A cluster of diagnostic Hsp68 amino acid sites that are identified in *Drosophila* from the *melanogaster* species group are concentrated around β-sheet residues involved with substrate binding

Mark Kellett and Stephen W. McKechnie

Abstract: The coding region of the *hsp68* gene has been amplified, cloned, and sequenced from 10 *Drosophila* species, 5 from the *melanogaster* subgroup and 5 from the *montium* subgroup. When the predicted amino acid sequences are compared with available Hsp70 sequences, patterns of conservation suggest that the C-terminal region should be subdivided according to predominant secondary structure. Conservation levels between Hsp68 and Hsp70 proteins were high in the N-terminal ATPase and adjacent β-sheet domains, medium in the α-helix domain, and low in the C-terminal mobile domain (78%, 72%, 41%, and 21% identity, respectively). A number of amino acid sites were found to be "diagnostic" for Hsp68 (28 of ~635 residues). A few of these occur in the ATPase domain (385 residues) but most (75%) are concentrated in the β-sheet and α-helix domains (34% of the protein) with none in the short mobile domain. Five of the diagnostic sites in the β-sheet domain are clustered around, but not coincident with, functional sites known to be involved in substrate binding. Nearly all of the Hsp70 family length variation occurs in the mobile domain. Within *montium* subgroup species, 2 nearly identical *hsp68* PCR products that differed in length are either different alleles or products of an ancestral *hsp68* duplication.

Key words: Hsp70, Hsp68, diagnostic sites, *Drosophila melanogaster*, *montium* subgroup.

Résumé : Les auteurs ont amplifié, cloné et séquencé la région codante du gène *hsp68* chez 10 espèces du genre *Drosophila*, 5 de celles-ci appartenant au sous-groupe *melanogaster* et les 5 autres au groupe *montium*. Lorsque les séquences prédites d'acides aminés sont comparées aux séquences des Hsp70, la distribution des motifs conservés suggère que la région C-terminale serait subdivisée en fonction de la structure secondaire prédominante. L'homologie entre les protéines Hsp68 et Hsp70 est élevée (78 % et 72 %, respectivement) dans le domaine N-terminal contenant l'activité ATPase et le domaine adjacent en feuillet β. L'homologie est moyenne (41 %) dans le domaine en hélice α et faible (21 %) au sein de la région C-terminale. Plusieurs acides aminés sont « diagnostiques » des Hsp68 (28 des ~635 résidus). Quelques-uns de ceux-ci se situent dans le domaine ATPase (385 résidus), mais la majorité (75 %) sont concentrés au sein des domaines en feuillet β et en hélice α (34 % de la protéine) et aucun ne se trouve dans le petit domaine mobile. Cinq des sites diagnostiques au sein du domaine en feuillet β sont groupés autour, sans coïncider cependant, avec des sites fonctionnels connus comme étant impliqués dans la liaison du substrat. Presque toute la variation en matière de longueur des protéines au sein de la famille des Hsp70 survient dans le domaine mobile. Au sein des espèces du sous-groupe *montium*, l'amplification PCR produit deux amplicons *hsp68* presque identiques mais dont la taille est différente. Ces deux produits constituent soit deux allèles ou encore les produits d'une duplication génique ancestrale.

Mots clés : Hsp70, Hsp68, sites diagnostiques, *Drosophila melanogaster*, sous-groupe *montium*.

[Traduit par la Rédaction]

Received 12 July 2004. Accepted 9 November 2004. Published on the NRC Research Press Web site at http://genome.nrc.ca on 18 March 2005.

Corresponding Editor: J.B. Bell.

M. Kellett and S.W. McKechnie.[1] Centre for Environmental Stress and Adaptation Research, School of Biological Sciences, Monash University, Victoria 3800, Australia.

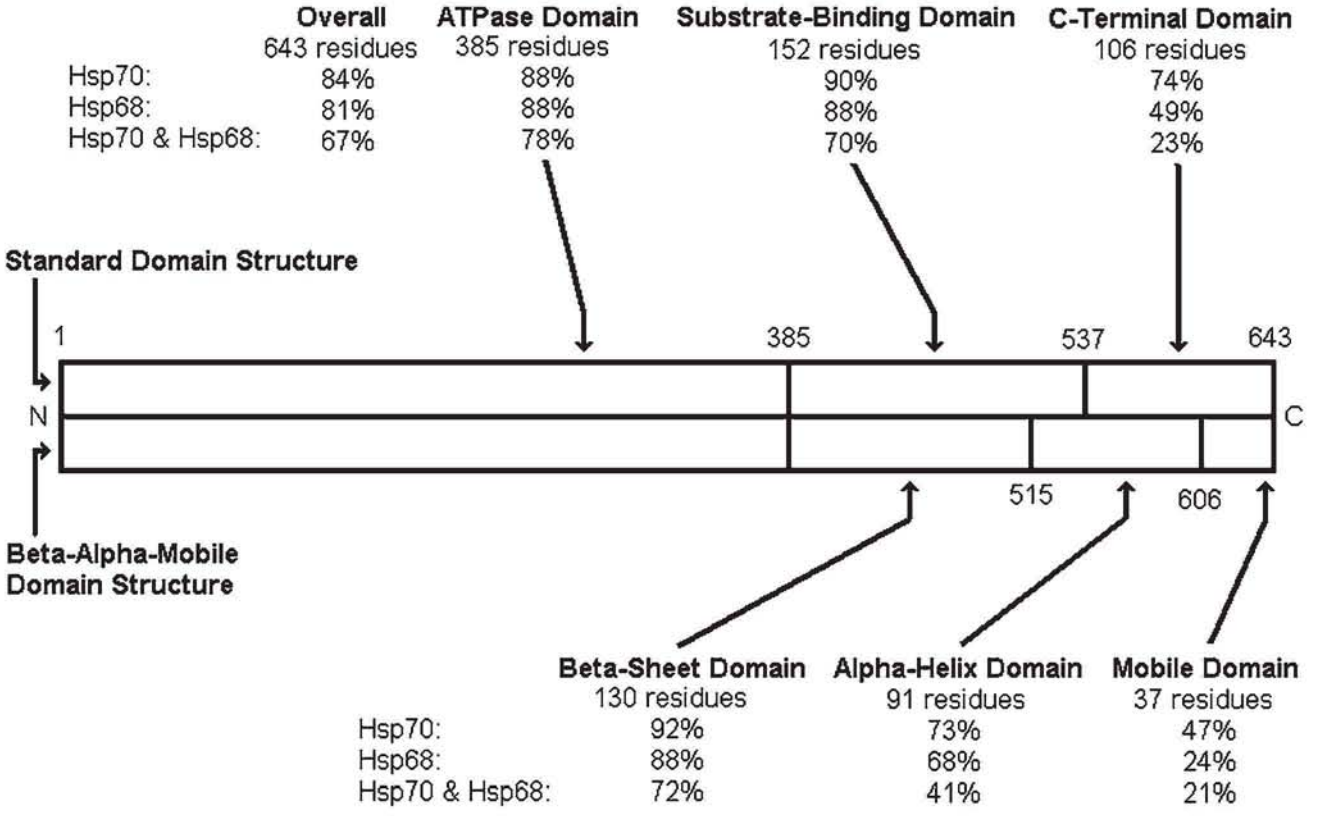
[1]Corresponding author (e-mail: stephen.mckechnie@sci.monash.edu.au).

Introduction

At least one copy of the *Hsp70* gene family has been found in every prokaryote or eukaryote tested (Gupta and Golding 1993). Members of this family encode a molecular chaperone, a protein that repairs denatured peptides, which is transiently expressed following cell damaging events such as extremes of temperature or exposure to toxins (Parsell and Lindquist 1993). The structure of this molecular chaperone is highly conserved, even between such disparate taxa as bacteria and vertebrates. At the N-terminal end of the peptide is the ATPase domain (Fig. 1), a large bilobed struc-

 doi: 10.1139/G04-113

Fig. 1. Domain structure and amino acid conservation of Hsp70 family proteins. Indicated are the standard domains and the beta–alpha–mobile domains described in this report. Listed below the domain names, from top to bottom, are the number of residues that make up the domain,% amino acid identity of the domain among 10 Hsp70 proteins,% amino acid identity among 12 Hsp68 proteins, and % amino acid identity between the Hsp70 and Hsp68 proteins.



ture (Flaherty et al. 1990). Distal to this is the substrate-binding domain, a structure that is made up of a groove for the binding of denatured peptides and an α-helical "latch" for regulating such binding (Zhu et al. 1996; Wang et al. 1998; Morhauser et al. 1999). At the extreme end is the C-terminal domain, a structure that has been shown to be non-essential for the protein to function as a molecular chaperone (Wang et al. 1993) but that is involved in inter- and intraprotein interactions (Bertelsen et al. 1999).

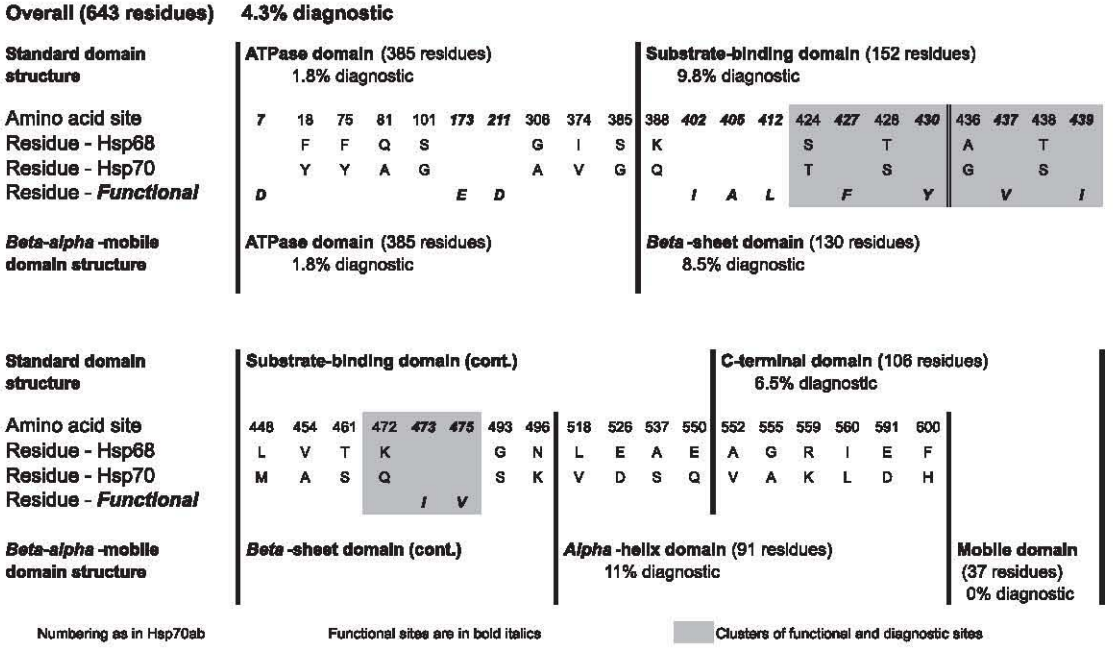
Spaced throughout each Hsp70 family protein are well-characterized amino acid residues that are involved with interactions with ATP and denatured protein substrate (Fig. 2). Three residues are hypothesized to be involved in interactions between the ATPase domain and ATP (Flaherty et al. 1990), while 9 residues in the substrate-binding domain have been determined by X-ray crystallography and NMR studies to be involved in binding to denatured peptides (Zhu et al. 1996; Wang et al. 1998; Morhauser et al. 1999).

Drosophila melanogaster has 2 distinct molecular chaperones of the Hsp70 family. The first, Hsp70, is encoded by 5 genes (Ish-Horowicz et al. 1979), reaches maximal protein level ~2 h after a cell-damaging event (Feder et al. 1997), and is turned over with a half-life of 1–2 h during a recovery period (Palter et al. 1986). In contrast, the second, Hsp68, is encoded by a single gene (Holmgren et al. 1981), reaches maximal protein level ~6 h after heat shock, and is maintained at this level at least 12 h into a recovery period (Palter et al. 1986). The genes that encode these proteins are believed to have evolved from a single ancestor by gene duplication (Bettencourt and Feder 2001), and their sequence has been confirmed by the *D. melanogaster* genome project (Kornberg and Krasnow 2000). There is now considerable evidence that Hsp70 plays a role in acquired thermotolerance, also called heat hardening (Feder and Krebs 1998; Michalak et al. 2001), and less, but still convincing, evidence that Hsp68 plays a similar role (McColl et al. 1996).

Many molecular and thermotolerance investigations into the evolution and function of Hsp70 and Hsp68 have used the model organism *D. melanogaster*. However, this is but one species in the closely related *melanogaster* species group, 9 of which are in the *melanogaster* subgroup and at least 60 of which are in the *montium* subgroup (Bock 1980). Examination of the Hsp70 family in *Drosophila vulcana*, a member of the *montium* subgroup, revealed that, on the basis of protein molecular masses, in situ Southern hybridizations, and chromosomal location, Hsp70 and Hsp68 were less differentiated in this species than they were in *D. melanogaster* (Pardali et al. 1996). Subsequent analysis suggested that this observation appeared to be general across the *montium* subgroup (Drosopolou et al. 1996).

In this study, the DNA sequence of the coding region of *hsp68* is determined in a number of the members of the *melanogaster* and *montium* subgroups. Comparisons of the encoded amino acid sequence of Hsp68 with those of Hsp70 allow us to identify unique features and amino acid sites of

Fig. 2. Location of Hsp70 family functional amino acid sites and Hsp68 diagnostic sites. Residue numbers are given relative to their position in Hsp70 Ab from *D. melanogaster* (accession No. J01103). Note that the numbers 7, 173, and 211 correspond to D10, E175, and D211, respectively, of Flaherty et al. (1990) and 402, 405, 412, 427, 430, 437, 439, 473, and 475 correspond to I401, M404, L411, F426, A429, V436, I438, I472, and V474, respectively, of Wang et al. (1998).



Hsp68 and to detect and speculate on the functional significance of interesting patterns of conservation throughout the protein. The underlying molecular causes of the observed differences in the structure of Hsp68 across taxa are also identified.

Materials and methods

Lines and maintenance

The following lines/species were used: *Drosophila simulans* (field collected from Melbourne, Victoria, Australia), *Drosophila sechellia* (stock No. 14021-0248.0), *Drosophila yakuba* (stock No. 14021-0261.0), *Drosophila teisseri* (stock 14021-0257.00), *Drosophila erecta* (stock No. 14021-0224.0), *Drosophila auraria* (stock No. 14028-0471.00), *D. vulcana* (stock No. 14028-0711.0), *Drosophila serrata* (field collected from Eacham, Queensland, Australia), and the cryptic species of *D. serrata*, hereinafter referred to as *Drosophila* species X (field collected from Lake Placid, Queensland, Australia; Schiffer et al. 2004). This species is currently being described by M. Schiffer and S. McEvery in a forthcoming paper. Forty pinned specimens (numbered 16474–16494 and 16537–16555) from the Lake Placid massbred culture have been lodged as types in the Australian Museum, Sydney, the Queensland Museum, Brisbane, and the Australian National Insect Collection, Canberra, and *Drosophila birchii* was field collected from Crystal Creek, Queensland, Australia. All numbered strains were obtained from the Tucson Species Center, Tuscon, Arizona. Culture and rearing of all lines were at constant temperature of 25 °C with 12 h light : 12 h dark cycles on potato–yeast–sugar–agar medium.

DNA extraction, PCR amplification, and cloning

Primers were generated by comparing *hsp68* of *D. melanogaster* (GenBank accession No. AF906275) with *hsp70* genes of *D. melanogaster* (GenBank accession Nos. JO1103 and JO1104) and *D. auraria* (GenBank accession No. X78403). To obtain *Hsp68* specificity, the reverse primer (5′-CTCGACGGTGGGACCCTT-3′) was placed overlapping into a sequence encoding the most C-terminal Hsp70-family conserved EEVD motif (Freeman et al. 1995), except for the 3′ *hsp68* unique sites. The forward primer (5′-ATGCCAGCCATTGGAATCGA-3′) was in a conserved N-terminal encoding sequence. Since primers covered terminal coding sequences, only 98% of each full protein coding sequence was included in the comparisons reported here.

Template DNA for the PCR was obtained by homogenizing single flies within 1.5-mL tubes containing 500 μL of a 0.5% aqueous solution of Biorad chelex 100 resin. After being incubated at 37 °C for 24 h and boiled for 8 min, the tube was centrifuged at 15 000 r/min for 3 min and the supernatant placed at –20 °C prior to use.

Immediately prior to amplification, 10 μL of template DNA solution was added to 20 μL of solution 1 (forward and reverse primers at 0.0125 μg/μL each and Promega dNTPs at 0.5 mmol/L each) and 20 μL of solution 2 (Roche 10× high-fidelity enzyme buffer diluted to 2.5× and Roche high-fidelity enzyme mixture at 2.5 U/μL) and capped with paraffin oil. Using an FTS-1 thermal cycler, the following PCR protocol was employed: an initial melt at 96 °C for 1 min, followed by 36 cycles of 94 °C for 15 s, 55 °C for 30 s, and 72 °C for 2 min, followed by a completion cycle at 72 °C for 5 min.

After being resolved on a 1% gel and visualized with ethidium bromide, PCR products of the expected size were

purified with the Promega Wizard Plus PCR Preps kit. The purified PCR product was cloned into the pGEM-T vector and the resulting construct amplified in Promega JM109 competent cells and purified with the Promega Wizard Plus Minipreps kit. The cloned PCR product was sequenced using an ABI PRISM dye-terminator cycle sequencing kit.

For each species examined in this study, an *hsp68* PCR product was independently amplified from 3 flies. From each PCR product, 3 *hsp68*/pGEM-T constructs were purified and sequenced. Deriving the sequence for each line/species from 9 clones amplified with an enzyme mix with proofreading capacity greatly reduced the probability of PCR errors influencing the final consensus sequence.

Sequence comparison

Characteristics of Hsp68 were determined using the Pepstats program from the ANGIS website (www.angis1.org.au), assuming that the 2% of *hsp68* left undetermined is the same sequence as *hsp68* from *D. melanogaster*. Protein sequence alignments were performed using the ClustalW v1.82 program from the EBI website (www.ebi.ac.uk).

The following published sequences were used in comparisons: *hsp70* Ab from *D. melanogaster* (accession No. J01103), *hsp70* Bb from *D. melanogaster* (accession No. J01104), *hsp70* Aa from *D. simulans* (accession No. AF295963), *hsp70* Ba from *D. simulans* (accession No. AF295971), *hsp70* Ab from *D. sechellia* (accession No. AF302416), *hsp70* Bb from *D. sechellia* (accession No. AF302420), *hsp70* Aa from *Drosophila orena* (accession No. AF302410), *hsp70* Ba from *D. orena* (accession No. AF302412), *hsp70* from *D. auraria* (accession No. X78403), and *hsp68* from *D. melanogaster* (accession No AF096273). *Drosophila pseudoobscura hsp70* (Contig3323_Contig6140) and *hsp68* (Contig7891_Contig7492A) were obtained by conducting BLAST searches of the *D. pseudoobscura* sequence database (www.hgsc.bcm.tmc.edu/projects/drosophila) using *hsp70* Bb and *hsp68* from *D. melanogaster*, respectively. Two copies of *hsp68* were identified in *D. pseudoobscura*, but one copy was disrupted by a break in the contig. Only the complete copy was used in this study.

Sequences derived in this study have the following accession numbers: *D. simulans hsp68*, AY365140; *D. sechellia hsp68*, AY369031; *D. yakuba hsp68*, AF277570; *D. teisseri hsp68*, AY369032; *D. erecta hsp68*, AF247554; *D. auraria hsp68*L, AY369033; *D. auraria hsp68*S, AF247553; *D. vulcana hsp68*, AF277569; *Drosophila* species X *hsp68*L, AY369034; *Drosophila* species X *hsp68*S, AY369035; *D. serrata hsp68*L, AF277568; *D. serrata hsp68*S, AY369037; *D. birchii hsp68*L, AY369036; *D. birchii hsp68*S, AY369038.

A cladogram was constructed using the Nona v1.0 program and the resulting tree was redrawn with the Winclada v0.9.99A (beta) program.

Results

Structure of Hsp68

In this study, 98% of the coding region of *hsp68* from 5 *melanogaster* subgroup species, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. teisseri*, and *D. erecta*, and 5 *montium* subgroup species, *D. auraria*, *D. vulcana*, *D. serrata*, *Drosophila* species X, and *D. birchii*, was cloned and sequenced. Discussion from here on relates largely to the structure of the predicted protein product of these sequences. We compare the structure and amino acid sequence of 10 available *Drosophila* Hsp70 proteins and 12 Hsp68 proteins (including *D. melanogaster* and *D. pseudoobscura*). In all, 28 diagnostic sites (amino acid residues that are conserved absolutely within all of the Hsp70 and Hsp68 considered in this report but not between them) were identified (Fig. 2). Diagnostic sites give an idea of what features distinguish one member of the Hsp70 family from the other and they made up 4.3% of the protein.

When the level of overall homology (percent amino acid identity) among members of the Hsp70 family are examined (Fig. 1), the protein is well conserved among both Hsp68 proteins (81% identity) and Hsp70 proteins (84% identity), although conservation decreases to 67% when the 2 types of Hsp70 family proteins are compared with each other. Higher, or lower, levels of conservation apply when specific domains are considered. The ATPase domain is highly conserved (88% identity) within both Hsp70 or Hsp68 groups alone, dropping in identity to 77% when these groups are compared with each other. The ATPase domain, just over half of the protein, has few diagnostic sites (1.8% of its 365-residue length). The substrate-binding domain is similarly highly conserved but has a high density of diagnostic sites (9.8% of its 152-residue length). The C-terminal domain is the least conserved, dropping to only 23% identity when the Hsp68 and Hsp70 groups are compared. However, it has an intermediate density of diagnostic sites (6.5% of its 106-residue length).

Different patterns of conservation become apparent if the domains of the protein are redefined (Fig. 1). The ATPase domain remains unchanged. The α-helical latch region of the substrate-binding domain is immediately adjacent to the α-helical region of the C-terminal domain. Here, we consider these 2 regions together as forming one continuous "α-helix domain", leaving the remainder of the substrate-binding domain, mostly composed of β-sheets, as the "β-sheet domain". The remaining distal end of the C-terminal domain, which is highly flexible, becomes the "mobile domain". Homology comparisons under this "beta–alpha–mobile" domain structure is of interest (Fig. 1). The β-sheet domain, comprising most of the standard substrate-binding domain, including the groove, remains highly conserved (although there is 72% identity between the Hsp68 and Hsp70 groups) and retains a relatively high density of diagnostic sites (8.5% of sites) (Fig. 2). The α-helix domain, although reasonably well conserved within either Hsp70 or Hsp68 proteins, is only loosely conserved between the Hsp68 and Hsp70 groups (41% identity). The α-helix domain appears to have the highest density of diagnostic sites in the protein (11% of its 91-residue length). The short mobile domain is poorly conserved (21% identity between Hsp70 and Hsp68) and has no diagnostic sites.

In no case was a diagnostic site at the same position as a site that had been previously determined to be involved in interactions between ATP and the ATPase domain (Flaherty et al. 1990) (Fig. 2). Nor did diagnostic sites occupy any sites involved with interactions between denatured peptides

Fig. 3. Amino acid sequence variation in the mobile domain across the Hsp70 family proteins. Numerals immediately following the species/gene code indicate the first depicted amino acid site number. Numerals on the right-hand side indicate the predicted amino acid length of the protein. Species codes: *mel*, *D. melanogaster*; *sim*, *D. simulans*; *mau*, *D. mauritiana*; *sech*, *D. sechcellia*; *yak*, *D. yakuba*; *tei*, *D. teisseri*; *ere*, *D. erecta*; *ore*, *D. orena*; *aur*, *D. auraria*; *vul*, *D. vulcana*; sppX, cryptic species of *D. serrata*; *ser*, *D. serrata*; *bir*, *D. birchii*. Specific genes are as indicated either 70 (Hsp70) or 68 (Hsp68), the latter being either the long (L) or short (S) form of the sequence. Residues that are identical to those found in the first sequence (Hsp70 Ab from *D. melanogaster*) are indicated by dots. Undetermined sequences are indicated with a question mark (?). EEVD indicates the C-terminal motif.

```
mel70Ab 606 TKMHQQ--GAGAG------AGGPGANCGQQAGGFG--GYSGRTVEEVD 643
mel70Bb 605 ......--....-------.................--....P...... 641
sim70Aa 605 ......GA....-----G-.................--....P...... 644
sim70Ba 605 ......GA...------A-.................--....P...... 643
mau70Ab 605 ......GA.....VGGPG-.................--....P...... 649
mau70Bb 605 ......GA...------A-.................--....P...... 643
ore70Aa 605 ......GA...------A-..-.............--....P...... 642
ore70Ba 605 ......AA...------A-..S.............--....P...... 643
aur70   605 ......GA..QAG-------.................--....P...... 643
pso70   605 ..I...GAA..---------PAA.............--.F..P...... 641
mel68   604 ....KGGG--DGQQAP---------.F.....-----..K.P...... 635
sim68   604 .....GGGG-DGQQAP---------.F.....-----..K???????? 636
sech68  604 .....GGGG-DGQQAP--NF.QQAP.F.....-----..K???????? 643
yak68   604 ....KGGG--DGQQAP---------.......-----..K???????? 635
tei68   604 ....KGGG--DGQQAP---------.......-----..K???????? 635
ere68   604 .....GGG--DGQQAP---------.......-----..K???????? 635
vul68   604 ......GGAG.D.QQG-------AP..........SG..K???????? 644
aur68L  604 ......GAAG.D.QHC-----------.......SG..K???????? 640
sppX68L 604 ......GAAG.DDHQG-------AT..........SG..K???????? 644
ser68L  604 ......GAAG.D.HQG-------AP..........SG..K???????? 644
bir68L  604 ......GAAG.D.HQG-------AP..........SG..K???????? 644
aur68S  604 ......GAAG.D.QQK-------AQ.------...----K???????? 634
sppX68S 604 ......GAAG.D.QQK-------AP.------...----K???????? 634
ser68S  604 ......GAAG.D.HQG-------AS.------...----K???????? 634
bir68S  604 ......GAAG.D.HQG-------AS.------...----K???????? 634
pso68   604 ......GGTGDGTQAP---------.......A.ANG..N.P..... 642
                                                    EEVD
```

and β-sheet or α-helix domains (Zhu et al. 1996; Wang et al. 1998; Morhauser et al. 1999). However, 5 of these diagnostic sites in the β-sheet domain are closely clustered around sites involved with interactions with denatured peptides, in 4 cases being immediately adjacent to them (Fig. 2).

There is some length variation among Hsp68 proteins across taxa. Assuming that the 2% of the gene that was not sequenced is conserved absolutely with *D. melanogaster*, the length of the encoded product ranges from 634 to 644 amino acid residues. This length variation is largely due to insertion/deletion sites in the terminal mobile domain (Fig. 3).

Long and short forms of Hsp68

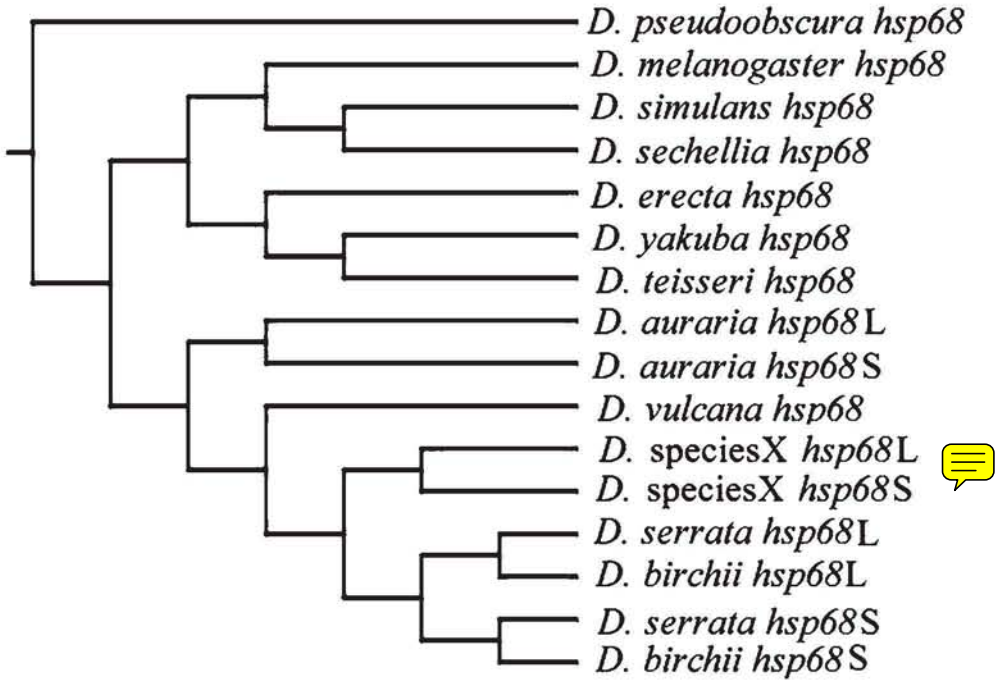
Two distinct PCR products were amplified from each member of the *montium* subgroup except *D. vulcana*. One was of a length similar to that of the PCR product from *hsp68* members of the *melanogaster* subgroup and will be referred to as the "short" form or *hsp68*S. The other was of a length comparable with that of *hsp70*, although its encoded product had greater similarity to Hsp68 and will be referred to as the "long" form or *hsp68*L. While in *D. vulcana*, only *hsp68*L was amplified, we cannot exclude the possibility that an Hsp68S occurs in this species, since only 3 clones from one strain were sequenced. When the 2 forms within each of 4 species are compared, they are virtually identical, the only consistent difference being 2 small insertions–deletions in the mobile domain (Fig. 3). In *D. auraria*, *hsp68*L appears to have undergone a subsequent deletion. A cladogram was generated from all of the available *hsp68* genes using *hsp68* from *D. pseudoobscura* as an outgroup. This cladogram was consistent with the recognized phylogenetic relationships noted for these taxa (Ashburner 1989) and placed *hsp68*L and *hsp68*S as sister taxa within all 4 species (Fig. 4).

Discussion

The well-characterised ATPase domain binds and hydrolyses ATP to drive the molecular chaperone process (Flaherty et al. 1990). This function might be expected to place considerable constraints upon evolution, and the high degree of conservation of this region and low density of diagnostic sites are consistent with this prediction.

Rather than the distal half of the molecule consisting of 2 domains (the central well-conserved domain and the C-terminal domain conserved at an intermediate level), this half of the protein is best envisaged as having 3 structure/function domains. These are the most N-terminal, well-conserved β-sheet domain, a central α-helix domain that is conserved at an intermediate level, and a short nonconserved C-terminal domain. Most of the Hsp68 diagnostic sites (75%

Fig. 4. Cladogram indicating the relationships of all *Drosophila hsp68* genes considered in this paper. Long and short forms of *hsp68* (see text) are indicated L and S, respectively.



of them) are densely and "equally" distributed in the adjacent β-sheet and α-helix domains (34% of the protein); none occur in the short terminal mobile domain.

The β-sheet domain (130 residues), involved in binding denatured peptide substrates in a groove, is conserved to a degree that is comparable with that of the ATPase domain but this region is rich in diagnostic sites. Although none of these sites overlap with a site determined to be involved in chaperone/denatured peptide binding (Zhu et al. 1996; Wang et al. 1998; Morhauser et al. 1999), they are concentrated in the region involved in this interaction, in some cases being immediately adjacent to such sites. In all cases but one, the substitutions are between residues of similar chemistry, although they have the potential to alter substrate specificity through van der Waal's interactions. In the case of residue 472, the change from the acidic glutamine in Hsp70 proteins to the basic residue lysine in Hsp68 proteins may have a greater potential to alter substrate specificity.

The shorter α-helix domain (91 residues) forms a rigid structure (Bertelsen et al. 1999) that acts physically by its positioning in the absence of ATP to prevent binding of denatured peptides (Wang et al. 1998; Morhauser et al. 1999). This region is less conserved than the β-sheet domain and markedly less conserved when the 2 members of the Hsp70 family are compared with each other; much of the divergence between Hsp70 and Hsp68 occurs in this short region (it has the highest density of diagnostic sites; Fig. 2). Since the predicted function of this domain appears to be structural rather than catalytic, its amino acid sequence may be under less constraint than the β-sheet in general; however, key functional differences between Hsp68 and Hsp70 associated physical structure may reside in this domain.

The short C-terminal mobile domain (37 residues) is extremely variable and uniformly rich in glycine and proline residues, which impart great flexibility to this section of the peptide (Bertelsen et al. 1999). No Hsp68 diagnostic sites occurred in this domain. The mobile domain has been hypothesized to be involved in inter- and intraprotein associations (Bertelsen et al. 1999) and the length of this flexible region may influence the energetics of such associations (Pontius 1993). The length and sequence variation in the mobile domain, both within and between species, have the potential to be subject to functional constraint.

The length variation detected between the long and short forms of Hsp68 within all species of the *montium* subgroup, save *D. vulcana*, occurred in the mobile domain. Although essentially identical to the long forms from the same species, the short forms shared 2 small deletions in the mobile domain (Fig. 3). The different forms are likely to be either different alleles of *hsp68*, representing a polymorphism that predates the species split, or the products of an early duplication of *hsp68* in this subgroup. Such duplications have been recorded in *hsp70* (Bettencourt and Feder 2001; Bettencourt and Feder 2002). If an early duplication occurred, we also need to postulate a number of gene conversion events because of the high levels of intraspecific homology between long and short forms (Fig. 4). This is a distinct possibility, since gene conversions are not uncommon if the duplications are in tandem (Bettencourt and Feder 2002). Intriguingly, *hsp68* seems to be tandemly duplicated in *D. pseudoobscura*, but one of the copies has been disrupted by a break in the contig. Analysis of the more distantly related *Drosophila virilis* favours the duplication hypothesis (Velikodvorskaia et al. 2005).

Our sequence data can help explain previous observations about *hsp68* homology levels and the size of the encoded protein across the *montium* subgroup. *Drosophila vulcana*'s Hsp68 was observed to have a molecular mass of 69 kDa,

larger than the 68 kDa reported for *D. melanogaster* (Pardali et al. 1996). Our data indicate that *D. vulcana*'s Hsp68 has a molecular mass of 70.3 kDa compared with 69.7 kDa for *D. melanogaster* and that much of this difference is accounted for by variation in the length of the mobile domain. Given the approximate nature of molecular mass determination on SDS–PAGE gels, these sequence data likely account for the observed size difference. The presence of 2 forms of *hsp68* in other *montium* subgroup species complicates extrapolating this finding to other taxa. In addition, *D. vulcana*'s *hsp68* sequence permitted *hsp70/hsp68* in situ Southern hybridization to persist at higher stringencies than was observed in *D. melanogaster* (Pardali et al. 1996). The latter observation was later extended to other members of the taxon (Drosopolou et al. 1996). While *hsp70* of *D. auraria* and the *hsp68* genes from *montium* subgroup species share on average 82% identity, *hsp70* and *hsp68* genes of *D. melanogaster* share only 78% identity. This difference in sequence similarity may account for the persistent hybridization in the *montium* subgroup when the *hsp70* sequence of *D. auraria* was used as the probe (Pardali et al. 1996; Drosopolou et al. 1996).

These comparisons highlight the nature of the difference between the protein products of the paralogous *hsp70* and *hsp68* genes in the *melanogaster* species group of *Drosophila* and support the idea that a major functional difference between them is in the substrate molecules upon which they act or in the dynamics of their interactions with substrates. That some of the diagnostic sites cluster close to established substrate-binding sites, but do not coincide with them, may provide clues for physical biochemists and biologists attempting to understand the functional diversity of these 2 chaperone families at the molecular, tissue, physiological, and ecological levels.

Acknowledgements

We thank Ary Hoffmann for providing a number of fly lines and species, Fred Govedich, Bonnie Bain, and Brian Bettencourt for advice with the analyses, and Warren Mann, Martin Burd, and an anonymous reviewer for valuable comments on the manuscript. We acknowledge support from the Australian Research Council via their Special Research Centre program.

References

Ashburner, M. 1989. *Drosophila*: a laboratory handbook. Cold Spring Harbour Laboratory Press, Plainview, N.Y. pp. 1161–1190.

Bertelsen, E.B., Zhou, H., Lowry, D.F., Flynn, G.C., and Dahlquist, F.W. 1999. Topology and dynamics of the 10 kDa C-terminal domain of DnaK in solution. Protein Sci. **8**: 343–354.

Bettencourt, B.R., and Feder, M.E. 2001. Hsp70 duplication in the *Drosophila melanogaster* species group: how and when did two become five? Mol. Biol. Evol. **18**: 1272–1282.

Bettencourt, B.R., and Feder, M.E. 2002. Rapid concerted evolution via gene conversion at the *Drosophila hsp70* genes. J. Mol. Evol. **54**: 569–586.

Bock, I.A. 1980. Current status of the *Drosophila melanogaster* species-group (Diptera). Syst. Entomol. **5**: 341–356.

Drosopolou, E., Konstantopoulou, I., and Scouras, Z.G. 1996. The heat shock genes in the *Drosophila montium* sub-group: chromosomal localisation and evolutionary implications. Chromosoma, **105**: 104–110.

Feder, M.E., and Krebs, R.A. 1998. Natural and genetic engineering of the heat-shock protein Hsp70 in *Drosophila melanogaster*: consequences for thermotolerance. Am. Zool. **38**: 503–517.

Feder, M.E., Blair, N., and Figueras, H. 1997. Natural thermal stress and heat shock protein expression in *Drosophila* larvae and pupae. Funct. Ecol. **11**: 90–100.

Flaherty, K.M., DeLuca-Flaherty, C., and McKay, D.B. 1990. Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. Nature (Lond.), **346**: 623–628.

Freeman, B.C., Myers, M.P., Schumacher, R., and Morimoto, R.I. 1995. Identification of a regulatory motif in Hsp70 that affects ATPase activity, substrate binding and interaction with HDJ-1. EMBO J. **14**: 2281–2292.

Gupta, R.S., and Golding, G.B. 1993. Evolution of the *hsp70* gene and its implications regarding relationships between archaebacteria, eubacteria and eukaryotes. J. Mol. Evol. **12**: 1063–1073.

Holmgren, R., Corces, V., Morimoto, R., Blackman, R., and Meselson, M. 1981. Sequence homologies in the 5′ regions of 4 *Drosophila* heat-shock genes. Proc. Natl. Acad. Sci. U.S.A. **78**: 3775–3778.

Ish-Horowicz, D., Pinchin, S.M., Schedl, P., Artavanis-Tsakonas, S., and Mirault, M.E. 1979. Genetic and molecular analysis of the 87A7 and 87C1 heat-inducible loci of *D. melanogaster*. Cell, **18**: 1351–1358.

Kornberg, T.B., and Krasnow, M.A. 2000. The *Drosophila* genome sequence: implications for biology and medicine. Science (Wash., DC), **287**: 2218–2220.

McColl, G., Hoffmann, A.A., and McKechnie, S.W. 1996. Response of two heat shock genes to selection for knockdown heat resistance in *Drosophila melanogaster*. Genetics, **143**: 1615–1627.

Michalak, P., Monkov, I., Helin, A., Lerman, D.N., Bettencourt, B.R., Feder, M.E., Korol, A.B., and Nevo, E. 2001. Genetic evidence for adaptation-driven incipient speciation of *Drosophila melanogaster* along the microcline contrast in 'Evolutionary Canyon', Israel. Proc. Natl. Acad. Sci. U.S.A. **98**: 13195–13200.

Morhauser, R.C., Hu, W., Wang, H., Pang, Y., Flynn, G.C., and Zuiderweg, E.R.P. 1999. High-resolution solution structure of the 18 kDa substrate-binding domain of the mammalian chaperone protein Hsc70. J. Mol. Biol. **289**: 1387–1403.

Palter, K.B., Watanabe, M., Stinson, L., Mahowald, A.P., and Craig, E.A. 1986. Expression and localisation of *Drosophila melanogaster* Hsp70 cognate proteins. Mol. Cell. Biol. **6**: 1187–1203.

Pardali, E., Feggou, E., Drosopoulou, E., Konstantopoulou, I., Scouras, Z.G., and Mavragani-Tsipidou, P. 1996. The Afrotropical *Drosophila montium* sub-group: Balbani ring 1, polytene chromosomes and heat shock response of *D. vulcana*. Genome, **39**: 588–597.

Parsell, D.A., and Lindquist, S. 1993. The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. Annu. Rev. Genet. **27**: 437–496.

Pontius, B.W. 1993. Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association. Trends Biochem. Sci. **18**: 181–186.

Schiffer, M., Carew, M.E., and Hoffmann, A.A. 2004. Molecular, morphological and behavioural data reveal the presence of a cryptic species in the widely studied *Drosophila serrata* species complex. J. Evol. Biol. **17**: 430–442.

Velikodvorskaia, V.V., Lyozin, G.T., Feder, M.E., and Evgen'ev, M.B. 2005. Unusual arrangement of the *hsp68* locus in the *virilis* species group of *Drosophila* implicates evolutionary loss of an *hsp68* gene. Genome, **48**: this issue.

Wang, H., Kurochin, A.V., Pang, Y., Hu, W., Flynn, G.C., and Zuiderweg, E.R.P. 1998. NMR solution structure of the 21 kDa chaperone protein DnaK substrate binding domain: a preview of chaperone–protein interaction. Biochemistry, **37**: 7929–7940.

Wang, T.F., Chang, J.H., and Wang, C. 1993. Identification of the peptide binding domain of Hsc70. J. Biol. Chem. **268**: 26049–26051.

Zhu, X., Zhao, X., Burkholder, W.F., Gragerov, A., Ogata, C.M., Gottesman, M.E., and Hendrikson, W.A. 1996. Structural analysis of substrate binding by the molecular chaperone DnaK. Science (Wash., DC), **272**: 1606–1614.