The sperm outer dense fiber protein is the 10th member of the superfamily of mammalian small stress proteins

Jean-Marc Fontaine,¹ Joshua S. Rest,² Michael J. Welsh,¹ and Rainer Benndorf¹

¹Department of Cell and Developmental Biology, Medical School, University of Michigan, 1335 Catherine Street, Ann Arbor, MI 48109-0616, USA

²Department of Ecology and Evolutionary Biology and Museum of Zoology, University of Michigan, 1109 Geddes Avenue, Ann Arbor, MI 48109-1079, USA

Abstract Nine proteins have been assigned to date to the superfamily of mammalian small heat shock proteins (sHsps): Hsp27 (HspB1, Hsp25), myotonic dystrophy protein kinase-binding protein (MKBP) (HspB2), HspB3, α A-crystallin (HspB4), α B-crystallin (HspB5), Hsp20 (p20, HspB6), cardiovascular heat shock protein (cvHsp [HspB7]), Hsp22 (HspB8), and HspB9. The most pronounced structural feature of sHsps is the α -crystallin domain, a conserved stretch of approximately 80 amino acid residues in the C-terminal half of the molecule. Using the α -crystallin domain of human Hsp27 as query in a BLAST search, we found sequence similarity with another mammalian protein, the sperm outer dense fiber protein (ODFP). ODFP occurs exclusively in the axoneme of sperm cells. Multiple alignment of human ODFP with the other human sHsps reveals that the primary structure of ODFP fits into the sequence pattern that is typical for this protein superfamily: α -crystallin domain (conserved), N-terminal domain (less conserved), central region (variable), and C-terminal tails (variable). In a phylogenetic analysis of 167 proteins of the sHsp superfamily, using Bayesian inference, mammalian ODFPs form a clade and are nested within previously identified sHsps, some of which have been implicated in cytoskeletal functions. Both the multiple alignment and the phylogeny suggest that ODFP is the 10th member of the superfamily of mammalian sHsps, and we propose to name it HspB10 in analogy with the other sHsps. The C-terminal tail of HspB10 has a remarkable low-complexity structure consisting of 10 repeats of the motif C-X-P. A BLAST search using the C-terminal tail as query revealed similarity with sequence elements in a number of *Drosophila* male sperm proteins, and mammalian type I keratins and cornifin- α . Taken together, the following findings suggest a specialized role of HspB10 in cytoskeleton: (1) the exclusive location in sperm cell tails, (2) the phylogenetic relationship with sHsps implicated in cytoskeletal functions, and (3) the partial similarity with cytoskeletal proteins.

INTRODUCTION

The superfamily of mammalian small heat shock proteins (sHsps) has several representatives in all organisms studied. By now, 9 proteins have been assigned to this group of proteins in humans (Kappe et al 2001). The first-known (since 1894) and best studied sHsps are the α -crystallins with the 2 polypeptides, α A-crystallin (α A-Cry, HspB4) and α B-crystallin (α B-Cry, HspB5) (Mörner 1894; MacRae 2000). Whereas α A-Cry occurs almost exclusively in the eye lens, α B-Cry additionally occurs in substantial amounts in cardiac, skeletal, and smooth muscle and in

varying amounts in a number of other tissues such as kidney and brain (Bhat and Nagineni 1989). Some 90 years after the discovery of the crystallins, Hsp27 (HspB1; also called Hsp25 in rodents and birds), the classic heat-inducible mammalian sHsp, was discovered (Hickey et al 1986). However, Hsp27 also occurs in the absence of stress in a number of tissues including cardiac, skeletal, and smooth muscles, brain, kidney, and testis (Pittenger et al 1992; Sugiyama et al 2000). In the past 10 years more proteins have been added to this list: Hsp20 (p20, HspB6) was found in α B-Cry preparations and occurs in considerable amounts in cardiac, skeletal, and smooth muscles (Kato et al 1994; Beall et al 1999; Sugiyama et al 2000). Myotonic dystrophy protein kinase-binding protein (MKBP, HspB2) was identified by sequencing the 5'-flank-

Received 7 February 2003; Accepted 10 February 2003.

Correspondence to: Rainer Benndorf, Tel: 734 615-5670; Fax: 734 763-1166; E-mail: rbenndo@umich.edu.

ing region of the α B-Cry gene (Iwaki et al 1997) and, independently, as a protein that binds to and activates the myotonic dystrophy protein kinase (Sugiyama et al 2000). MKBP occurs in striated muscles. HspB3 and cvHsp (HspB7) were identified on the basis of expressed sequence tag (EST) analysis, and both proteins occur essentially in striated muscles (Boelens et al 1998; Krief et al 1999; Sugiyama et al 2000). Hsp22 (HspB8, H11, E21GI) has been identified independently in 4 laboratories as an estrogen-regulated protein in the breast cancer-derived cell line MCF-7 (Charpentier et al 2000), as a viral protein kinase-related protein (Smith et al 2000), as a protein interacting with Hsp27 (Benndorf et al 2001), and by EST analysis (Kappe et al 2001). In addition to muscles and MCF-7 cells, Hsp22 also occurs in other tissues, eg, placenta, prostate, colon, brain, and keratinocytes (Aurelian et al 2001; Benndorf et al 2001; Kappe et al 2001). HspB9, which is expressed exclusively in germ cells in the testis during certain stages of spermatogenesis, was also identified on the basis of an EST analysis (Kappe et al 2001). The distinguishing feature of sHsps is a conserved stretch of approximately 80 amino acid residues in the C-terminal part of the molecule, the so-called α -crystallin domain. Other sequence characteristics are the less conserved N-terminal domain, the variable central region, and the variable region of C-terminal tails.

Here, we show that another protein, the sperm outer dense fiber protein (ODFP), is the 10th mammalian member of this protein superfamily, on the basis of multiple alignment and phylogenetic analysis. ODFP (HspB10) has all the structural characteristics typical for this protein superfamily but is distinguished from the other sHsps by an extended N-terminal domain and by a C-terminal tail of low-complexity structure.

MATERIALS AND METHODS

BLAST search and multiple alignment

A BLASTP search using the α -crystallin domain of human Hsp27 (position 118–213 in Fig 1) and Swissprot as database was performed at the National Center for Biotechnology Information website (<http://www.ncbi.nlm.nih.gov/BLAST/>). Multiple alignment was performed using CLUSTALW (Thompson et al 1994). The resulting alignment of the 10 sequences was edited manually to discern additional common sequence elements. Amino acid groupings according to physicochemical properties were done according to Livingstone and Barton (1993): small (Sm), tiny (Ti), hydrophobic (Hy), aliphatic (Al), aromatic (Ar), polar (Pl), charged (Ch), positive (Ps), negative (Ne), and proline (P). The BLASTP search using the C-terminal tail of HspB10 (position 214–262) as query

was performed without the low-complexity filter to identify repeats of the C-X-P motif in other proteins.

The accession numbers of the used protein sequences are human Hsp27 (HspB1), 19855073; human MKBP (HspB2), 6016268; human HspB3, 6016270; human α A-Cry (HspB4), 1706112; human α B-Cry (HspB5), 117385; human Hsp20 (HspB6), 22096351; human cvHsp (HspB7), 12643877; human Hsp22 (HspB8), 7644380; human HspB9, 15082238; human ODFP (HspB10), 13249342. More accession numbers are given in Figure 2 and the legend of Figure 3.

Phylogenetic analysis

Profile-hidden Markov models (HMMs) in HMMR 2.2g were used to characterize, search for, and align diverse sHsp superfamily members (Eddy 1996, 1998; Karchin and Hughey 1998). The N-terminal (position 59–91 in Fig 1) and α -crystallin (position 118–213 in Fig 1) domains of a multiple alignment of the 9 human sHsps and ODFP (similar to that shown in Fig 1) were used to construct a profile sHsp-HMM. The nonredundant protein database (GenBank CDS translations+PDB+SwissProt+PIR, 1 November 2002) was queried with sHsp-HMM for significant matches. From this analysis, proteins with bit scores of 4 or greater and expectation values of 0.05 or less were retained because sequences with these scores can be statistically considered members of the sHsp family. We aligned these resulting sequences with sHsp-HMM.

To estimate phylogeny, a Bayesian inference approach (Yang and Rannala 1997; Mau et al 1999) with Metropolis-coupled Markov chain Monte Carlo, or (MC)³, was used, to approximate the posterior probabilities (PPs) of the trees in MrBayes 3.0alpha for the above HMM alignment (Huelsenbeck and Ronquist 2001). The WAG substitution matrix (Whelan and Goldman 2001), which outperforms other general models for amino acid sequence evolution, was used, and the α -parameter for the γ -distribution of rates was estimated. The search was run twice, starting from random trees with 4 simultaneous Markov chains and sampling every 100 generations. A burn-in period of 2×10^6 generations was necessary for the chains to reach stationarity, and the chains were run for an additional 2×10^6 generations to sample the PP landscape. The proportion of searches in which any given node (set of relationships) is found during the chain is an approximation of its PP and provides an indication of support for that node based on the data set.

RESULTS

BLASTP search for proteins similar to human Hsp27

A BLASTP search with the α -crystallin domain of human Hsp27 as the query sequence (position 118–213 in Fig 1)

Fig. 1. Multiple alignment of the known 10 human small heat shock proteins. The alignment was done using CLUSTALW and improved by manual editing. Identical amino acid residues are highlighted in black or gray if they occur in at least 6 or 5 of the 10 sequences, respectively. Similar amino acid residues (E/D, A/G, H/F/W/Y, S/T, I/L/V, H/R/K) are highlighted in gray if they occur in at least 1 of the remaining sequences or in at least 5 of the 10 sequences. Sequence motifs in the variable regions common to at least 2 sequences are in bold and underlined. The corresponding consensus motifs are given in the bottom line using the amino acid groupings as described in the Materials and Methods.

identified all the known mammalian sHsps. Also indicated as a protein with sequence similarity was the ODFP of the mammals human, pig, cow, rat, and mouse (Burfeind and Hoyer-Fender 1991; Hofferbert et al 1993; Hoyer-Fender et al 1995; Kim et al 1995). ODFP was previously identified as a cysteine-proline-rich protein in rat, and later in human, by screening a rat testis complementary deoxyribonucleic acid library for transcripts homologous to the *Drosophila melanogaster* MST(3)CGP gene family, which is expressed in the fly male germ line (Burfeind and Hoyer-Fender 1991). ODFP is expressed in testis of mammals but not in other tissues such as spleen, kidney, or brain. ODFP is the major protein of the outer dense fibers of the axoneme in the midpiece and principal piece of sperm tails (Gastmann et al 1993; Hoyer-Fender et al 1995).

Multiple alignment of human ODFP with the 9 known human sHsps

Human ODFP was aligned with the 9 known human sHsps, using the CLUSTALW program, and then was further adjusted by hand. In this alignment, ODFP obviously shares similarity with the other sHsps. In addition to the initial methionine, 3 amino acid residues (indicated by an asterisk in Fig 1) are conserved throughout all 10 sequences. As in the earlier analyses (de Jong et al 1998), based on this pattern of identical and similar amino acid residues, the different domains and regions of sHsps can be defined: the conserved α -crystallin domain (position 118–213), the more variable N-terminal domain (position 48–91), the variable central region (position 92–117), and the variable region of the C-terminal tails (position 214–

Fig. 2. Phylogeny of Bilateria small heat shock proteins (sHsps) according to Bayesian inference. A hidden Markov model was built of the 10 human sHsps and used to search the nonredundant protein database and align the resulting proteins. Bayesian inference phylogenetic analysis was conducted to ascertain the evolutionary relationships among these sHsps. Support values in the form of Bayesian posterior probabilities are shown next to each node. The tree is midpoint rooted, consistent with using the Arthropod and Nematode sHsps as outgroups. Clades with triangles have been collapsed for brevity; however, the number of proteins in that clade is indicated in parentheses. Relevant taxonomic groupings, gene names, and GenInfo identifiers are supplied where relevant. See text for additional details.

262). Characteristic sequence elements in the α -crystallin domain are the L-D-V-Ps-H-F-S/T-P-Ne-Ne-Al-X-V-like motif (position 135–147), the E-R-X-D-Ch-H-G-Hy-Al-S-R-X-F-X-Ps-Ps-Y-like motif (position 168–184), and the L-P-X-X-V-D-P-X-X-Al-S/T-X-X-L-S-like motif (position 189–203). The N-terminal domain, although less con-

served, also includes motifs common to most of the sequences, including the motif P-Hy (position 49–50), the P-S-R-L-like motif (position 59–62), the F-G-X-X-G-L-like motif (position 67–71), and the P-Ar-Ar-Al-R-P-like motif (position 86–91).

In the 2 highly variable regions, the central region and

human ODFP, C-term. tail	202	CYPCTSPCSPCSPCNPNP-CSPCNP-CSFYDPCNPNPCYPCGSRFSCRKMIL
<i>Drosophila</i> MST84	12	CSPCGGPCGPGCGPGCGSCCSPCGSCCAPWGPGCGPGCCGGCGPCGLCG
human ODFP, C-term. tail	202	CYPCTSPCSPCSPCNPNP----PCS--PCNPNCSFYDPCNPNPCY---CGSRFSCRKMIL
human keratin	362	SEDCLNLPNPNPCATTNACSKPIGPCLSNPCTSCVPPAPCTPCAPRPRCG---PCNSFVR
human ODFP, C-term. tail	202	CYPCTSPCSPCSPCSP--CNPCN--PCSPCN--PCSP--YDPCNPNPCY--PCGSRFSCRKMIL
mouse cornifin- α	41	CHPVPEPCNPKGPEPCHPKAPEPCHPKAPEPCNPKVPEPCQPKVPEPCQPKVPEPCNPK

Fig. 3. Structure of the C-terminal tail. The C-terminal tail of human HspB10 was aligned with sequence stretches of proteins retrieved from the database: *Drosophila* male-specific sperm protein MST84DB (400212), human keratin type I (6016413), mouse cornifin- α (3023555). Identical amino acid residues are highlighted in gray. Numbers refer to positions in protein sequences.

the region of the C-terminal tails, there is almost no similarity according to the applied criteria (see legend of Fig 1). However, in these regions there are several sequence elements that are common to at least 2 of the sHsps. These elements are Hy-R-Sm, P-S-Hy, A-X-P-A, A-A-Q, and Hy-D-S/T in the central region and R-H-X-X-T-E, N-E-AI, I-X-AI, R-E-E-K-P, E-A-A, G-S-Ps, and A-A(X)-K in the region of the C-terminal tails (Fig 1). Similarly, in the N-terminal domain and the far N-termini there are the elements M-D-AI-Sm-I, S-C-Hy-Hy, and D-Hy-Ar-P-Ar, each shared by 2 sequences.

This alignment shows that ODFP has all the structural features characteristic of sHsps: the α -crystallin domain, the N-terminal domain, the central region, and the C-terminal tail. ODFP also shares a number of sequence elements in the variable regions (S-C-Hy-Hy, D-Hy-Ar-P-Ar, Hy-R-Sm, G-S-Ps) with Hsp22, Hsp27, and HspB9 (Fig 1). Unlike the other sHsps, ODFP has a unique sequence stretch of approximately 20 additional amino acid residues near the N-terminus, which may justify the formal separation of the far N-termini from the core N-terminal domain. In these N-terminal tails of most of the aligned sHsps, there are the motifs Hy-P-Hy (position 7–9) and Hy-Ps-Ps (position 17–19), as shown in Figure 1. The C-terminal tail of ODFP has a remarkably regular sequence (structure of low complexity) consisting of 10 repeats of the sequence motif C-X-P (Figs 1 and 3; see below).

Phylogeny of the Bilateria sHsps (including ODFP)

An sHsp-HMM was used to model the aligned N-terminal (position 59–91) and C-terminal (position 118–213) domains of the human sHsps (see Fig 1). The sHsp-HMM identified 245 significant matches in the nonredundant protein database, including all 10 of the sequences in the original alignment. Only a single representative was retained for identical sequences in the alignment. We scrutinized the resulting sequences and removed any that we determined were duplicates due to sequencing variations or truncations. The resulting alignment contains 167 sHsps from Bilateria (Metazoa: Animalia: Eumetazoa: Bilateria), including arthropods (19 from *D. melanogaster*),

nematodes (12 from *Caenorhabditis elegans*), platyhelminthes, and chordates. This alignment was used to build a phylogeny using Bayesian inference. An abbreviated version of this phylogeny illustrating the overall topology of Bilateria sHsps while detailing the relationships between ODFP and neighboring clades is shown in Figure 2. Details and analysis of all Bilateria sHsps represented in this tree will be published elsewhere (J. S. Rest et al, in preparation). The phylogeny was rooted at the midpoint according to branch lengths obtained from the Bayesian inference. This rooting places a clade of arthropod sHsps and a clade of nematode sHsps at the base of the tree, consistent with the sister relationship between nematodes and arthropods (Ecdysozoa) (Peterson and Eernisse 2001) with respect to other Bilateria, although the presence of several arthropod and nematode clades in the in-group requires further investigation of the record of gene duplications with respect to the position of the root (J. S. Rest et al, in preparation). Mammalian ODFPs form a clade with 2 sHsp-like proteins (HR29) from 2 ascidians, *Ciona intestinalis* and *Halocynthia roretzi* (Chordata: Urochordata: Ascidiacea). The monophyly of this ODFP group is supported with a PP of 1.0 from the Bayesian inference. This ODFP group is sister to the Hsp30 group (1.0 PP) from fish (Euteleostomi: Actinopterygii: Neopterygii: Teleostei: Euteleostei) and frogs (Amphibia: Batrachia: Anura). These 2 clades together are sister to an sHsp clade (1.0 PP) of 3 nematode (Nematoda: Chromadorea) and 2 insect (Arthropoda: Insecta: Diptera) species. These 3 clades form a monophyletic group (0.85 PP) and are in an unresolved polytomy with 4 other lineages: (1) an sHsps clade (0.98 PP) composed of 5 flatworm (Platyhelminthes) species and, interestingly, of mammalian HspB9, (2) a clade of 4 sHsps (1.0 PP) from *C. elegans* (Nematoda: Chromadorea), (3) an sHsp from the brine shrimp, *Artemia franciscana* (Arthropoda: Crustacea), and (4) a large clade (0.79 PP) composed of Euteleostomi (including Mammalia) sHsps (α A-Cry, α B-Cry, Hsp20, Hsp27, Hsp22, HspB3, cvHsp, MKBP). The monophyly of the 5 clades of this polytomy, with respect to the nematode and arthropod out-groups, is supported by a PP of 0.66. The crown position of the ODFPs in this phylogeny of sHsps strongly suggests that ODFP is itself an sHsp.

In addition to the phylogenetic analysis, we tested whether ODFPs significantly fit an sHsp-HMM built with the 9 human sHsps but without ODFP. ODFPs from *C. intestinalis* and *H. roretzi* significantly fit this HMM (data not shown), further indicating that ODFPs are part of the sHsp family.

Structure of the C-terminal tail of ODFP

The primary structure of the C-terminal tail of ODFP is noteworthy. It has a low-complexity structure consisting of 10 repeats of the motif C-X-P. A BLASTP search using the C-terminal tail (position 214–262 in Fig 1) of ODFP as query showed that no other protein has a sequence element with an equivalent regular repeat of this C-X-P motif. BLASTP identified a number of *D. melanogaster* male-specific sperm proteins coded for by the MST84 gene cluster (Kuhn et al 1988), several mammalian type I keratins (Fink et al 1995), and mouse cornifin- α (Kartasova et al 1996) as similar to the C-terminal tail of ODFP. An alignment of the human ODFP C-terminal tail with sections of representative proteins identified with BLASTP is shown in Figure 3.

DISCUSSION

We show that mammalian ODFP belongs to the superfamily of sHsps. In addition to a BLASTP search and multiple alignment, we used an HMM and Bayesian phylogenetic analysis to support this relationship. Both the N-terminal and the α -crystallin domains of human sHsps (omitting variable regions) were used to build the HMM. HMMs are effective in identifying the positions of amino acids that describe conserved first-order structure of the family, generating a multiple alignment that reveals these conserved regions, and discriminating between family and nonfamily members in a sequence database search. These properties have advantages over those of BLAST or FASTA searches when producing an alignment of conserved protein families. The phylogenetic method we used, Bayesian inference, has advantages over other methods of phylogenetic inference in interpretation of results, consistency (Wilcox et al 2002), and computational speed (Larget and Simon 1999), although additional analysis to test the position of the root and consistency with other methods are warranted (J. S. Rest et al, in preparation).

The sHsp-HMM identified, in a search of the entire nonredundant database, 167 sHsp-like proteins. All these matching proteins are from the Bilateria, or triploblast animals (Metazoa: Animalia: Eumetazoa: Bilateria) including Chordata, Nematoda, Platyhelminthes, and Arthropoda, and they were used to generate the phylogeny in Figure 2. The lack of sHsps from plants, fungi, bacte-

ria, or archaea in this data set is due to the inclusion of the more variable N-terminal domain in the sHsp-HMM.

As shown in Figure 2, the mammalian ODFPs form a well-supported clade with 2 HR29 proteins from urochordate ascidians that have been described before as similar to α -crystalline-Hsp27 family members (Vandenberghe et al 2001). HR29 proteins are contained in preparations of microfilaments from ascidian body wall muscle and are believed to be involved in stabilization of myofibrils (Takagi et al 1993). This phylogenetic relationship of the ODFP genes is consistent with the sister relationship between urochordates (Chordata: Urochordata) and craniates (Chordata: Craniata), although the 2 urochordates here are paraphyletic, which may be an artifact of the analysis (Nelson 1994). The ODFP-HR29 clade is sister to the clade of the Hsp30 group from frogs and fish. Hsp30 in the bullfrog, *Rana catesbeiana*, is believed to play a role in developmental transitions of the liver (Helbing et al 1996). Thus, sHsp-like proteins in sister clades of ODFP may play a role in cytoskeletal organization and in development. The ODFP-HR29-Hsp30 clade is sister to a clade of 2 insect and 3 nematode sHsps, including Hsp25 from *C. elegans*. Hsp25 of this nematode has been implicated in the organization and maintenance of the myofilament lattice and adherens junctions (Ding and Candido 2000). Because ODFPs (HspB10) are nested within a group of sHsps that are implicated in cytoskeletal functions, a similar function is suggested for these proteins.

To date, 5 highly conserved mammalian ODFPs from human, pig, cow, rat, and mouse have been described. ODFP is one of the main components of the outer dense fibers that surround the axoneme of the midpiece and the principal piece of sperm tails. Although their functions are not defined, it is believed that the outer dense fibers are involved in the maintenance of elastic structures and the elastic recoil of the sperm tails (Baltz et al 1990). Based on the abundance and highly specific expression in the outer dense fibers, ODFP has been suggested to play a role in these events.

A further argument for a cytoskeletal role of ODFP (HspB10) in sperm cells comes from the unusual primary structure of the C-terminal tail. Although the regularity of the C-X-P motif of the tail is unique in the protein databases, a few proteins (which otherwise are not related to ODFP) share similarity with this C-X-P repeat. This similarity is most likely convergent rather than due to shared common ancestry (homology). The best match is the short proline- and cysteine-rich *Drosophila* male-specific sperm protein (gene locus MST84, accession 400212). This protein is similar to another *Drosophila* male-specific sperm protein from the same gene locus (accession 72136) that was used to detect the ODFPs in several mammals (Kuhn et al 1988; Burfeind and Hoyer-Fender 1991). The

fact that these *Drosophila* proteins are expressed exclusively in sperm cells suggests that these C-X-P-like motifs are of functional importance in these cells. The second sequence group retrieved from the database were keratins, a well-characterized group of intermediate filament proteins, which contribute substantially to the special properties of epidermal cells, skin, and its derivatives. The third retrieved sequence was cornifin- α , the function of which is to serve as a cross-bridging protein in cell envelopes, and it is assumed that the mechanical attributes of cell envelopes are determined in part by the cornifin- α content (Kartasova et al 1996). Thus, the similarity of the C-terminal tail of ODFP with a *Drosophila* sperm cell-specific protein on the one hand and mammalian keratins and cornifin- α on the other hand suggests that ODFP is an sHsp with a specialized role in cytoskeletal structures of sperm cells.

In summary, the specific expression and subcellular localization of ODFP (HspB10), its phylogenetic position, and the unusual primary sequence of its C-terminal tail point to a cytoskeletal role of this protein. Thus, ODFP (HspB10) might share this role with other, better-studied sHsps (eg, Hsp27, α B-Cry) that are known to be involved in the organization of cytoskeletal elements such as microfilaments and intermediate filaments and in the function of muscles (Bitar et al 1991; Beall et al 1999; Yamboliev et al 2000). A better understanding of the multifaceted associations of the sHsps with different cytoskeletal proteins and systems will greatly benefit our knowledge of the function of sHsps.

ACKNOWLEDGMENT

This work was supported by National Institutes of Health grant P01 ES11188 to M.J.W. (Principal Investigator) and R.B.

REFERENCES

- Aurelian L, Smith CC, Winchurch R, Kulka M, Gytoku T, Zaccaro L, Chrest FJ, Burnett JW. 2001. A novel gene expressed in human keratinocytes with long-term in vitro growth potential is required for cell growth. *J Invest Dermatol* 116: 286–295.
- Baltz JM, Williams PO, Cone RA. 1990. Dense fibers protect mammalian sperm against damage. *Biol Reprod* 43: 485–491.
- Beall A, Bagwell D, Woodrum D, Stoming TA, Kato K, Suzuki A, Rasmussen H, Brophy CM. 1999. The small heat shock-related protein, Hsp20, is phosphorylated on serine 16 during cyclic nucleotide-dependent relaxation. *J Biol Chem* 274: 11344–11351.
- Benndorf R, Sun X, Gilmont RR, et al. 2001. Hsp22, a new member of the small heat shock protein superfamily, interacts with mimic of phosphorylated Hsp27 (32 P-Hsp27). *J Biol Chem* 276: 26753–26761.
- Bhat SP, Nagineni CN. 1989. Alpha B subunit of lens-specific protein alpha-crystallin is present in other ocular and non-ocular tissues. *Biochem Biophys Res Commun* 158: 319–325.
- Bitar KN, Kaminski MS, Hailat N, Cease KB, Strahler JR. 1991. Hsp27 is a mediator of sustained smooth muscle contraction in response to bombesin. *Biochem Biophys Res Commun* 181: 1192–1200.
- Boelens WC, Van Boekel MA, De Jong WW. 1998. HspB3, the most deviating of the six known human small heat shock proteins. *Biochim Biophys Acta* 1388: 513–516.
- Burfeind P, Hoyer-Fender S. 1991. Sequence and developmental expression of a mRNA encoding a putative protein of rat sperm outer dense fibers. *Dev Biol* 148: 195–204.
- Charpentier AH, Bednarek AK, Daniel RL, Hawkins KA, Laflin KJ, Gaddis S, MacLeod MC, Aldaz CM. 2000. Effects of estrogen on global gene expression: identification of novel targets of estrogen action. *Cancer Res* 60: 5977–5983.
- de Jong WW, Caspers GJ, Leunissen JA. 1998. Genealogy of the alpha-crystallin-small heat-shock protein superfamily. *Int J Biol Macromol* 22: 151–162.
- Ding L, Candido EP. 2000. Hsp25, a small heat shock protein associated with dense bodies and M-lines of body wall muscle in *Caenorhabditis elegans*. *J Biol Chem* 275: 9510–9517.
- Eddy SR. 1996. Hidden Markov models. *Curr Opin Struct Biol* 6: 361–365.
- Eddy SR. 1998. Profile hidden Markov models. *Bioinformatics* 14: 755–763.
- Fink P, Rogers MA, Korge B, Winter H, Schweizer J. 1995. A cDNA encoding the human type I hair keratin hHal. *J Biochim Biophys Acta* 1264: 12–14.
- Gastmann O, Burfeind P, Gunther E, Hameister H, Szpirer C, Hoyer-Fender S. 1993. Sequence, expression, and chromosomal assignment of a human sperm outer dense fiber gene. *Mol Reprod Dev* 36: 407–418.
- Helbing C, Gallimore C, Atkinson BG. 1996. Characterization of a *Rana catesbeiana* Hsp30 gene and its expression in the liver of this amphibian during both spontaneous and thyroid hormone-induced metamorphosis. *Dev Genet* 18: 223–233.
- Hickey E, Brandon SE, Sadis S, Smale G, Weber LA. 1986. Molecular cloning of sequences encoding the human heat-shock proteins and their expression during hyperthermia. *Gene* 43: 147–154.
- Hofferbert S, Burfeind P, Hoyer-Fender S, Lange R, Haidl G, Engel W. 1993. A homozygous deletion of 27 basepairs in the coding region of the human outer dense fiber protein gene does not result in a pathologic phenotype. *Hum Mol Genet* 2: 2167–2170.
- Hoyer-Fender S, Burfeind P, Hameister H. 1995. Sequence of mouse Odf1 cDNA and its chromosomal localization: extension of the linkage group between human chromosome 8 and mouse 15. *Cytogenet Cell Genet* 70: 200–204.
- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.
- Iwaki A, Nagano T, Nakagawa M, Iwaki T, Fukumaki Y. 1997. Identification and characterization of the gene encoding a new member of the alpha-crystallin/small Hsp family, closely linked to the alphaB-crystallin gene in a head-to-head manner. *Genomics* 45: 386–394.
- Kappe G, Verschuure P, Philipsen RL, Staaldunin AA, van de Boogaart P, Boelens WC, de Jong WW. 2001. Characterization of two novel human small heat shock proteins: protein kinase-related HspB8 and testis-specific HspB9. *Biochim Biophys Acta* 1520: 1–6.
- Karchin R, Hughey R. 1998. Weighting hidden Markov models for maximum discrimination. *Bioinformatics* 14: 772–782.
- Kartasova T, Darwiche N, Kohno Y, et al. 1996. Sequence and expression patterns of mouse SPR1: correlation of expression with epithelial function. *J Invest Dermatol* 106: 294–304.
- Kato K, Goto S, Inaguma Y, Hasegawa K, Morishita R, Asano T. 1994. Purification and characterization of a 20-kDa protein that

- is highly homologous to alpha B-crystallin. *J Biol Chem* 269: 15302–15309.
- Kim Y, Adham IM, Haack T, Kremling H, Engel W. 1995. Molecular cloning and characterization of the bovine and porcine outer dense fibers cDNA and organization of the bovine gene. *Biol Chem Hoppe-Seyler* 376: 431–435.
- Krief S, Faivre JF, Robert P, et al. 1999. Identification and characterization of cvHsp. A novel human small stress protein selectively expressed in cardiovascular and insulin-sensitive tissues. *J Biol Chem* 274: 36592–36600.
- Kuhn R, Schäfer U, Schäfer M. 1988. Cis-acting regions sufficient for spermatocyte-specific transcriptional and spermatid-specific translational control of the *Drosophila melanogaster* gene *mst(3)gl-9*. *EMBO J* 7: 447–454.
- Larget B, Simon DL. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol Biol Evol* 16: 750–759.
- Livingstone CD, Barton GJ. 1993. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput Appl Biosci* 9: 745–756.
- MacRae TH. 2000. Structure and function of small heat shock/alpha-crystallin proteins: established concepts and emerging ideas. *Cell Mol Life Sci* 57: 899–913.
- Mau B, Newton MA, Larget B. 1999. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Biometrics* 55: 1–12.
- Mörner CT. 1894. Untersuchung der Proteinsubstanzen in den lichtbrechenden Medien des Auges. *Hoppe Seyler's Z Physiol Chem* 18: 61–106.
- Nelson JS. 1994. *Fishes of the World*, 3rd ed. John Wiley & Sons, New York.
- Peterson KJ, Eernisse DJ. 2001. Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA gene sequences. *Evol Dev* 3: 170–205.
- Pittenger GL, Gilmont RR, Welsh MJ. 1992. The low molecular weight heat shock protein (Hsp27) in rat Sertoli cells: evidence for identity of Hsp27 with a germ cell-responsive phosphoprotein. *Endocrinology* 130: 3207–3215.
- Smith CC, Yu YX, Kulka M, Aurelian L. 2000. A novel human gene similar to the protein kinase (PK) coding domain of the large subunit of herpes simplex virus type 2 ribonucleotide reductase (ICP10) codes for a serine-threonine PK and is expressed in melanoma cells. *J Biol Chem* 275: 25690–25699.
- Sugiyama Y, Suzuki A, Kishikawa M, et al. 2000. Muscle develops a specific form of small heat shock protein complex composed of MKBP/HspB2 and HspB3 during myogenic differentiation. *J Biol Chem* 275: 1095–1104.
- Takagi T, Yasunaga H, Nakamura A. 1993. Structure of 29-kDa protein from ascidian (*Halocynthia roretzi*) body wall muscle. *J Biochem (Tokyo)* 113: 321–326.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680.
- Vandenberghe AE, Meedel TH, Hastings KE. 2001. mRNA 5'-leader trans-splicing in the chordates. *Genes Dev* 15: 294–303.
- Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 18: 691–699.
- Wilcox TP, Zwickl DJ, Heath TA, Hillis DM. 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Mol Phylogenet Evol* 25: 361–371.
- Yamboliev IA, Hedges JC, Mutnick JL-M, Adam LP, Gerthoffer WT. 2000. Evidence for modulation of smooth muscle force by the p38 MAP kinase/Hsp27 pathway. *Am J Physiol Heart Circ Physiol* 278: H1899–H1907.
- Yang ZH, Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. *Mol Biol Evol* 14: 717–724.