J Mol Evol (1995) 41:414–420

JOURNAL OF MOLECULAR EVOLUTION

A *Drosophila hsp70* Gene Contains Long, Antiparallel, Coupled Open Reading Frames (LAC ORFs) Conserved in Homologous Loci

Irene Konstantopoulou,[1] Christos A. Ouzounis,[2] Elena Drosopoulou,[1] Minas Yiangou,[1] Paschalis Sideras,[3] Chris Sander,[2] Zacharias G. Scouras[1]

[1] Department of Genetics, Development and Molecular Biology, School of Biology, Aristotle University of Thessaloniki (AUTH), GR-54006 Thessaloniki, Greece
[2] Protein Design Group, European Molecular Biology Laboratory, D-69012 Heidelberg, Germany
[3] Unit for Applied Cell and Molecular Biology, University of Umea, S-90187 Umea, Sweden

Received: 22 November 1994 / Accepted: 3 February 1995

Abstract. A clone isolated from a *Drosophila auraria* heat-shock cDNA library presents two long, antiparallel, coupled (LAC) open reading frames (ORFs). One strand ORF is 1,929 nucleotides long and exhibits great identity (87.5% at the nucleotide level and 94% at the amino acid level) with the *hsp70* gene copies of *D. melanogaster*, while the second strand ORF, in antiparallel in-frame register arrangement, is 1,839 nucleotides long and exhibits 32% identity with a putative, recently identified, NAD^+-dependent glutamate dehydrogenase (NAD^+-GDH). The overlap of the two ORFs is 1,824 nucleotides long. Computational analysis shows that this LAC ORF arrangement is conserved in other *hsp70* loci in a wide range of organisms, raising questions about possible evolutionary benefits of such a peculiar genomic organization.

Key words: Gene structure — Heat shock — *hsp70* — Antiparallel ORFs — *Drosophila*

Introduction

Overlapping genes in general, and open reading frames (ORFs) in particular, sharing the same DNA sequence were until recently considered to be rare in eukaryotes. Almost all of the known cases so far involve short coding-region overlaps, presence of ORFs in introns, and 3′ untranslated regions (UTRs) (Henikoff et al. 1986; Spencer et al. 1986; Adelman et al. 1987; Thompson-Jager and Domdey 1990; Merino et al. 1994; Boles and Zimmermann 1994). Recently, a peculiar arrangement of long antiparallel overlapping coding regions producing mRNA transcripts for a heat-shock 70 cognate gene (*hsc70*) and putative NAD^+-dependent glutamate dehydrogenase (NAD^+-GDH) was reported in a freshwater mold, *Achlya klebsiana* (LeJohn et al. 1994).

hsp70 is the most prominent member of the Hsp70 gene family, coding for the major heat-shock protein *hsp70*, one of the most-conserved proteins throughout evolution: human *hsp70* is 73% identical to *D. melanogaster hsp70* and 47% identical to *E. coli DnaK* (Boorstein et al. 1994). *hsp70* is known to act as a molecular chaperone and to play a major role in thermotolerance (Parsell and Lindquist 1993). Other family members in *Drosophila* are *hsp68*, a single-copy gene, and the *hsc70* (cognate) genes, which, however, exhibit constitutive and not heat-inducible expression (Lindquist and Craig 1988).

Among the *Drosophila* species studied so far, *D. auraria*, a *montium* subgroup species of the *melanogaster* species group (Lemeunier et al. 1986), poses some interesting features—namely, a large number of reverse tandem chromosomal duplications in which two well-

Correspondence to: Z.G. Scouras

formed Balbiani rings, structurally similar to those of chironomids, are included (Scouras and Kastritsis 1984; Kastritsis et al. 1986), and the clustered arrangement of the β-tubulin genes (Scouras et al. 1994). Here, we provide evidence that in a double-stranded DNA derived from the *hsp70* gene locus of *D. auraria,* a remarkably long overlap of antiparallel coupled (LAC) ORFs does exist, involving an *hsp70* gene copy. After computational analysis, we show that the same gene arrangement also extends in other homologous loci in a wide range of organisms.

Materials and Methods

Drosophila *Strains. D. auraria,* strain 17, isolated by sib mating from an original stock, No. 3040.11b, from the collection of the University of Texas (Scouras and Kastritsis 1984; Kastritsis et al. 1986; Scouras et al. 1986), and *D. melanogaster* (Canton S, C/S) were used in the present study. The stocks were maintained in 16-h light, 8-h dark, at a temperature of 21 ± 1 °C, on yeast glucose medium.

Construction of a Heat-Shock cDNA Library. Third-instar larvae of *D. auraria* were heat shocked (37 °C, 40 min); total RNA was isolated by the guanidinium/phenol/chloroform method (Chomczynski and Sacchi 1987). Poly(A)$^+$RNA was purified by the oligo(dT)-cellulose method (Aviv and Leder 1972). Poly(A)$^+$RNA was converted to cDNA using the Amersham cDNA synthesis kit following the manufacturer's instructions. Double-stranded cDNAs were end-modified with *Eco*RI adapters, cloned into λZAP II arms (Stratagene, LaJolla, CA), and in vitro packaged using the Promega Lambda DNA packaging system (Promega, Madison WI). The cDNA library contained 3×10^5 independent recombinant λ phage clones.

DNA Probes. Clone 56H8 (Shedl et al. 1978) containing one *hsp70* gene copy from locus 87A, and clone pDM300 (McGarry and Lindquist 1985), containing two copies from *hsp70* locus 87C of *D. melanogaster,* were used in the present study. In addition, several clones isolated from a heat-shock cDNA library of *D. auraria* bearing *hsp70*-specific sequences were also used. (See below.)

Clone Isolation and Characterization. hsp70 clones were isolated from a heat-shock cDNA library of *D. auraria.* Screening of the cDNA library was performed using clone 56H8. All plaque and Southern hybridizations were performed at 65 °C according to Church and Gilbert (1984). Probe was labeled with ^{32}P-dCTP as precursor (Dupont Radiochemicals) using the random priming method (Feinberg and Vogelstein 1983). Isolated clones were in vivo excised from λ ZAP II vector to pBluescript plasmid following the instructions by Stratagene. Several restriction endonucleases, single, double, complete and partial digestions, as well as Southern hybridizations were used for clone mapping (Pissios and Scouras 1993; Scouras et al. 1994). Reactions were carried out according to manufacturers (Boehringer-Mannheim/New England Biolabs, Inc.). Autoradiography was performed using Fuji RX-100 film and Kodak D-19 developer.

DNA Sequencing Analysis. Sublones and short synthetic primers were used for nucleotide sequencing which was performed according to the Sanger et al. (1977) method using the Pharmacia LKB T7 DNA polymerase sequencing kit with ^{35}S-dATP as precursor.

In Situ Hybridizations. In situ hybridizations on the salivary gland polytene chromosomes were performed at 65 °C as previously described (Scouras et al. 1994) using as probes both the *D. melanogaster* and the *D. auraria hsp70*-specific sequences.

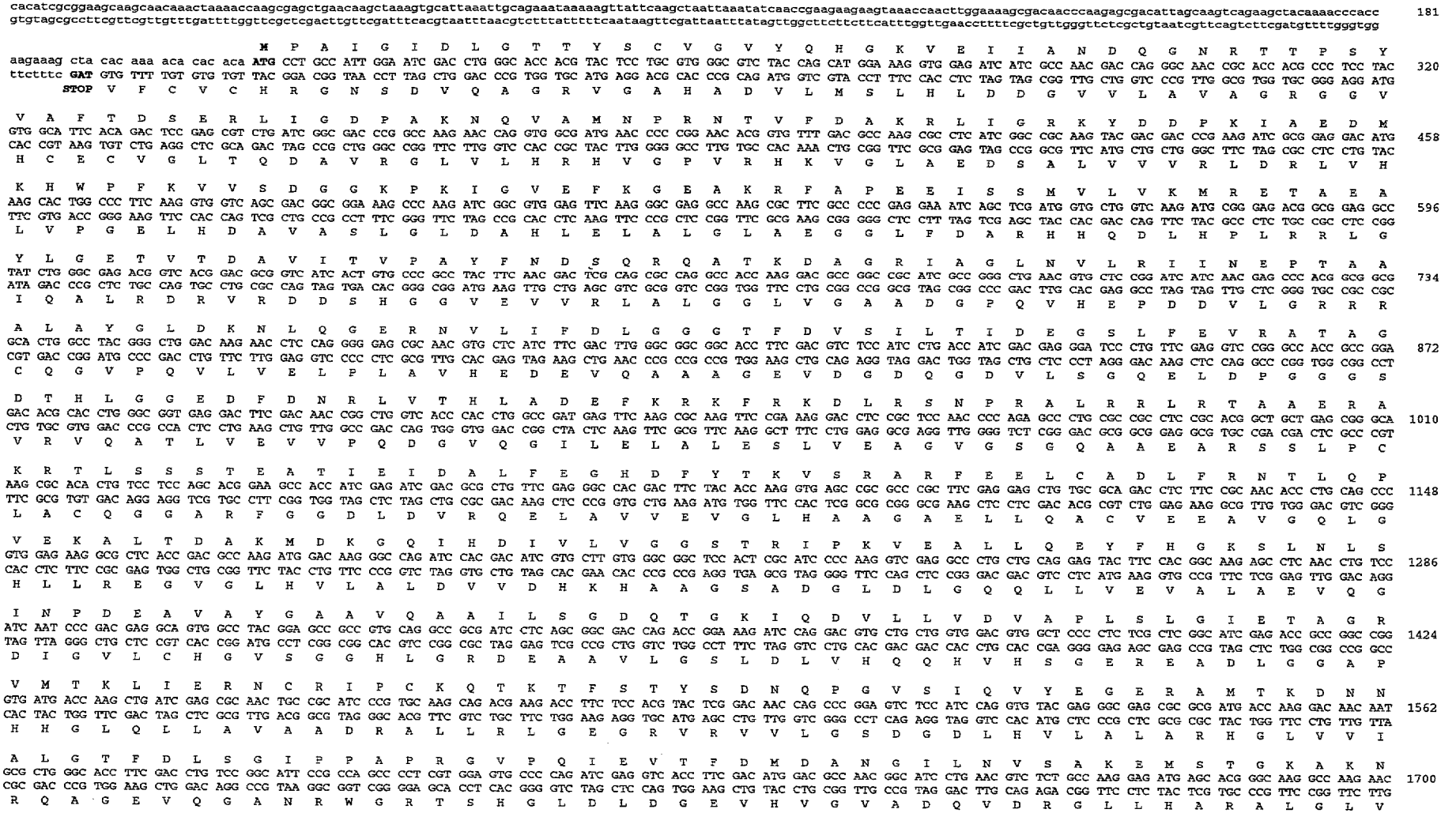
In Vitro Translation Experiment. The cDNA clone pda17-70.1 was used as template in an in vitro translation experiment which was performed using the Promega in vitro translation kit with T3 and T7 polymerases according to manufacturer's instructions.

Tree Production. Tree produced by CLUSTALW (Higgins et al. 1992) using PHYLIP (Felsenstein 1988) and drawn by TREETOOL (copyright U. Illinois). The names on the tree are as in SWISS-PROT database (Bairoch and Boeckmann 1992) and 3D is the bovine cognate protein for which the structure of the *N*-terminal domain has been solved (Flaherty et al. 1990).

Results and Discussion

In the course of examining the heat-shock response of *D. auraria,* a number of clones were isolated and characterized, after restriction endonuclease analysis, from a heat-shock cDNA library. The longer clone, pda17-70.1, was subjected to sequencing and found to be 2,351 base pairs (bp) long (Fig. 1). The one strand of pda17-70.1 contains an ORF 1929 nucleotides (nt) long (positions 207–2135, predicted amino acid sequence length of 643), which strongly resembles the *hsp70* gene of *D. melanogaster* (polytene band 87C1), exhibiting 87.3% identity at the nucleotide level and 94.3% identity of the predicted amino acid sequences. Most of the differences at the nucleotide level represent silent substitutions, while approximately 42 changes are both single- and double-base mutations resulting in 34 amino acid changes. Clone pda17-70.1 strongly hybridizes in situ with a band at region 32A (2L chromosome arm) of *D. auraria,* where also clones 56H8 and clone pDM300 of *D. melanogaster* hybridize (data not shown). Locus 32A has previously been found to be the most prominent heat-shock locus of *D. auraria,* both in terms of puff size as well as transcription induction level (Scouras et al. 1986). Dendrograms derived from multiple sequence alignments for members of this family also strongly suggest that the present gene from *D. auraria* most closely resembles the corresponding heat-inducible forms of *D. melanogaster* (Fig. 2).

Surprisingly, the complementary strand (c-strand) of the same clone contains a large ORF 1,839 nt long (complement positions 188–2030, predicted amino acid sequence length of 613) (Fig. 1). The first codon (ATG) of this anti-*hsp70* ORF is located opposite codon 607 of the *hsp70* ORF (in-frame register), while the stop codon (complement positions 188/190) is found at −18 nt from the beginning of the *hsp70* ORF (Fig. 1). Thus, clone pda17-70.1 exhibits an 1,824-nt overlap (608 aa, positions 207–2030) of the two long antiparallel ORFs found at the complementary strands. Recently, LeJohn et al. (1994) showed that exon X of the gene for a putative NAD$^+$-dependent glutamate dehydrogenase (NAD$^+$-



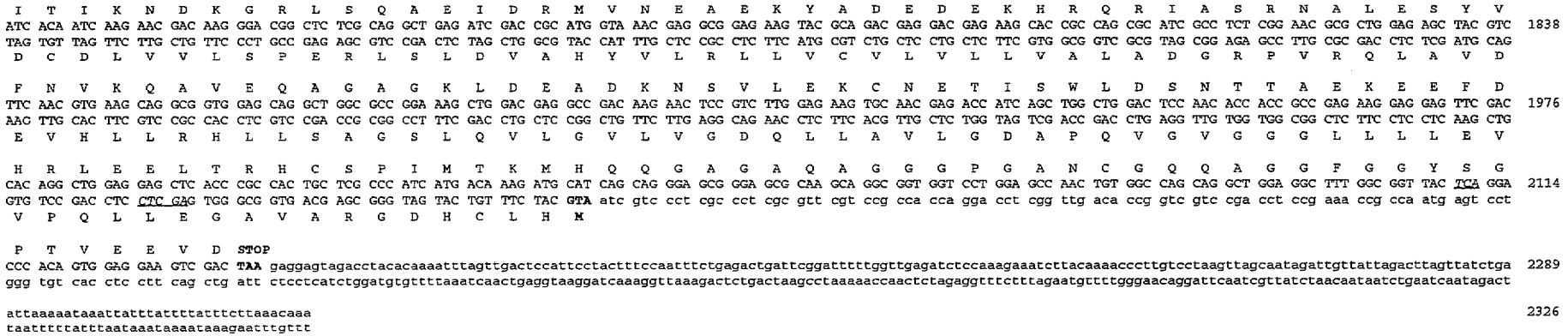


Fig. 1. DNA and amino acid sequences of both strands of the cDNA clone pda17-70.1. *Uppercase letters* signify predicted translated region; *lowercase letters* stand for untranslated regions (in the case of the anti-*hsp70* ORF, possibly intronic sequences). *Bold letters* identify both the start and stop codons of the two ORFs. The acceptor site plus the first codon coding for leucine (AG CTC, complement positions 1989–1993) and the forbidden codon outside the overlap (TCA, in antisense TGA: stop codon, complement positions 2109–2111) are *underlined.* The last 25 nt coding for the poly-A tail is not shown. The nucleotide sequence data reported in this paper will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X78403.

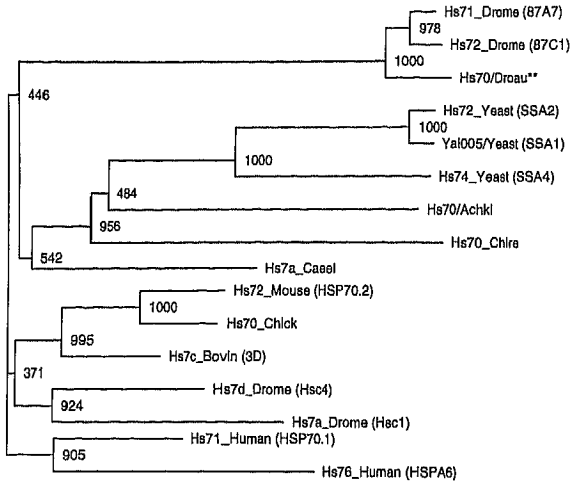


Fig. 2. Dendrogram of protein sequences for representative members of the Hsp70 family which includes the newly sequenced *hsp70* from *D. auraria* (marked with two *asterisks*). Bootstrap values are shown for corresponding branches, total number of bootstrap operations being 1,000. Obviously, the *D. auraria* sequence is the orthologue of the corresponding protein in *D. melanogaster.* Names are as in SWISS-PROT database and references therein; common names in *brackets.* Two recently sequenced members (*Achlya hsc70* and yeast *SSA1,* extensively discussed in the text) are given provisional names. The cognate *hsp70* proteins (constitutively expressed) are mainly located at the lower part of the tree (e.g., *Hs7a, Hs7d D. melanogaster*). 3D is the bovine cognate protein.

GDH) is found in a similar antiparallel arrangement with the *hsc70* (cognate) gene in *Achlya klebsiana,* a freshwater mold (Oomycota). After database searches (Altschul et al. 1990; Bork et al. 1992) of the anti-*hsp70* ORF from *D. auraria,* we have found that it bears 32% amino acid identity over 625 residues with exon X of the NAD^+-GDH (here called anti-*hsp70* protein) of *A. klebsiana.* The fact that the c-strand ORF in *D. auraria* is found in a similar arrangement to exon X of the anti-*hsp70* gene in *Achlya* (*D. auraria* 1857 nt vs. *A. klebsiana* 1866 nt) supports the possibility that it represents (part of) a functional gene. Yet, whether the anti-*hsp70* protein represents a NAD^+-GDH remains doubtful for the following two reasons: first, it seems that the *A. klebsiana* RNA transcript has not been directly associated with the expected GDH activity; and second, this anti-*hsp70* gene is less than 20% identical to the *Saccharomyces cerevisiae* and *Neurospora crassa* NAD^+-GDH proteins which share over 40% identity. It should be noted that the fungal NAD^+-GDH protein family is distinct from the GDH gene of *D. melanogaster* (Papadopoulou and Louis 1990), whose sequence (C. Louis, personal communication) is homologous to the GLF dehydrogenases (Bairoch 1992; Benachenhou-Lahfa et al. 1993).

The presence of a start codon at position +2027 in *D. auraria* anti-*hsp70* ORF may not represent a translation initiation site, because based on the similarity to the anti-*hsp70* gene from *A. klebsiana,* the exon boundary ap-

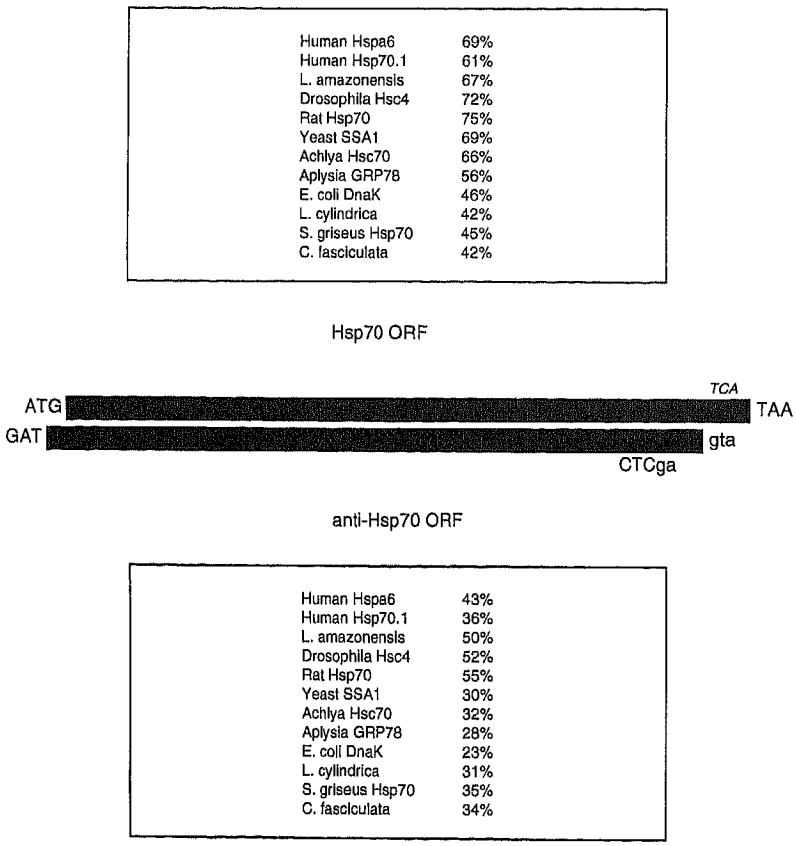


Fig. 3. The LAC ORFs in *D. auraria* with the main features highlighted and sequence similarity levels with other similar cases found. The *upper bar* represents the *hsp70* gene (sequence in the sense mode, read from left to right): ATG is start, *TCA* is the forbidden codon, TAA is the stop codon. The *lower bar* represents the anti-*hsp70* ORF (sequence in the antisense mode, read from right to left): atg is start (but predicted not to be translated), agCTC is the acceptor site plus the first codon coding for leucine, and TAG is the stop codon. Percent identity values of the *D. auraria* proteins to corresponding homologous proteins (for which an intact ORF is present at the complementary strand of *hsp70* genes) are shown at the top and bottom of the two ORFs. An alignment of the anti-*hsp70* family is available upon request (ouzounis@embl-heidelberg.de). Entries have the following accession numbers: human *Hspa6:* S78631, human *Hsp70.1:* X51758, *L. amazonensis:* L14605, *D. melanogaster Hsc4:* L01500, rat *Hsp70:* Z27118, yeast SSA1: L22015, *Achlya Hsc70:* U02504, *Aplysia* GRP78: Z15041, *E. coli DnaK:* D10765, *L. cylindrica:* S40718, *S. griseus:* D14499, *C. fasciculata:* M95682.

pears to be at complement position +1994. Thus, the acceptor site of the predicted exon X of NAD^+-GDH (the anti-*hsp70* ORF) in *D. auraria,* based on the similarity to the *Achlya* sequence, should be located at complement positions +1995 onward. Indeed, a consensus acceptor site (AG dinucleotide) is found prior to the exon boundary at complement positions +1994/1995 (Fig. 1). In vitro translation of the *D. auraria* anti-*hsp70* ORF produces a protein product of the expected size (67 kD, data not shown).

To investigate the generality of this phenomenon in the same genus, we examined the *D. melanogaster hsp70* gene-copy complementary strands for a possible ORF, but in spite of the high identity with *D. auraria,* four stop codons are found along these sequences. The only *D. melanogaster* Hsp70 gene containing a potential coding sequence for the anti-*hsp70* protein along the entire complementary strand is *Hsc4,* the only cognate gene that lacks introns (Perkins et al. 1990). The conservation of this LAC ORF arrangement suggests that at least one anti-*hsp70* gene may also be functional in this *Drosophila* species.

Another homologous ORF, 215 amino acids long, named YAL004, located on chromosome I of *S. cerevisiae* (Clark et al. 1994), was also identified in an antiparallel arrangement with the heat-shock gene SSA1. This ORF is 54% identical over its length to the predicted anti-*hsp70* protein of *A. klebsiana* and 30% identical to the *D. auraria* ORF (Fig. 3). However, its short length suggests that it is either part of a possibly larger gene, as in *A. klebsiana,* or a shorter, independently expressed gene. The upstream regions of the *Achlya* gene are apparently unique to that species, and not present in other yeast genera. Therefore, the predicted amino acid sequence for the nine exons of *A. klebsiana* NAD^+-GDH remains singular and may not code for the expected protein, yet it is transcribed into mRNA and presumably translated (LeJohn et al. 1994).

Although the ORFs for *hsp70* in *D. auraria* and *Hsc4* in *D. melanogaster* are 76% identical at nucleotide and 72% identical at amino acid levels, the antisense ORFs potentially coding for the anti-*hsp70* ORF are only 52% identical at the amino acid level (Fig. 3). Given that the fungal sequences display the same level of sequence similarity, this implies that the newly defined anti-*hsp70* class has fewer functional requirements than the *hsp70* proteins themselves and exhibits a faster rate of evolutionary change (Fig. 3). In theory, this may be expected, because changes in the third codon position (wobble nucleotide for many codons) of *hsp70* will result in a first codon change for the antiparallel gene. The unusual codon usage for the *D. auraria* anti-*hsp70* ORF with respect to the *Drosophila* codon usage pattern (D^2 is 5.93 vs *D. melanogaster Hsc4* being 0.41 and *D. auraria hsp70* being 0.90) (Grantham et al. 1981) is indicative of these structural constraints for antisense coding.

Thus, all available evidence, including the previous identification of a characterized transcript, the additional sequencing of a homologous short ORF in *S. cerevisiae,* the presence of a predicted acceptor site at the expected position in two *Drosophila* species, the location of a single forbidden codon (complement position +2110/2112) outside the overlap of *D. auraria* (Figs. 1, 3), and the conservation of at least one anti-*hsp70* ORF in a wide range of organisms (Fig. 3; C.A. Ouzounis, Z.G. Scouras, and C. Sander, in preparation), suggests the possibility that this c-strand-coded ORF represents (part of) a functional gene. If this is true, then together with the *A. klebsiana* finding, two experimental observations now exist, reported for distantly related organisms, of such a rare gene arrangement involving a gene of possibly unknown function and different members of the Hsp70 family. This remarkable gene arrangement calls for more efforts toward a deeper understanding of antiparallel genes, and of their function and evolution.

Acknowledgments. We thank F.C. Kafatos and M. Ashburner for reading the manuscript and making valuable suggestions, E. Nikolakaki for her help in various phases of this work, D. Papadopoulou and C. Louis for sharing unpublished data, and S. Artvanis-Tsakonas for feedback and comments. Parts of this work have been presented at the EMBO Workshop on *Drosophila,* held in Kolymbari, Crete (Greece), during June 1994. Some of the data in this report are part of a doctoral thesis to be submitted to the School of Biology of the Aristotole University of Thessaloniki by I.K. The work was supported by a grant from the Greek Agency of Research and Technology to Z.G.S.

References

Adelman JP, Bond CT, Douglass J, Herbert E (1987) Two mammalian genes transcribed from opposite strands of the same DNA locus. Science 235:1514–1517

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Aviv H, Leder P (1972) Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc Natl Acad Sci USA 69:1408–1412

Bairoch A (1992) PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acids Res 20:2013–2018

Bairoch A, Boeckmann B (1992) The SWISS-PROT protein sequence data bank. Nucleic Acids Res 20:2019–2022

Benachenhou-Lahfa N, Forterre P, Labedan B (1993) Evolution of glutamate dehydrogenase genes: evidence for two paralogous protein families and unusual branching patterns of the archaebacteria in the universal tree of life. J Mol Evol 36:335–346

Boles E, Zimmermann FK (1994) Open reading frames in the antisense strands of genes coding for glycolytic enzymes in *Saccharomyces cerevisiae.* Mol Gen Genet 243:363–368

Boorstein RW, Ziegelhoffer T, Craig EA (1994) Molecular evolution of the HSP70 multigene family. J Mol Evol 38:1–17

Bork P, Ouzounis C, Sander C, Scharf M, Schneider R, Sonnhammer E (1992) Comprehensive sequence analysis of the 182 predicted open reading frames of yeast chromosome III. Protein Sci 1:1677–1690

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium cyanate-phenol-chlorophorm extraction. Anal Biochem 162:156–159

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci USA 81:1991–1995

Clark MW, Keng T, Storms RK, Zhong W, Fortin N, Zeng B, Delaney S, Ouellette FBF, Barton AB, Kaback DB, Bussey H (1994) Sequencing of chromosome I of *Saccharomyces cerevisiae:* analysis of the 42 Kbp SPO7-CENI-CDC15 region. Unpublished

Feinberg AP, Fogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Felsenstein J (1988) PHYLIP: phylogeny inference package. Cladistics 5:355–356

Flaherty KM, De Luca-Flaherty CR, McKay DB (1990) Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. Nature 346:623–628

Grantham R, Gautier C, Gouy M, Jacobzone M, Mercier R (1981) Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res 9:r43–r74

Henikoff S, Keene MA, Fechtel K, Fristrom JW (1986) Gene within a gene: nested *Drosophila* genes encode unrelated proteins on opposite DNA strands. Cell 44:33–42

Higgins DG, Bleasby AJ, Fuchs R (1992) CLUSTAL V: improved software for multiple sequence alignment. Comput Appl Biosci 8:189–191

Kastritsis CD, Scouras ZG, Ashburner M (1986) Duplications in the polytene chromosomes of *Drosophila auraria.* Chromosoma 93: 381–385

LeJohn HB, Cameron LE, Yang B, Rennie SL (1994) Molecular characterization of an NAD-specific glutamate dehydrogenase gene inducible by L-glutamine: antisense gene pair arrangement with L-glutamine-inducible heat shock 70-like protein gene. J Biol Chem 269:4523–4531

Lemeunier F, David JR, Tsacas L, Ashburner M (1986) The *melanogaster* species group. In: Ashburner M, Carson HL, Thompson J (eds) The genetics and biology of *Drosophila.* Academic Press, London. vol 3, pp 147–256

Lindquist S, Craig EA (1988) The heat shock proteins. Annu Rev Genet 22:631–677

Merino E, Balbas P, Puente JL, Bolivar F (1994) Antisense overlapping open reading frames in genes from bacteria to humans. Nucleic Acid Res 22:1903–1908

Papadopoulou D, Louis C (1990) The gene coding for glutamate dehydrogenase in *Drosophila melanogaster.* Biochem Genet 28:337–346

Parsell DA, Lindquist S (1993) The function of heat shock proteins in stress tolerance: degradation and reactivation of damaged proteins. Annu Rev Genet 27:437–496

Perkins LA, Doctor JS, Zhang K, Stinson L, Perrimon N, Craig EA (1990) Molecular and developmental characterization of the heat shock cognate 4 gene of *Drosophila melanogaster.* Mol Cell Biol 10:3232–3238

Pissios P, Scouras ZG (1993) Mitochondrial DNA evolution in the *montium* species subgroup of *Drosophila.* Mol Biol Evol 10:375–382

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schedl P, Artavanis-Tsakonas S, Steward R, Gehring WJ, Mirault M-E, Goldschmidt-Clermont M, Moran L, Tissieres A (1978) Two hybrid plasmids with *Drosophila melanogaster* DNA sequences complementary to mRNA coding for the major heat shock protein. Cell 14:921–929

Scouras ZG, Kastritsis (CD) 1984) Balbiani rings and puffs of the polytene chromosomes of *Drosophila auraria.* Chromosoma 89: 96–106

Scouras ZG, Karamanlidou GA, Kastritsis CD (1986) The influence of heat shock on the puffing pattern of *Drosophila auraria* polytene chromosomes. Genetica 69:13–218

Scouras ZG, Milioni D, Yiangou M, Duchene M, Domdey H (1994) The β-tubulin genes of *Drosophila auraria* are arranged in a cluster. Curr Genet 25:84–87

Spencer CA, Gietz RD, Hodgetts RB (1986) Overlapping transcription units in the dopa decarboxylase region of *Drosophila.* Nature 322: 279–281

Thompson-Jäger S, Domdey H (1990) The intron of the yeast actin gene contains the promoter for the antisense RNA. Curr Genet 17:269–273