostelium HSP genes in yeast (10). We have recently shown that C. elegans hsp16 genes are correctly expressed when transfected into mouse cells in culture, i.e. they are inducible under the same conditions which induce endogenous mouse heat shock genes, and the transcripts are initiated at the correct positions (11).

We have now used this heterologous expression system to answer two questions that are raised by the unusual properties of the hsp16 introns: are the short introns accurately excised from the hsp16 transcripts in mouse cells, and can the splicing apparatus of the mouse cells function effectively under conditions that induce the expression of heat shock genes?

MATERIALS AND METHODS

Construction of hsp16 gene pairs and insertion into transfection vectors

Methods for restriction enzyme digestion of DNA, fragment purification by agarose gel electrophoresis and electroelution, ligation, and cloning in E. coli were essentially as described (12). Restriction enzymes and polymerases were obtained from Bethesda Research Laboratories, New England Biolabs, Pharmacia, or Boehringer Mannheim.

The assembly of the wild-type hsp16 gene pair as a 1926 bp BamHI fragment in the transfection vectors pPN1 and pPN3 has been described previously (11). This 1926 bp fragment was excised from pPN1 by digestion with BamHI and inserted into the unique BamHI site of pCGBPV9 ΔB5 (13), with the hsp16-48 gene adjacent to the NPT II gene of the vector.

For the construction of the hsp16 gene pair bearing a duplication of the hsp16-1 intron, the 1926 bp BamHI fragment was inserted into the BamHI site of pUC13 (14), with the hsp16-1 gene adjacent to the HindIII site of the vector. The following fragments were then excised from this pUC13-hsp16 construct: 1) the 695 bp DdeI-SalI fragment containing the intron and 3' exon of the hsp16-1 gene and 2) the 3967 bp HindIII-HpaI fragment containing the pUC13 sequences, the complete hsp16-48 gene, and the 5' exon and intron of the hsp16-1 gene. The excised fragments were made blunt-ended by treatment with the Klenow fragment of DNA Polymerase I and all four deoxynucleotides, purified by agarose gel electrophoresis, and ligated together. The structure of the resulting pUC13-hsp16ID construct was determined by restriction endonuclease site mapping, and the exact sequence of the duplicated hsp16-1 intron was determined by the chain termination method of Sanger (15) after transferring the 203 bp Sau3A1 fragment of the hsp16ID gene pair into M13mp18 (16). The 1988 bp BamHI fragment containing the hsp16ID gene pair was then excised from the pUC13 vector and inserted into the BamHI site of pCGBPV9 ΔB5, with the hsp16-48 gene adjacent to the NPT II gene of the transfection vector.

Transfection

C127 cells (10^5) were transfected with 1 μg of pCGBPV9 ΔB5-hsp16 or 1 μg of pPN3-hsp16 plus BPV DNA by the calcium phosphate coprecipitation method and selected for resistance to the drug G418 as described previously (11). Cell lines were established from single G418-resistant colonies.

Induction of gene transcription and isolation of nucleic acids

Adult C. elegans were grown in liquid culture as described (17). The worms were collected by centrifugation, purified by flotation in 30% sucrose, resuspended in basal S medium (100 mM NaCl, 50 mM potassium phosphate, pH 6.0) in an aerated Roux bottle, and heat shocked by immersion in a 33°C water bath for 2 h. After harvesting by centrifugation, which exposes the worms to room temperature for about 30 min., they were suspended in 250 mM sucrose, 10 mM $MgCl_2$, 1 mM ethylene glycol-bis-(β-amino ethyl ether) N,N'-tetra-acetic acid (EGTA), 10 mM Tris-Cl pH 8.0, frozen in liquid nitrogen, and ground with a mortar and pestle. RNA was purified from the homogenate as described (18). The heat shock response was induced in subconfluent C127 cell cultures by transferring the culture flasks to a water bath at the appropriate temperature. For recovery from heat shock, the flasks were moved to an incubator chamber at 37°C. The cells were lysed with sodium dodecyl sulfate (SDS) and digested with Proteinase K, and nucleic acids were isolated as described previously (11). For arsenite induction, sodium arsenite was added to the medium in a culture

flask to a final concentration of 0.1 mM. After 90 min at 37°C the medium containing arsenite was replaced with fresh medium and the cells were returned to 37°C for 60 min.

Analysis of the subcellular distribution of unspliced hsp16 transcripts

C127 cells from two 25 cm^2 flasks were detached by rinsing in 75 mM NaCl, 25 mM Na EDTA pH 7.5 and collected by centrifugation. After one rinse in ice cold 0.14 M NaCl, cells were again pelleted and resuspended in 75 μl of 0.14 M NaCl. To the cell suspension was then added 1.0 ml of Buffer K (1.0 mM Tris-HCl pH 8.0, 1.0 mM EGTA, 1.0 mM $MgCl_2$, 1.0 mM dithiothreitol), and 1 μl (36 units) of the ribonuclease inhibitor, RNasin (Boehringer Mannheim). The cells were then lysed by the gradual addition, with constant mixing, of 25 μl of 10% Nonidet P40 (final concentration 0.25%). The supernatant (cytosolic) and pellet (nuclear) fractions were collected after centrifugation at 500xg for 5 min and RNA was extracted as follows: To the supernatant, an equal volume of 20mM sodium acetate pH 5.0, 2% SDS was added, and the resulting solution was extracted for 3 min at 60°C with an equal volume of phenol. The aqueous phase was removed following centrifugation (maximum speed in a bench top clinical centrifuge for 4 min), and re-extracted with an equal volume of phenol as before. The nuclear pellet fraction was treated in the same way, except that the pellet was first suspended in 1.0 ml of ice cold Buffer K prior to the addition of sodium acetate-SDS and phenol extraction. RNA was precipitated from aqueous phases overnight at -20°C after the addition of sodium acetate to 0.2 M and 2.5 volumes of 95% ethanol. The precipitate was collected by centrifugation at 8000 xg for 10 min, washed once with cold 70% ethanol, air dried and dissolved in 20 μl of sterile diethylpyrocarbonate (DEP) treated water. All solutions used for RNA extraction were prepared from DEP treated water as described (12).

Structural analysis of hsp16 gene pairs in transfected cells

Total nucleic acid from transfected cells was digested with RNase A and BamHI, separated by electrophoresis through agarose gels, and transferred to nitrocellulose filters as described (3). The 1926 bp BamHI fragment containing the wild-type hsp16 gene pair was labelled by nick translation (19) and hybridized to the DNA on the nitrocellulose filters. The filters were washed as described (3), and exposed to Kodak XAR film with Cronex intensifying screens (Dupont).

Structural analysis of hsp16 transcripts by S1 nuclease protection

Single-stranded, continuously labelled DNA was synthesized from inserts in M13mp vector templates by extension of annealed primers, excised as a discrete

fragment by restriction endonuclease digestion, and purified by electrophoresis as described previously (11).

For mapping the hsp16-1 splice sites, the 1045 bp BglII fragment of the wild type hsp16 gene pair and the 1107 bp BglII fragment of the hsp16ID gene pair were inserted into the BamHI sites of M13mp8 and M13mp18, respectively, with the hsp16-1 sequences adjacent to the primer annealing site in the vector. After primer extension, the probe fragments were excised with XbaI.

For mapping the hsp16-48 5' splice site, the 1045 bp BglII fragment was inserted into M13mp8 with the hsp16-48 sequences adjacent to the primer annealing site, and the probe was excised at the HpaI site within the hsp16-1 sequences.

For mapping the hsp16-48 3' splice sites, the 377 bp SspI fragment of hsp-16-48 was inserted into the SmaI site of M13mp18 with the 3' exon sequences adjacent to the primer annealing site, and the probe was excised by digestion at the SstI site in the vector.

The purified single-stranded probes were hybridized to cellular nucleic acid and digested with S1 nuclease, and the lengths of protected probe fragments were determined by electrophoresis through denaturing polyacrylamide gels and autoradiography as described (11).

RESULTS

Introduction of C. elegans hsp16 genes into mouse cells

The structure of the gene pair used in the present studies is shown in Figure 1A. The hsp16-1 and hsp16-48 genes of C. elegans are organized as two closely linked, divergently transcribed pairs (3). Each gene contains a single intron of 52 base pairs (hsp16-1) or 55 base pairs (hsp16-48). These hsp16 genes are efficiently transcribed in response to heat shock after being introduced into mouse C127 fibroblasts on bovine papillomavirus (BPV)-derived vectors (11).

A 1926 bp fragment of DNA bearing a hsp16-1 + hsp16-48 gene pair was inserted into the BamHI site of the transfection vector pCGBPV9ΔB5 (13) or pPN3 (see Materials and Methods). Both of these vectors include a gene conferring resistance to the drug G418, and BPV plasmid maintenance sequences that allow the establishment of the vector DNA to stable copy numbers of between ten and one hundred in transfected cells (20,21). The BPV genes encoding the trans-acting factors required for extrachromosomal replication of DNA linked to plasmid maintenance sequences are contained within pCGBPV9ΔB5, and were cotransfected in the case of pPN3. Mouse C127 cells were transfected with the

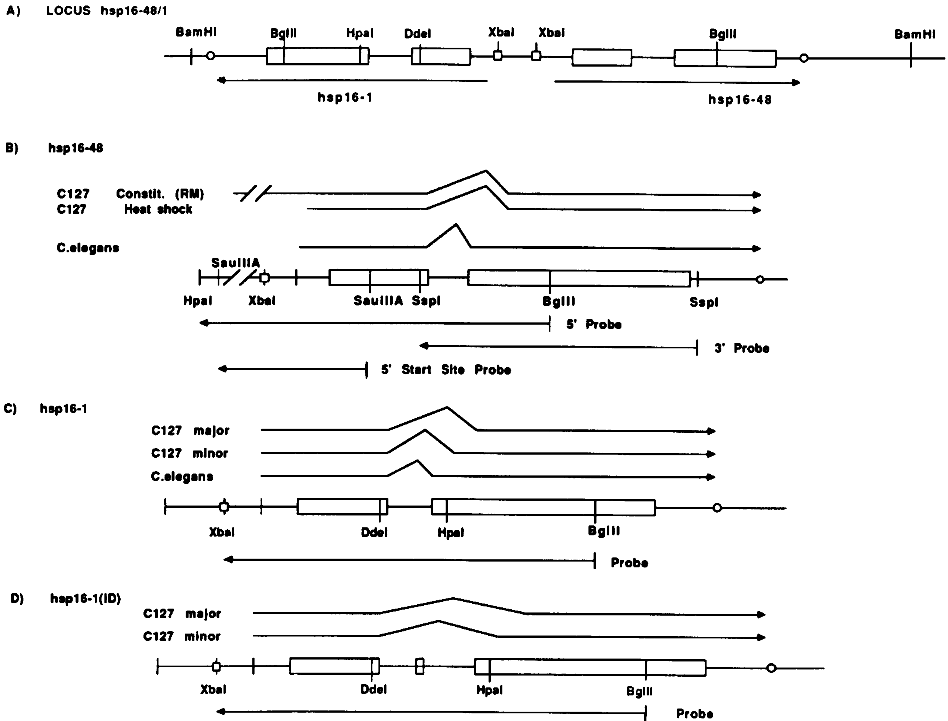

Figure 1. Structures of hsp16 genes and transcripts. Genes are represented by continuous lines, with the small boxes indicating heat-inducible promoters, the large boxes indicating protein-coding regions, and the circles indicating polyadenylation sites. A) The 1926 bp BamHI fragment containing the hsp16 gene pair, with the direction of transcription indicated by the arrows below the genes. B-D) Transcript structures and splice site mapping probes. The transcripts are represented by the arrows above the genes, with the kinks in the arrows indicating the regions removed by splicing. The constitutively synthesized hsp16-48 transcript initiating at the cryptic upstream promoter in cell line RM is also shown. The single-stranded probes used to locate splice sites by S1 nuclease protection are shown below the genes.

vector-borne hsp16 gene pair by the calcium phosphate coprecipitation method and selected for growth in the presence of G418. The resulting colonies were expanded into cell lines and tested for the presence of intact hsp16 gene pairs by Southern blot hybridization. From this analysis, the copy numbers of the hsp16 genes were estimated to be between twenty and one hundred in the cell lines used for the analysis of splicing (data not shown).

Splicing of the wild-type hsp16 genes in mouse cells

The various forms of spliced and unspliced hsp16 transcripts were identified by their protection of continuously labelled, single-stranded DNA probes

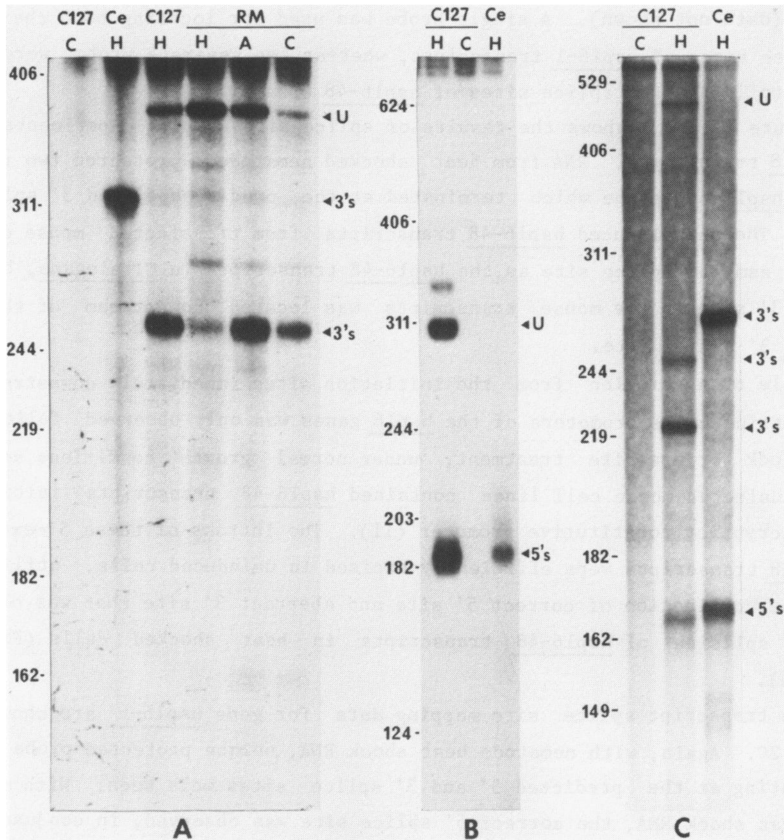

Figure 2. Splice sites in hsp16 transcripts. RNA from C. elegans (Ce) or nucleic acid from C127 cells transfected with the hsp16 gene pair (C127) was hybridized to the following probes (see Fig. 1): the hsp16-48 probe extending between the two SspI sites (Panel A), the hsp16-48 probe extending from the BglII site to the HpaI site (Panel B), or the hsp16-1 probe extending from the BglII site to the XbaI site (Panel C). After digestion with S1 nuclease, the probe fragments were separated by electrophoresis through denaturing 6% polyacrylamide gels. H,A, and C indicate the use of nucleic acid from heat induced, arsenite induced, or uninduced cells, respectively. C127 cells were heat shocked for 1h at 42.5°C, followed by 30 min of recovery at 37°C. Arsenite induction of C127 cells and heat induction of C. elegans are described in Methods. Nucleic acid from the constitutive C127 cell line is indicated by RM in panel A. U, probe protected by unspliced transcripts; 5's and 3's, probe protected by 5' and 3' portions of spliced transcripts, respectively. The size markers are end-labelled HpaII fragments of pBR322.

from degradation by nuclease S1. The probes used for this purpose are illustrated in Figure 1. The 3' regions of spliced transcripts were distinguished from 5' regions by their ability to protect the 5' portions of end-labelled

probes (data not shown). A single probe was used for locating both the 5' and 3' splice sites of hsp16-1 transcripts, whereas two separate probes were required to locate the splice sites of hsp16-48 transcripts.

Figure 2A and B shows the results of splice site mapping experiments for hsp16-48 transcripts. RNA from heat shocked nematodes protected two regions of the hsp16-48 probe which terminated at the predicted 5' and 3' splice sites. The heat-induced hsp16-48 transcripts from transfected mouse cells had the same 5' splice site as the hsp16-48 transcripts in C. elegans, but the single 3' site in the mouse transcripts was located downstream of the C. elegans 3' splice site.

While transcription from the initiation sites immediately downstream of the heat-inducible promoters of the hsp16 genes was only observed following heat shock or arsenite treatment, under normal growth conditions some of the transfected mouse cell lines contained hsp16-48 transcripts initiating from a cryptic, constitutive promoter (11). The introns of these 5'-extended hsp16-48 transcripts were efficiently excised in uninduced cells, utilizing the same combination of correct 5' site and aberrant 3' site that was observed in the splicing of hsp16-48 transcripts in heat shocked cells (Fig. 2, panel A).

The transcript splice site mapping data for gene hsp16-1 are shown in Figure 2C. Again, with nematode heat shock RNA, unique protected probe bands terminating at the predicted 5' and 3' splice sites were seen. With mouse cell heat shock RNA, the correct 5' splice site was observed, in conjunction with a pair of 3' splice sites located downstream from that used in C. elegans. While only the spliced forms of the hsp16 transcripts were found in C. elegans, some completely unspliced transcripts of both hsp16 genes could be detected in mouse cells.

The locations of the splice sites mapped in these experiments are shown in Figure 3. The two 3' splice sites of the hsp16-1 transcripts in transfected mouse cells are at AG dinucleotides 73 and 100 nucleotides downstream of the 5' splice site, the distal 3' site being much more frequently used. The 3' splice site of the hsp16-48 transcripts in mouse cells is located 113 nucleotides away from the 5' splice site, at the more proximal of a pair of AG dinucleotides.

Effect of duplication of an hsp16 intron on the selection of 3'splice sites in mouse cells

The exclusive use in transfected mouse cells of downstream AG dinucleotides as 3' splice sites in hsp16 transcripts, in preference to the correct 3'

Figure 3. Intron sequences and splice sites. The 5' and 3' hsp16 splice sites used in C. elegans and transfected C127 cells are indicated by the vertical arrows. Sequences resembling the branch site consensus sequence (YTRAY) are underlined. The potential but unused 3' splice site within the second intron of the hsp16-1(ID) gene is overlined. (H/D) indicates the ligated HpaI and DdeI sites in the hsp16-1(ID) construct. Numbering of the sequence starts at the first base in the intron.

splice sites used with 100% efficiency in C. elegans, suggested that the very short lengths of the hsp16 introns might be incompatible with the splicing apparatus of mouse cells. Deletion studies of a rabbit β-globin gene intron (22) and of an adenovirus E1A gene intron (23), and expansion of the SV40 small t intron to lengths greater than its natural 66 nucleotides (47) indicate a requirement for a minimal intron length of approximately 70-80 nucleotides for efficient splicing in mammalian cells. The predominant spliced hsp16 transcripts in mouse cells had lost introns that were significantly longer than this minimum length, while the 3' splice site of the hsp16-1 transcript which defined an intron close to the minimum length was inefficiently utilized.

In order to test the effect of intron length on 3' site selection, we constructed an hsp16-1 gene containing an enlarged intron. Rather than insert extraneous DNA sequences into the natural intron, we chose to tandemly duplicate it as shown in Figure 1D. This approach was taken in order to minimize the possibility that the introduction of foreign sequences might lead to the generation of novel splice sites or affect the efficiency of splicing in unpredictable ways. The sequence of the duplicated intron is shown in Fig. 3. In hsp16-1 transcripts containing the duplicated intron, a variety of splic-

ing modes could potentially occur, involving combinations of the 5' splice sites and the natural or aberrant 3' splice sites.

The hsp16-1 construct containing the intron duplication (hsp16-1ID) was inserted into the BPV vector pCGBPV9ΔB5 as part of a 1988 bp BamHI fragment which also contained the wild-type hsp16-48 gene in its normal relationship to hsp16-1. Mouse cells were transfected with vector DNA and G418 resistant clones were selected as before. A resulting cell line containing approximately 500 copies of the intact hsp16-1ID + hsp16-48 gene pair (determined by Southern blotting, data not shown) was used to determine the mode of splicing of the hsp16-1ID transcripts. Both the 5' and 3' splice sites proved to be identical in hsp16-1 transcripts carrying either single or duplicated introns, i.e. the natural 5' site at the boundary of the first exon was ligated to one of the two aberrant 3' sites within the protein-coding sequences of the second exon (Fig. 4). Neither the distal natural 3' splice sites nor AG dinucleotides lying between the proximal and distal natural 3' sites in the hsp16-1ID transcripts were utilized for splicing, despite being located more than 80 bases away from the 5' splice site. The doubling of the 5' splice band seen in this Figure (and in Fig. 6) likely resulted from excessive digestion by the S1 nuclease. Splicing of the hsp16-48 transcripts was the same in mouse cells transfected with gene pairs containing either wild-type or intron-duplicated hsp16-1 (data not shown).

Effect of heat shock conditions on hsp16 transcription and splicing in mouse cells

In the above studies a significant degree of variation in the ratio of unspliced to spliced transcripts was observed between different experiments. This prompted us to examine systematically the effects of heat shock temperature and length of the recovery period on transcription and processing of the hsp16 genes in the transfected mouse cells. Figure 5 presents the results of these experiments for the hsp16-48 gene. Similar results were obtained for hsp16-1 (data not shown).

For a one hour heat shock with no recovery period, increasing the temperature of the shock from 41.5°C to 43.5°C resulted in a dramatic increase in the amount of hsp16-48 transcripts (Fig. 5, lanes 1-3). However, the proportion of these transcripts which were spliced declined with increasing heat shock temperatures. Raising the heat shock temperature to 44.5°C for one hour caused a significant decrease in the number of hsp16-48 transcripts, although some spliced transcripts were still seen (Fig. 5, lane 4). Extending the period of heat shock at 42.5°C to 2 hr resulted in a small increase in the number

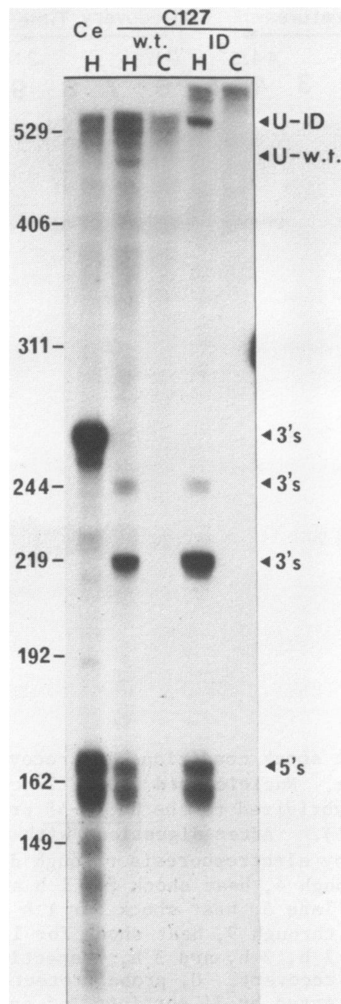

Figure 4. Comparison of splice sites in hsp16-1 transcripts bearing single or duplicated introns. RNA from C. elegans (Ce) or nucleic acid from C127 cells transfected with either the wild type (C127 WT) or intron-duplicated hsp16 gene pair (C127 ID) was hybridized to the appropriate hsp16-1 probe extending from the BglII site to the XbaI site (Fig. 1). After digestion with S1 nuclease, the probe fragments were separated by electrophoresis through denaturing 6% polyacrylamide gels. H and C indicate the use of nucleic acid from heat shocked or uninduced cells, respectively. C127 cells were heat shocked for 1 h at 42.5°C, followed by 30 min of recovery at 37°C, while C. elegans were heat shocked as described in Methods. U-w.t. and U-ID, probe protected by unspliced wild type and intron-duplicated hsp16-1 transcripts, respectively; 5'S and 3'S, probe protected by 5' and 3' portions of spliced transcripts, respectively. The size markers are end-labelled HpaII fragments of pBR322.

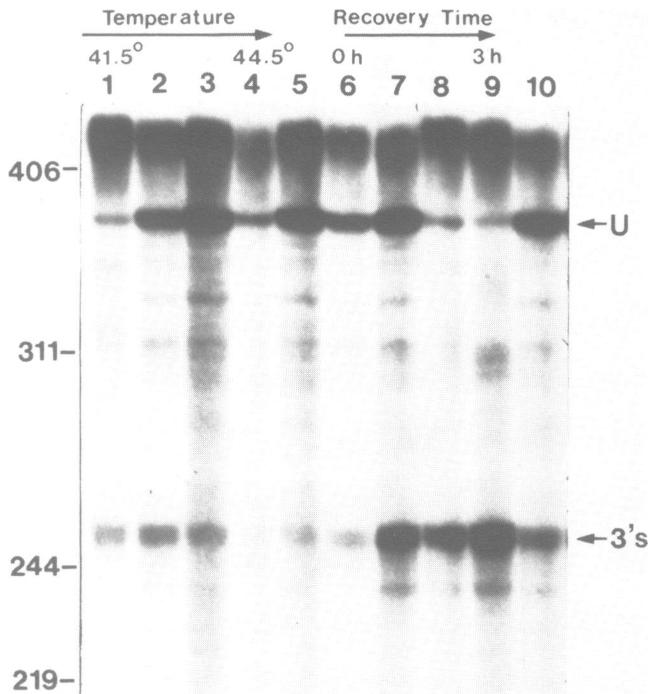

Figure 5. Effects of heat shock conditions and recovery times on hsp16-48 transcription and splicing. Nucleic acid from C127 cells transfected with the hsp16 gene pair was hybridized to the hsp16-48 probe extending between the two SspI sites (Fig. 1). After digestion with S1 nuclease, the probe fragments were separated by electrophoresis through denaturing 4% polyacrylamide gels. Lanes 1 through 4, heat shock for 1 h at 41.5°C, 42.5°C, 43.5°C, and 44.5°C, respectively; lane 5, heat shock for 1 h at 43.5°C followed by 1h recovery at 37°C; lanes 6 through 9, heat shock for 1 h at 42.5°C followed by recovery at 37°C for 0 h, 1 h, 2 h, and 3 h, respectively; lane 10, heat shock for 2 h at 42.5°C with no recovery. U, probe protected by unspliced transcripts; 3' S, probe protected by 3' portions of spliced transcripts. 5' portions of spliced transcripts are not detected by this probe. The size markers are end-labelled HpaII fragments of pBR322.

of hsp16-48 transcripts without significantly affecting the proportion of the transcripts that were spliced (Fig. 5, lane 10).

When the transfected mouse cells were returned to 37°C after a 1 hr heat shock at 42.5°C, the amounts of hsp16-48 transcripts increased several-fold over the next hour and then remained approximately constant for two more hours (Fig. 5, lanes 6-9). During the recovery period at 37° C the proportion of spliced transcripts continually increased, to almost 100% after three hours. Raising the temperature of heat shock from 42.5°C to 43.5°C for 1 hour appears

to reduce the efficiency of both transcription and splicing, resulting in no increase in the level of either total or spliced hsp16-1 transcripts during a subsequent 1 hour recovery period (Fig. 5, lane 5).

Thus, high levels of hsp16-48 transcripts are induced and accumulate during the period of heat shock, but the efficiency of splicing is low under optimal induction conditions. The unprocessed transcripts can all be spliced during a subsequent recovery period at normal temperature, if the heat shock conditions are not too severe.

Treatment of the mouse cells with 100 μM arsenite for 1 hr induced transcription of hsp16-48 and inhibited subsequent splicing to the same extent as did a 42.5°C heat shock for 1 hr (data not shown). This suggests that inhibition of splicing may be coupled to the induction of the heat shock response in mouse cells rather than being simply due to a direct effect of high temperatures.

Intracellular localization of unspliced hsp16 transcripts in mouse cells

The distribution of hsp16-48 transcripts within transfected C127 cells was examined to determine whether their export to the cytoplasm was dependent on splicing. Under heat shock conditions that resulted in the accumulation of a large amount of unspliced hsp16 transcripts, the nuclei contained both spliced and unspliced hsp16-48 transcripts in roughly equal amounts, while the hsp16-48 transcripts in the cytoplasm of the same cells were almost exclusively in the spliced form (Figure 6). Therefore, the transport of hsp16 transcripts to the cytoplasm is dependent on intron removal as is the case for endogenous mRNA precursors containing introns (24), and the heat shock conditions do not prevent this segregation of spliced and unspliced transcripts.

DISCUSSION

When a fragment of *C. elegans* DNA containing the pair of small heat shock genes hsp16-1 and hsp16-48 is transfected into mouse cells, the *C. elegans* genes are correctly expressed under mammalian heat shock conditions (11). The hsp16 genes are initiated at their normal transcription start sites, in response to activation of their upstream heat shock promoter elements (11).

In contrast to the high degree of conservation in the transcription regulatory sequences, the splicing of the nematode gene transcripts did not follow its natural mode in a heterologous host. No correctly spliced transcripts for either gene were detectable in mouse cells. However, specific spliced RNA products were observed for each gene. These arose as the result of correct

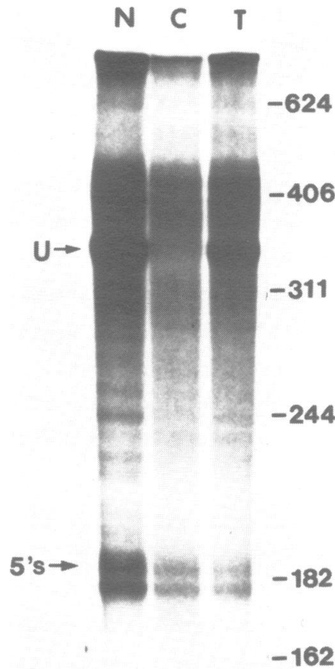

Figure 6. Distribution of hsp16 transcripts within transfected C127 cells. RNA prepared from the nuclei (N), cytoplasm (C), or from whole cells (T) transfected with the hsp16 gene pair was hybridized to a hsp16-48 probe extending from the EcoRI site 44 bp downstream of the BglII site (Fig. 1 B) to the XbaI site. After digestion with S1 nuclease, the probe fragments were separated by electrophoresis through denaturing 6% polyacrylamide gels. The cells were heat shocked for 1 h at 42.5°C, followed by 30 min of recovery at 37°C. U, probe protected by unspliced transcripts; 5'S, probe protected by 5' portions of spliced transcripts. The size markers are end-labelled HpaII fragments of pBR322.

cleavage at the 5' splice sites, coupled with the use of alternate 3' sites downstream of the natural sites used in C. elegans. In the case of hsp16-1, two alternate 3' sites were used, corresponding to AG dinucleotides located 21 and 48 nucleotides downstream of the natural site, the latter being used at the higher frequency. In the case of hsp16-48, a single 3' site was used which is 58 nucleotides downstream of the natural one. The aberrant mode of splicing of the hsp16 transcripts in transfected mouse cells did not result

from heat- or arsenite- induced damage to the splicing apparatus, since the same pattern of splicing was also observed in hsp16-48 transcripts produced from a cryptic upstream promoter that was active under normal growth conditions.

A common mechanism for removing introns from protein-coding transcripts appears to be used by most eukaryotes (24,25). Splicing occurs via a two step process involving three separate sites within the transcript. The 5' splice site is first cleaved and the resulting free end of the intron is attached to an adenosine residue at the branch site within the intron through a 5'-2' phosphodiester bond. In the second step, the 5' and 3' splice sites are ligated together, and the intron sequences are released as a lariat-shaped structure.

Consensus sequences for all three splicing sites have been derived from comparisons of many different sequences. The 5' site consensus sequence is AG/GTARGT (24-26) with the slash indicating the position of cleavage. R signifies a purine residue. The 5' splice sites of the hsp16 genes closely resemble this sequence (Fig. 3). The 3' site consensus sequence is Y_nNCAG/G, where Y_n is a series of pyrimidines which is typically about 11 nucleotides long and may contain several interspersed purines (24-26). The hsp16-48 3' splice site is a good match to this consensus sequence but the hsp16-1 3' site has a pyrimidine stretch that extends for only six bases. The relatively small number of branch sites defined so far for higher eukaryotes have the consensus sequence YTRAY, where the underlined A is the site of branch formation (27-29). None of the other bases in this sequence are strictly conserved, and sites with no resemblance to the branch consensus can serve as branch sites when the natural branch site has been removed (30,31). Sequences similar to the branch consensus are found within the hsp16 introns (Fig. 3).

All known branch sites are located within 18 to 50 bases of the 3' splice site and the minimum functional spacing between the 5' splice site and the branch site appears to be about 50 bases in mammalian cells (30). These spacings fix a minimum intron size of about 70 bases, which is in agreement with the size limit determined by the analysis of deleted and expanded introns (22,23,47). Thus the hsp16 introns with lengths of 52 or 55 bases cannot have sufficient spacing between the 5', branch, and 3' sites to permit their natural mode of splicing in mammalian cells. The 5' to branch spacing appears to be adequate because the correct 5' site can be cleaved. Therefore, the spacing between the branch site utilized in the transfected mouse cells and the natural 3' splice site of the hsp16 transcripts must be too short. This

would necessitate the observed use of alternate 3' splice sites located further downstream.

In the hsp16-1 transcripts with the tandemly duplicated intron, the spacing between the upstream natural 5' splice site and the downstream natural 3' splice site is 114 bases, and so it should have been possible to ligate these two sites together, particularly since introns with lengths of 73 and 100 bases could be excised from the wild-type hsp16-1 transcripts in the transfected mouse cells. However, this mode of splicing was not observed; instead, the upstream 5' splice site was ligated to the same alternate downstream 3' sites that were used for splicing of the hsp16-1 transcripts with single introns. It seems that the alternate downstream 3' sites were being selected in preference to the natural 3' site of the hsp16-1 transcripts on the basis of their sequences rather than their permissible distance from the 5' splice site. This preference cannot be rationalized solely on the basis of the sequences that are immediately adjacent to the 3' splice sites. Both the natural 3' splice site and a sequence within the hsp16-1 intron, which is 86 bases downstream from the 5' splice site when the intron is duplicated, are much better matches to the consensus sequence than is the minor downstream 3' site that is actually used in the mouse cells. Possible branch site sequences can be found within the prescribed limits upstream of both the natural 3' splice sites, the potential 3' site within the hsp16-1 intron, and the aberrant 3' sites used in mouse cells (Fig. 3). Differences in the efficiency of utilization of these branch sites may have dictated the selection of 3' splice sites.

Among most of the eukaryotic systems examined to date, the mechanism of splice site recognition and cleavage has been highly conserved. The introns of human β-globin mRNA are correctly spliced in Xenopus oocytes (37), and the fission yeast Schizosaccharomyces pombe correctly excises the intron from SV40 small-t antigen transcripts (38). A notable exception to splicing compatibility is the budding yeast, Saccharomyces cerevisiae. This organism is incapable of correctly removing introns from transcripts of higher eukaryotic genes (39-41). This seems to be due to the stringent requirement for the presence of the branch consensus sequence TACTAAC near the 3' splice site (42-44).

The inability of mammalian cells to correctly recognize the 3' splice sites of C. elegans introns, and the generally small size of the nematode introns, suggests that the splicing components of mammals and nematodes have diverged significantly. Nematodes have apparently developed a mechanism for

handling short spacings between splice sites as an adaptation to the presumed selective advantage (as yet unknown) conferred on these organisms by drastically reduced intron sizes. Ultimately, a comparison of the mechanisms of splicing between these groups of organisms may help to elucidate important features of this process in higher eukaryotes.

It is difficult to assess the efficiency of splicing of the C. elegans genes in the transfected mouse cells because inhibition of splicing is a consequence of the conditions which induce transcription of the hsp16 genes. While the transcription apparatus of the mouse cells is still functional at temperatures between 42°C and 44°C, the splicing capabilities of the cells are drastically curtailed. When normal growth conditions are restored after heat shock, splicing resumes with increasing efficiency while transcription of the hsp16 genes is repressed. These results parallel recent findings in Drosophila, demonstrating an inhibition of intron splicing at high heat shock temperatures (32). Heat shock genes which contain introns cannot be effectively expressed as long as splicing is inhibited by the conditions that induce their transcription. This explains the absence of introns in most heat shock genes. Two other heat shock genes which do contain introns are the Drosophila hsp83 gene and the human hsp26 gene (45). The former gene is maximally transcribed under moderate heat shock conditions which do not significantly inhibit splicing (32), while the pattern of transcription and splicing of the latter gene is unknown. It is possible that C. elegans has developed a splicing apparatus that is not inhibited by heat shock. However, the pattern of protein synthesis in heat shocked C. elegans indicates that splicing of hsp16 transcripts may also be inhibited under extreme induction conditions in their normal cellular environment. Synthesis of the 16,000 Da heat shock proteins declines with increasing heat shock temperatures (33), while the synthesis of other heat shock proteins continues unabated. Furthermore we have recently isolated and characterized, from C. elegans heat shock mRNA, an hsp16 cDNA which contains an unspliced intron sequence (4). It is not known whether the hsp16 genes are efficiently transcribed at the higher heat shock temperatures, or whether the genes encoding the other heat shock proteins of C. elegans contain introns.

Heat shock does not result in a failure to recognize the presence of introns in transcripts, because unspliced hsp16-48 transcripts were still specifically sequestered in the nucleus during heat shock. Rather, the actual splicing reactions are probably subject to inhibition. Splicing inhibition is also observed after the induction of heat shock gene transcription by

arsenite, and so it might be a specific feature of the heat shock response, mediated by the factors that initiate the response or by the products of heat shock gene expression. In Drosophila, some heat shock proteins are associated with ribonucleoprotein complexes (34-36) and prior induction of heat shock proteins increases the efficiency of splicing during a subsequent period of heat shock (32). Therefore, one of the roles of heat shock proteins may be to protect the splicing apparatus during heat shock and other forms of stress, and to promote its recovery following a return to normal conditions.

ACKNOWLEDGEMENTS

We wish to thank Roger Graham for a preparation of C. elegans heat shock mRNA. The vector pCGBPV9ΔB5 was kindly provided by P. Matthias. This work was supported by the Medical Research Council of Canada and the B.C. Health Care Research Foundation. R.J.K. and R.H.R. held MRC of Canada Studentships.

REFERENCES

1. Craig, E.A. (1985) Crit. Rev. Biochem. 18, 239-280.
2. Lindquist, S. (1986) Annu. Rev. Biochem. 55, 1151-1191.
3. Russnak, R.H., and Candido, E.P.M. (1985) Mol. Cell. Biol. 5, 1268-1278.
4. Jones, D., Russnak, R.H., Kay, R.J. and Candido, E.P.M. (1986) J. Biol. Chem., in press.
5. Corces, V., Pellicer, A., Axel, R., and Meselson, M. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 7030-7042.
6. Mirault, M.E., Southgate, R. and Delward, E. (1982) EMBO J. 1, 1279-1285.
7. Pelham, H.R.B. (1982) Cell 30, 517-528.
8. Burke, J.F. and Ish-Horowicz, D. (1982) Nucleic Acids Res. 10, 3821-3830.
9. Ayme, A., Southgate, R. and Tissières A. (1985) J. Mol. Biol. 182, 469-475.
10. Cappello, J., Zuker, C. and Lodish, H. (1984) Mol. Cell. Biol. 4, 591-598
11. Kay, R.J., Boissy, R.J., Russnak, R.H., and Candido, E.P.M. (1986) Mol. Cell. Biol. 6, 3134-3143.
12. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). "Molecular Cloning: A Laboratory Manual". Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
13. Matthias, P.D., Bernard, H.U., Scott, A., Brady, G., Hashimoto-Gotoh, T. and Schütz, G. (1983) EMBO J. 2, 1487-1492.
14. Vieira, J. and Messing, J. (1982). Gene 19, 259-268.
15. Sanger, F., Nicklen, S. and Coulson, R.A.R. (1977) Proc. Nat. Acad. Sci. U.S.A. 74, 5463-5467.
16. Messing, J. (1983) Methods Enzymol. 101, 20-78.
17. Brenner, S. (1974). Genetics 77, 71-94.
18. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299.
19. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
20. Di Mario, D., Treisman, R. and Maniatis, T. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 4030-4034.

21. Sarver, N., Byrne, J.C., and Howley, P.M. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 7147-7151.
22. Wieringa, B., Hofer, E. and Weissmann, C. (1984) Cell 37, 915-925.
23. Ulfendahl, P.J., Pettersson, U. and Akusjärvi, G. (1985) Nucleic Acids Res. 13, 6299-6315.
24. Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A. (1986) Annu. Rev. Biochem. 55, 1119-150.
25. Rogers, J.H. (1984) Int. Rev. Cytol. Suppl. 17, 1-71.
26. Mount, S.M. (1982) Nucleic Acids Res. 10, 459-472.
27. Ruskin, B., Kraimer, A.R., Maniatis, T. and Green, M.R. (1984) Cell 38, 317-331.
28. Zeitlin, S. and Efstratiadis, A. (1984) C 39, 589-602.
29. Reed, R. and Maniatis, T. (1985) Cell 41, 95-105.
30. Ruskin, R., Greene, J.M. and Green, M.R. (1985) Cell 41, 833-844.
31. Padgett, R.A., Konarska, M.M., Aebi, M., Hornig, H., Weissmann, C. and Sharp, P.A. (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 8349-8353.
32. Yost, H.J. and Lindquist, S. (1986) Cell 45, 185-193.
33. Snutch, T.P. and Baillie, D.L. (1983) Can. J. Biochem. Cell Biol. 61, 480-487.
34. Kloetzel, P-M. and Bautz, E.K.F. (1983) EMBO J. 2, 705-710.
35. Arrigo, A.-P., Darlix, J.-L., Khandjian, E.W., Simon, M. and Spahr, P.-F. (1985) EMBO J. 4, 399-406.
36. Schuldt, C. and Kloetzel, P.-M. (1985) Develop. Biol. 110, 65-74.
37. Green, M.R., Maniatis, T. and Melton, D.A. (1983) Cell 32, 681-694.
38. Käufer, N.F., Simanis, V. and Nurse, P. (1985) Nature 318, 78-80.
39. Beggs, J.D., van den Berg, J., van Ooyen, A. and Weissmann, C. (1980) Nature 283, 835-840.
40. Langford, C., Nellen, W., Niessing, J. and Gallwitz, D. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 1496-1500.
41. Watts, F., Castle, C. and Beggs, J. (1983) EMBO J. 2, 2085-2091.
42. Langford, C.J., Klinz, F.J., Donath, C. and Gallwitz, D. (1984) Cell 36, 645-653.
43. Langford, C.J. and Gallwitz, D. (1983) Cell 33, 519-527.
44. Pikielny, C.W., Teem, J.L. and Rosbash, M. (1983) Cell 34, 395-403.
45. Hickey, E., Brandon, S.E., Potter, R., Stein, G., Stein, J. and Weber, L. (1986) Nucleic Acids Res. 10, 4127-4145.
46. de Wet, J.R., Wood, K.V., DeLuca, M., Helinski, D.R. and Subramani, S. (1987) Mol. Cell. Biol. 7, 725-737.
47. Fu, N.-Y. and Manley, J.L. (1987) Mol. Cell. Biol. 7, 738-748.