ELSEVIER

The HSP90 family of genes in the human genome: Insights into their divergence and evolution

Bin Chen [a,b,*], William H. Piel [a], Liming Gui [c], Elspeth Bruford [d], Antónia Monteiro [a]

[a] *Department of Biological Sciences, The State University of New York at Buffalo, NY 14260, USA*
[b] *College of Plant Protection, Southwest University, Chongqing 400716, P.R. China*
[c] *School of Biology, The University of Leeds, Leeds LS2 9JT, UK*
[d] *The Galton Laboratory, University College London, London NW1 2HE, UK*

Abstract

HSP90 proteins are important molecular chaperones. Transcriptome and genome analyses revealed that the human HSP90 family includes 17 genes that fall into four classes. A standardized nomenclature for each of these genes is presented here. Classes HSP90AA, HSP90AB, HSP90B, and TRAP contain 7, 6, 3, and 1 genes, respectively. HSP90AA genes mapped onto chromosomes 1, 3, 4, and 11; HSP90AB genes mapped onto 3, 4, 6, 13 and 15; HSP90B genes mapped onto 1, 12, and 15; and the *TRAP1* gene mapped onto 16. Six genes, *HSP90AA1*, *HSP90AA2*, *HSP90N*, *HSP90AB1*, *HSP90B1* and *TRAP1*, were recognized as functional, and the remaining 11 genes were considered putative pseudogenes. Amino acid polymorphic variants were detected for genes *HSP90AA1*, *HSP90AA2*, *HSP90AB1, HSP90B1*, and *TRAP1*. The structures of these genes and the functional motifs and polymorphic variants of their proteins were documented and the features and functions of their proteins were discussed. Phylogenetic analyses based on both nucleotide and protein data demonstrated that HSP90(AA + AB + B) formed a monophyletic clade, whereas TRAP is a relatively distant paralogue of this clade.
© 2005 Elsevier Inc. All rights reserved.

Keywords: HSP90; TRAP; Gene; Human; Genome; Evolution; Nomenclature; Polymorphism

Introduction

HSP90 proteins, named according to the 90-kDa average molecular mass of their members, are highly conserved molecular chaperones that account for 1–2% of all cellular proteins in most cells under non-stress conditions [1]. HSP90 proteins have key roles in signal transduction, protein folding, protein degradation, and morphological evolution. A fully functional HSP90 protein normally associates with other co-chaperones, playing an important role in the folding of newly synthesized proteins or stabilizing and refolding denatured proteins after stress [2–4]. Its expression is associated with many types of tumors including breast cancer and pancreatic carcinoma, human leukemia, and systemic lupus erythematosus, as well as multidrug resistance [1]. HSP90 inhibition provides a recently developed, important pharmacological platform for anticancer therapy [5].

HSP90 proteins can be found in the cytosol, nucleoplasm, endoplasmic reticulum (ER), mitochondria, and chloroplasts [1,6,7]. Most Eubacteria have a single homologue of HSP90 known as HtpG (high-temperature protein G), whereas Archaebacteria lack a HSP90 representative [8]. All eukaryotes possess cytosolic members, called HSP90 (90 kDa heat-shock protein), in the *sensu stricto*. There are two major cytosolic isoforms of HSP90, HSP90AA1 (inducible form) and HSP90AB1 (constitutive form) [1,9,10]. These two isoforms are the result of a gene duplication approximately 500 million years ago [11]. A recent report added another cytosolic isoform to the HSP90 family, HSP90N, which is associated with cellular transformation [12]. The ER paralogue, generally called Grp94 (94-kDa glucose-regulated protein), is present in all eukaryotes except fungi, which seem to have lost it, and is suggested to have originated via gene duplication very early in evolution [13,14]. A chloroplast homologue that is most

similar in sequence to ER Grp94 has been found in three plant species, suggesting that it originated from the common ancestor of Eubacteria and chloroplasts [8,15]. The mitochondrial paralogue, TRAP1 (tumor necrosis factor receptor-associated protein 1), is most closely related to Eubacterial HtpG sequences, which that suggests it originated from a HtpG-like ancestor [8,16]. As a distinctive feature, TRAP1 possesses a unique LxCxE motif that is absent in all other HSP90 family members [17].

The total number and chromosomal localization of the genes encoding members of the HSP90 family in the human genome are still unknown. Earlier efforts that contributed to identifying and mapping the human HSP90 family of genes were fragmented or misleading due to the use of an incomplete genome assembly, and to the limitation of the experimental methods used ([18] for HSP90AA, [19] for HSP90AB, [20] for HSP90B (GrpP94)). In this study, we (i) determine the number of HSP90 gene copies in the human genome using transcriptome analysis and gene prediction methods, (ii) map the genes on the human genome, (iii) identify amino acid (a.a.) polymorphisms, (iv) predict individual gene function, (v) analyze the evolution of the gene family members, and (vi) propose a standardized nomenclature for the members of this gene family. This nomenclature has been jointly devised and approved by the HUGO Gene Nomenclature Committee (http://www.gene.ucl.ac.uk/nomenclature/). Our aim is to establish a genomic framework of information for the human HSP90 gene family, which may also facilitate and stimulate the study of this gene family across all organisms.

Results

Diversity and mapping of human HSP90 gene family

From the 188 mRNAs and 96 protein sequences retrieved from database searches, 64 mRNAs and 41 protein sequences came from different submissions and were confirmed to belong to the family through similarity comparison after sequence alignment. Thirty-six of these protein sequences were originally submitted to the GenBank/EMBL Data Libraries with an accompanying mRNA, whereas 4 protein sequences (JQ0129, T46243, HHHU86, and HHHU84) had no corresponding mRNA, but were confirmed to belong to the family because of 100% identity to protein sequences AAA36024, CAB66478, NP_005339, and NP_031381, respectively. The last protein sequence without corresponding mRNA, A55877, was 91% identical to NP_057376. All of the mRNAs mapped to five chromosomal sites (*HSP90AA1* and *HSP90N* mapped to the same site), suggesting the presence of a minimum of six expressed genes (Table 1). A TBLASTN search against the human genome assembly identified 77 hits on chromosomes from protein queries and 43 hits from nucleotide queries. In order to reduce prediction error, we filtered out predictions of gene sequences with an exon probability of less than 50% in the GENSCAN calculation or with a coding region (CDS) shorter than 200 bp. We concluded that there are a total of 17 genes belonging to the HSP90 family in the human genome

(Table 1). These genes were all confirmed by nucleotide and protein homology comparisons. Six of these genes were also supported by the presence of corresponding mRNA sequences, as described above, 2 were previously predicted by NCBI annotation, and 9 are newly predicted in this study.

These 17 genes could be clearly divided into four classes by phylogenetic analysis in the present study (see the following section), which comprise 7 genes, 6 genes, 3 genes, and 1 gene, respectively (Table 1, Figs. 4 and 5). We did not find any genes similar to bacterial HtpG nor to plant chloroplast homologues. Earlier known members of the human HSP90 family have been given a variety of names: our *HSP90AA1* is synonymous with HSP90, HSP90AA, HSP90-α, HSPN, LAP2, HSP86, HSPC1, HSPCA, Hsp89, HSPCAL4, and FLJ31884; our *HSP90AB1* is synonymous with HSP90AB, HSP90-β, HSPC2, HSPCB, and D6S182; our *HSP90B1* is synonymous with TRA1, HSP90AA, Grp94, CGP, and GP96; and *TRAP1* (alias *HSP90L*) is synonymous with Hsp75. Obviously, the previous nomenclature is quite confusing and there is a need to name the new members of the HSP90 family discovered in the study. Here we propose a new nomenclature for the gene family (Table 1). We divided the gene family into the four classes mentioned above and then named genes based on class name, with the exception of *HSP90N*. *HSP90N* was originally named Hsp89-α-δ-N [21] and found in the present study to be a chimera of two genes: the first 1–105 bp of the CDS (1–551 in mRNA) are identical to the sequence of the CD47 gene, mapped to chromosome 3 q13.12, whereas the following 91–1620 bp (545–2733 of mRNA) are identical to *HSP90AA1*.

The present study proposes a new nomenclature system for the HSP90 gene family (Table 1) with the root HSP90A indicating cytosolic HSP90 protein, HSP90B indicating endoplasmic reticulum HSP90 protein, and TRAP indicating mitochondrial HSP90 protein. HSP90A was divided into two classes: HSP90AA for conventional Hsp90-α and HSP90AB for Hsp90-beta. The number following the root/class names encodes the gene in the class, and the "P" at the end of the gene name indicates a possible pseudogene. This system has not been used for other gene families or species and it may serve as a model for the nomenclature of other chaperone gene families (e.g., HSP100, HSP70, HSP60, small HSP) that have status similar to that of the HSP90 gene family.

These 17 genes were mapped to 10 different chromosomes (Fig. 1). The 7 HSP90AA genes are located on chromosomes 1, 3, 4, 11, and 14; the 6 HSP90AB genes are on chromosomes 3, 4, 6, 13, and 15; the 3 HSP90B (Grp94) genes are on chromosomes 1, 12, and 15; and the single *TRAP1* gene is on chromosome 16. The chromosomal locations of the family members have no clear correlation with the sequence similarities among these members (Figs. 4 and 5).

Nucleotide and amino acid polymorphism

Considerable nucleotide polymorphism, in both introns and exons, was detected for genes *HSP90AA1*, *HSP90AA2*, *HSP90AB1*, *HSP90AB2*, *HSP90B1*, and *TRAP1*, for which we had multiple mRNA sequences. In total, 27, 24, 9, and 14

mRNAs belonging to the HSP90AA, HSP90AB, HSP90B, and TRAP1 classes, respectively, were analyzed. Exon polymorphisms led to both silent and replacement substitutions. Two gene variants resulted from codon deletion and six from replacement substitutions. In this study, we carefully considered these six gene variants from replacement substitutions and believe that they most probably are not due to sequencing errors due to a number of nucleotide replacement substitutions in each sequence. However, we cannot exclude the possibility of sequencing errors in databases. Nine multiple splicing variants were also detected for these genes (Fig. 2). For each gene variant due to codon deletion, replacement substitutions, and/or splicing variants, exon numbers are shown in Fig. 2 and Table 1, and polymorphism types are summarized in Table 1 of the Supplementary data available online.

Protein features and conservative regions

The amino acid sequences of these 17 genes can be reliably aligned and the alignment of representative sequences from each of the four classes is shown in Fig. 3. Each sequence can be divided in to five domains, i.e., the N-terminal domain (HSP90AA1-1 residues 1–236), charged domain 1 (237–271), middle domain (272–617), charged domain 2 (618–628), and C-terminal domain (629–732) [22]. The N-terminal domain has the highest sequence conservation across all HSP90 sequences [23]. The middle domain is essentially involved in ATPase activity [24] and the cooperation of multiple functional domain is essential for active, chaperone-medicated folding [25]. Among the four representative sequences in Fig. 3, HSP90AA1-1 and HSP90AB1-1 are most similar, with 86% identity and 93% similarity, and both terminate with the functional motif MEEVD, which is characteristic of cytosolic HSP90 [14]. HSP90AB sequences can be easily distinguished from HSP90AA sequences due to the absence of 5 and 3 a.a. in the N-terminal domain and changed domain 1, respectively. HSP90B (Grp94) sequences are the longest of the family, characterized by three unique deletions and three unique insertions. HSP90B1-1 has 42% identity and 59% similarity with HSP90AA1-1 and 43% identity and 60% similarity with HSP90AB1-1. TRAP1 sequences are the shortest, lacking the charged domain 1 and containing five unique deletions and four unique insertion sites. TRAP1-1 has 26% identity and 45% similarity with either HSP90AA1-1 or HspSP90B1-1 and has 28% identity and 47% similarity with HSP90B1-1. The functionally important residues K112 and F369 do not exist in TRAP1-1 and S231 and S263 occur only in HSP90AA1-1 and HSP90AB1-1. In classes HSP90AA and HSP90AB, no signal peptide was detected; however, 21 residues of signal peptide in HSP90B1 (Grp94), and 26–32 residues in TRAP1, were predicted with higher than 80% probability. The molecular mass ranges from 11.74 to 98.11 kDa in HSP90AA, 8.38 to 84.84 kDa in HSP90AB, 33.16 to 90.16 kDa in HSP90B (Grp94), and 54.43 to 76.50 kDa in TRAP1 for mature proteins (Table 1). Those genes with smaller molecular mass are all predicted as putative pseudogenes (Table 1 and Discussion).

In addition to the domain structure and functionally important residues mentioned above, various conservative and/or functional regions shared by all HSP90 members may be fulfilling HSP90-specific functions. InterProScan and ScanProsite searches explored such regions (Table 1, Fig. 3). The "HSP90 protein" sequence (residues 196–732 in HSP90AA1-1, Document No. PF00183 of the InterPro database) exists in all of these amino acid sequences. In most of the genes studied, there are several HSP90 smaller "signature" sequences (residues 18–61, 88–123, 131–153 and 182–218, InterPro doc PR00775 and PS00298; residues 38–47, PROSITE doc PS00298), an ATP-binding domain (residues 40–193, ATPases, InterPro doc PF02518 and SM00387), and a glutamic acid-rich motif (residues 223–268, PROSITE doc PS50313). The four-helical cytokine region (residues 518–668, InterPro doc SSF47266) is present in HSP90AA, HSP90AB, and HSP90B (Grp94) sequences, with the exception of four transcripts, but is absent in TRAP1. In addition, the ER targeting sequence KDEL (residues 800–803 in HSP90B1-1, InterPro doc IPR000886) exists in HSP90B. Other domains/motifs that occur in some of these HSP90 a.a. sequences include a lysine-rich domain (PROSITE doc PS50318), a protein kinase C phosphorylation site (PROSITE doc PS0005), a casein kinase II phosphorylation site (PROSITE doc PS0006), an N-glycosylation site (PROSITE doc PS0001), a tyrosine kinase phosphorylation site (PROSITE doc PS0007), a tyrosine sulfation site (PROSITE doc PS0003), a cAMP- and cGMP-dependent protein kinase phosphorylation site (PROSITE doc PS0004), an N-myristoylation site (PROSITE doc PS0008), a bipartite nuclear targeting site (PROSITE doc PS00015), a leucine zipper domain (PROSITE doc PS00029), and an amidation site (PROSITE doc PS00009). These conservative regions/domains/motifs are only predicted in this study and have not been proven functionally significant.

Phylogenetic analysis

Maximum parsimony analysis of protein data (36 sequences) resulted in a strict consensus tree of 618 equally most-parsimonious trees, with tree length = 2154, consistency index (CI) = 0.9271, homoplasy index (HI) = 0.0729, retention index (RI) = 0.9638, and rescaled consistency index (RC) = 0.8936 (Fig. 4). From the 895 a.a. characters used, 118 characters were constant, 115 were parsimony-uninformative, and 662 were parsimony-informative. The analysis of protein-coding nucleotide data (36 sequences) led to a strict consensus tree of 512 equally most-parsimonious trees (5853 steps, CI = 0.6966, HI = 0.3034, RI = 0.8690, RC = 0.6053) (Fig. 5). Five hundred one of the total 2745 nucleotide characters were constant, 315 were parsimony-uninformative, and 1929 were parsimony-informative.

The two tree topologies strongly support the nodes for HSP90AA + HSP90AB, HSP90B (Grp94), TRAP, HSP90AA + HSP90AB + HSP90B, and complete ingroup with at least 96% bootstrap values. HSP90AA + HSP90AB sequences share a more recent common ancestor with HSP90B (Grp94) sequences than with TRAP, hence our proposed HSP90B

Table 1
Human HSP90 family of genes

| Gene	Gene variant[a]	Accession No.[b] a.a.	Accession No.[b] mRNA	Chromosomal location[c]	mRNA[d]	Exon No.[e]	a.a. length[f]	Mass (kDa)	HSP90 signature[g]	HSP90 protein[g]	ATPase[g]	4-helical cytokine[g]	Gln-rich region[g]	Tissue/cell type
HSP90AA														
HSP90AA1	HSP90AA1-1	NP_005339.2	NM_005348.2	14- q32.32	FLM	10 (13)	732 (732)	84.67	+	+	+	+	+	Various
	HSP90AA1-2	HSP90AA1-2P[N]	AK056446.1	14- q32.32	FLM	12 (12)	854 (854)	98.11	+	+	+	+	+	NT2 cell
	HSP90AA1-3	CAD62296.1	BX247955.1	14- q32.32	FLM	6 (7)	413 (413)	49.36	+/−	+/−	+/−	+	+/−	Thymus
	HSP90AA1-4	HSP90AA1-4P[N]	CR596117.1	14- q32.32	FLM	10 (11)	732 (732)	84.66	+	+	+	+	+	Fetal brain
	HSP90AA1-5	AAH023006.1	BC023006.1	14- q32.32	FLM	9 (9)	635 (635)	73.60	+/−	+	+/−	+	+	Breast
HSP90AA2	HSP90AA2-1[P]	XP_084514.6	XM_084514.9[E]	11- p14.1	FLC	2	757 (757)	87.04	+	+	+	+	+	N/A
	HSP90AA2-2	AAA36024.1	M30627.1	11- p14.1	5′M	1 (1)	312 (312)	35.67	+	+/−	+	+	−	Lamnda gt10
HSP90AA3P	HSP90AA3P[P]	XP_496420.1	XM_496420.1[E]	1+ q23.1	FLC	4	435 (435)	50.05	+/−	+	+/−	+	+/−	N/A
HSP90AA4P	HSP90AA4P[NP]	HSP90AA4PP[N]	HSP90AA4PN[N]	4+ q35.2	FLC	4	418 (418)	47.71	+/−	+/−	+/−	−	+/−	N/A
HSP90AA5P	HSP90AA5P[NP]	HSP90AA5PP[N]	HSP90AA5PN[N,E]	3+ q27.1	FLC	4	334 (334)	38.74	+/−	+/−	+/−	−	+/−	N/A
HSP90AA6P	HSP90AA6P[NP]	HSP90AA6PP[N]	HSP90AA6PN[N]	4- q33	FLC	2	101 (101)	11.74	−	+/−	−	−	+/−	N/A
HSP90N	HSP90N	AAC25497.1	AF028832.1	3- q13.12 14- q32.32	FLM	1+7 (1+7)	539 (539)	63.25	+/−	+	+/−	+	+	CEM cell and pancreas
HSP90AB														
HSP90AB1	HSP90AB1-1	NP_031381.2	NM_007355.2	6+ p21.1	FLM	11 (12)	724 (724)	83.26	+	+	+	+	+	Various
	HSP90AB1-2	CAB66478.1	AL136543.1	6+ p21.1	FLM	11 (12)	737 (737)	84.84	+	+	+	+	+	Amygdala
	HSP90AB1-3	BAB15121.1	AK025370.1	6+ p21.1	3′M	9 (9)	571 (571)	66.37	+/−	+	+/−	+	+	Colon
	HSP90AB1-4	AAF8792.1	AF275719.1	6+ p21.1	FLM	10 (10)	632 (632)	72.87	+/−	+	−	+	+	pancreas
HSP90AB2P	HSP90AB2P-1[NP]	HSP90AB2P-1P[N]	HSP90AB2P-1N[N]	4+ p15.33	FLC	4	422 (422)	49.12	+/−	+/−	+/−	+	+/−	N/A
	HSP90AB2P-2[P]	XP_496686.1	XM_496686.1	4+ p15.33	FRG	1	142 (142)	16.89	+/−	+/−	−	+	−	N/A
HSP90AB3P	HSP90AB3P[NP]	HSP90AB3PP[N]	HSP90AB3PN[N]	4+ q22.1	FLC	4	597 (597)	68.32	+/−	+	+/−	+	+	N/A
HSP90AB4P	HSP90AB4P[NP]	HSP90AB4PP[N]	HSP90AB4PN[N]	15- q21.3	FLC	3	505 (505)	58.26	+/−	+/−	+/−	+	+/−	N/A
HSP90AB5P	HSP90AB5P[NP]	HSP90AB5PP[N]	HSP90AB5PN[N]	3+ p12.3	FLC	1	74 (74)	8.38	−	+/−	−	−	+/−	N/A
HSP90AB6P	HSP90AB6P[NP]	HSP90AB6PP[N]	HSP90AB6PN[N]	13- q32.1	FLC	6	361 (361)	41.46	+/−	+/−	+/−	+/−	+/−	N/A
HSP90B														
HSP90B1	HSP90B1-1	NP_003290.1	NM_003299.1	12+ q23.3	FLM	18 (18)	782 (803)	90.17	+	+	+	+/−	+	Various
	HSP90B1-2	AAH09195.1	BC009195.2	12+ q23.3	FLM	8 (9)	294 (315)	33.16	+	+/−	+	+/−	−	Lung
	HSP90B1-3	AAK74072.1	AY040226.1	12+ q23.3	3′M	17 (17)	782 (782)	90.19	+	+	+	+/−	+	Liver tumor
	HSP90B1-4	HSP90B1-4P[N]	AK025459.1	12+ q23.3	FLM	18 (18)	781 (802)	90.04	+	+	+	+/−	+	Hepatoma
HSP90B2P	HSP90B2P[NP]	HSP90B2PP[N]	HSP90B2PN[N]	15+ q26.3	FLC	3	399 (399)	45.86	+/−	+	−	−	+	N/A
HSP90B3P	HSP90B3P[NP]	HSP90B3PP[N]	HSP90B3PN[N]	1+ p22.1	FLC	3	453 (453)	51.72	+/−	+	+/−	+/−	+/−	N/A
TRAP														
TRAP1	TRAP1-1	NP_057376.1	NM_016292.1	16- p13.3	FLM	18 (18)	672 (704)	76.50	+	+	+	−	+	Various
	TRAP1-2	BAC04139.1	AK093344.1	16- p13.3	FLM	15 (18)	547 (579)	62.54	+	+/−	+	−	+/−	Testis
	TRAP13	AAC02679.1	AF043254.1	16- p13.3	FLM	19 (19)	669 (698)	76.27	+	+	+	−	+	Trachea
	TRAP1-4	TRAP1-4P[N]	AK057620.1	16- p13.3	3′M	13 (13)	469 (495)	54.43	+/−	+	+/−	−	+	N/A
	TRAP1-5	AAH18950.1	BC018950.2	16- p13.3	FLM	18 (18)	672 (704)	76.60	+	+	+	−	+	Brain
	TRAP1-6	AAH01455.2	BC001455.2	16- p13.3	3′M	18 (18)	672 (699)	76.45	+	+	+	−	+	Placenta
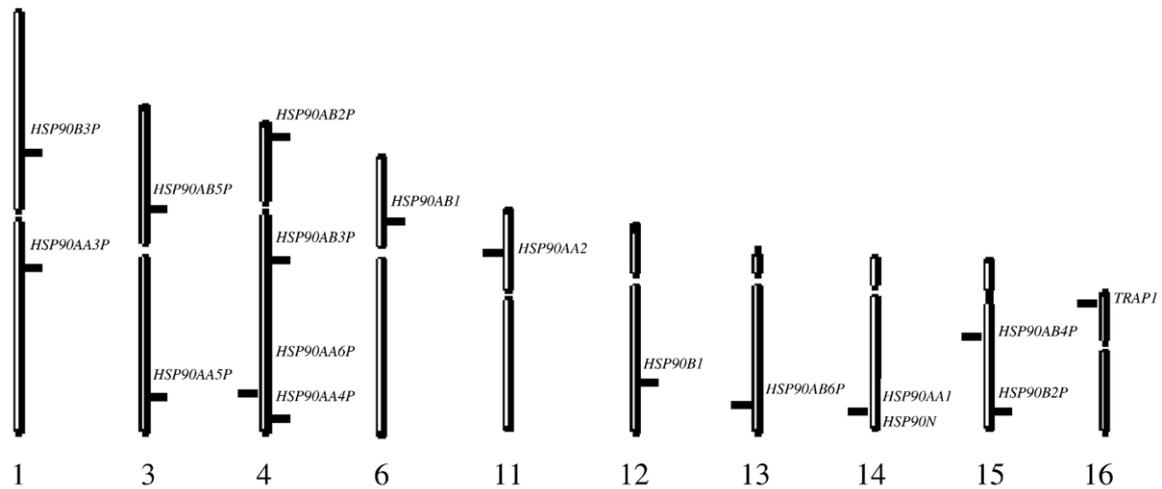
Fig. 1. Chromosomal location of the human HSP90 gene family. Genes mapped to the forward strand are marked on the right side of the chromosomes with horizontal bars, whereas those on the reverse strand are marked on the left side. Gene names are shown on the right side of the corresponding chromosome and chromosome numbers are shown at the bottom.

nomenclature for these Grp94 genes. Nodes for the HSP90AA and HSP90AB members are only moderately supported (88 and 53% for protein data and 63 and 52% for nucleotide data, respectively), whereas there is a clearly supported monophyly of HSP90AA + HSP90AB as a group (96 and 100% bootstrap values for protein and nucleotide data, respectively). Within classes, the genes or gene variants *HSP90AA3P* + *HSP90AA4P* + *HSP90AA5P*, *HSP90AA2*, *HSP90AB1*, and *HSP90B1* are consistently grouped into separate branches in both topologies, but differences exist between the two topologies for the remaining sequences, despite the position of these not being strongly supported in either case.

Discussion

The present study represents the first genome-wide study of the gene family across all organisms. There are a total of 17 genes belonging to this family in the human genome, and their chromosomal locations, a.a. and nucleotide polymorphisms, gene structure, and protein features are documented. These genes were clearly divided into four classes by both sequence alignment and phylogenetic analysis and were named HSP90AA, HSP90AB, HSP90B, and TRAP.

HSPs were originally identified as proteins whose expression is markedly increased following a heat shock and were traditionally classified by molecular weight [26]. The major

HSPs in eukaryotic cells have the approximate sizes of 60, 70, and 90 kDa and are referred to as HSP60, HSP70. and HSP90, respectively. HSP90AA and HSP90AB are heat shock proteins induced by elevated temperature, whereas HSP90B (Grp94) is glucose-regulated and induced by glucose starvation [1]. HSPs generally lack introns, which facilitates their rapid expression while avoiding incorrect RNA spicing due to heat stress [27]. Interestingly, unlike genes encoding other HSPs, HSP90 and TRAP genes contain a large number of introns (e.g., 10 introns in *HSP90AA1-1*, 11 in *HSP90AB1-1*, and 18 in *TRAP1-1*). HSP90 splicing was inhibited after severe heat shock in *Drosophila* cells [28]. This structural feature found in HSP90 and TRAP gene sequences implies a unique function distinguishable from the other HSPs but this has yet to be investigated.

Among the 17 HSP90 genes, 6 genes, *HSP90AA1*, *HSP90AA2*, *HSP90N*, *HSP90AB1*, *HSP90B1*, and *TRAP1*, are supported by corresponding mRNAs, so they appear be functional. The predicted genes *HSP90AA6P*, *HSP90AB5P*, and *HSP90B2P* encode between 74 and 399 a.a. residues, split between 1 to 3 exons, and lack the HSP90 signature motif sequence (except *HSP90B2P*), the ATPase domain, and the four-helical cytokine regions (Table 1). Thus, these genes are most probably pseudogenes and hence we have added a P suffix to the gene symbol. The remaining 8 predicted genes, *HSP90AA3P*, *HSP90AA4P*, *HSP90AA5P*, *HSP90AB2P*,

Notes to Table 1:
   [a] Includes splice variants and /or a.a. polymorphisms (see text). [P], predicted by genome annotation; [NP], predicted in this study; unmarked patterns are identified from cloning studies.
   [b] [N], New sequences identified/deduced in this study and available at Supplementary Table 2 in the online version of this article. [E], containing a stop codon in the middle of mRNA.
   [c] +, Forward strand; −, reverse strand.
   [d] FLM, full-length mRNA; 5′M, 5′ mRNA; 3′M, 3′ mRNA; FLC, fulL-length CDs; FRG, mRNA fragment.
   [e] Total number of exons in CDs. The numbers in parentheses include exons in untranslated regions. 1 + 7 for HSP90N: 1 exon mapped to 3- q13.12 and 7 to 14- q32.32.
   [f] Length of mature a.a. with that of precursor in parentheses.
   [g] Conservative and functionalj regions of proteins. +, complete; −, absent; +/−, incomplete.
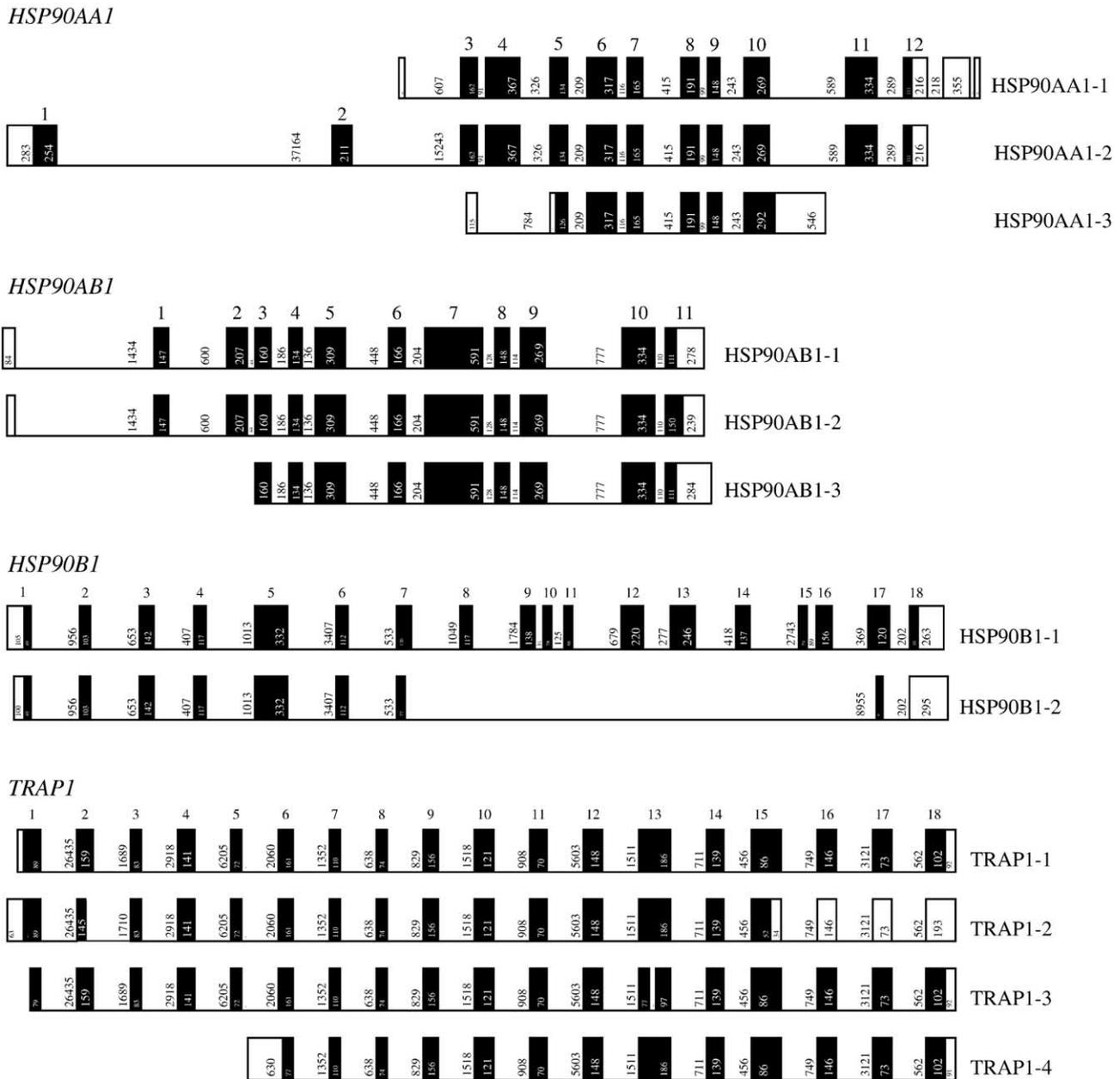
Fig. 2. Exon splicing patterns of gene variants of *HSP90AA1*, *HSP90AB1*, *HSP90B1*, and *TRAP1* drawn to scale with the exception of large introns. The filled rectangles depict protein–coding sequences, with exon numbers above the rectangles, and open rectangles represent untranslated regions. Introns are the open spaces between these rectangles. The numbers of nucleotides are shown inside each region.

HSP90AB3P, *HSP90AB4P*, *HSP90AB5P*, and *HSP90B3P*, encode the HSP90 signature and both the HSP90 protein and ATPase regions, despite these latter being incomplete (except in *HSP90AA3P*, *HSP90AB3P*, and *HSP90B3P*, which encode the complete HSP90 a.a. region). The H-Invitational human gene database contains 41,118 full-length mRNAs that have been clustered into 21,037 gene candidates [29]. Compared to the total number of 30,000–40,000 estimated genes for the human genome [30], this number of gene candidates comprises 52.59–70.12% of the total number of estimated human genes. Because our mRNA search utilized the H-Inv human genome database, and no mRNA was found to match these 8 genes, this suggests that these genes are also possible pseudogenes.

Our research identified abundant nucleotide polymorphisms and alternative splicing patterns in the HSP90 gene family, which may lead to various protein variants with possibly different functions. In addition, depending on the surrounding regulatory regions, each of these proteins may have acquired novel temporal and spatial expression patterns. In the HSP90AA and HSP90AB classes, the HSP90AA1-1 and HSP90AB1-1 proteins are largely cytosolic [1], have ATPase activity, and are involved in the folding of cell regulatory proteins and the refolding of stress-denatured polypeptides [22]. HSP90AA1-1 is somewhat inducible, whereas HSP90AB1-1 is more constitutively expressed [1]. HSP90AA1-4 differs from HSP90AA1-1 by a single residue. HSP90AA1-2 is a much

```
                                                                      ▼
HSP90AA1-1 ----------------------------------------MPEETQTQDQPMEEEEVETFAFQAEIAQLMSLIINTFYSNKEIFLRELISNSSDALDKIR  60
HSP90AB1-1 ----------------------------------------MPEEVHHG-----EEEVETFAFQAEIAQLMSLIINTFYSNKEIFLRELISNASDALDKIR  55
HSP90B1-1  19-RADDEVDVDGTVEEDLGKSREGSRTDDEVVQREEEAIQLDGLNASQIRELREKSEKFAFQAEVNRMMKLIINSLYKNKEIFLRELISNASDALDKIR 116
TRAP1-1    31-PILCPRRTTAQLGPRRNPAWSLQAGRLFSTQTAEDKEEPLHSIISSTESVQGSTSKHEFQAETKKLLDIVARSLYSEKEVIRELISNASDALEKLR  128
                                                                                        ########

                      ▼ ▼                          ▼                                    ▼
HSP90AA1-1 YETLTDPSKLDSGKELHINLIPNKQDRTLTIVDTGIGMTKADLINNLGTIAKSGTKAFMEALQAG-----ADISMIGQFGVGFYSAYLVAEKVTVITKHN 155
HSP90AB1-1 YESLTDPSKLDSGKELKIDIIPNPQERTLTLVDTGIGMTKADLINNLGTIAKSGTKAFMEALQAG-----ADISMIGQFGVGFYSAYLVAEKVVVITKHN 150
HSP90B1-1  LISLTDENALSGNEELTVKIKCDKEKNLLHVTDTGVGMTREELVKNLGTIAKSGTSEFLNKMTEAQEDGQSTSELIGQFGVGFYSAFLVADKVIVTSKHN 216
TRAP1-1    HKLVSLGQALP---EMEIHLQTNAEKGTITIQDTGIGMTQEEIVSNLGTIARSGSKAFLDALQNQ---AEASSKIIGQFGVGFYSAFMVADRVEVYSRSA 222

                                              N-terminal domain                        ▼     charged domain 1
HSP90AA1-1 DDEQYAWESSAGGSFTVRTDTGEPMGRGTKVILHLKEDQTEYLEERRIKEIVKKHSQFIGYPITLFVEKERDKEVSDDEAEEKEDKEEEKEKEEKKESED 254
HSP90AB1-1 DDEQYAWESSAGGSFTVRADHGEPIGRGTKVILHLKEDQTEYLEERRVKEVVKKHSQFIGYPITLYLEKEREKEISDDEAEEEKG---EKEEEDKDDEE 246
HSP90B1-1  NDTQHIWESDSNEFSVIADPRGNTLGRGTTITLVLKEEASDYLELDTIKNLVKKYSQFINFPIYVWSSKTETVEEP-----------MEEEEAAKEE 302
TRAP1-1    APGSLGYQWLSDGSGVFEIAEASGVRTGTKIIIHLKSDCKEFSSEARVRDVVTKYSNFVSFPLYLNGRR----------------------- 291

             ▼
HSP90AA1-1 KPEIEDVGSDEEEEKKDGDKKKKKKIKEKYIDQEELNKTKPIWTRNPDDITNEEYGEFYKSLTNDWEDHLAVKHFSVEGQLEFRALLFVPRRAPFDLFE 353
HSP90AB1-1 KPKIEDVGSDEEDDSGKDKKKKTKKIKEKYIDQEELNKTKPIWTRNPDDITQEEYGEFYKSLTNDWEDHLAVKHFSVEGQLEFRALLFIPRRAPFDLFE 345
HSP90B1-1  KEESDDEAAVEEEE--EEKKPKTKKVEKTVWDWELMNDIKPIWQRPSKEVEEDEYKAFYKSFSKESDDPMAYIHFTAEGEVTFKSILFVPTSAPRGLFD 399
TRAP1-1    ------------------------MNTLQAIWMMDPKDVGEWQHEEFYRYVAQAHDKPRYTLHYKTDAPLNIRSIFYVPDMKPSMFDV 355

                           ▼                         ▼  ▼
HSP90AA1-1 NR--KKKNNIKLYVRRVFIMDNCEELIPEYLNFIRGVVDSEDLPLNISREMLQQSKILKVIRKNLVKKCLELFTELAE-DKENYKKFYEQFSKNIKLGIH 450
HSP90AB1-1 NK--KKKNNIKLYVRRVFIMDSCDELIPEYLNFIRGVVDSEDLPLNISREMLQQSKILKVIRKNIVKKCLELFSELAE-DKENYKKFYEAFSKNLKLGIH 442
HSP90B1-1  EYGSKKSDYIKLYVRRVFITDDFHDMMPKYLNFVKGVVDSDDLPLNVSRETLQQHKLLKVIRKKLVRKTLDMIKKIAD-DKYN-DTFWKEFGTNIKLGVI 497
TRAP1-1    SR--ELGSSVALYSRKVLIQTKATDILPKWLRFIRGVVDSEDLPLNLSRELLQESALIRKLRDVLQQRLIKFFIDQSKKDAEKYAKFFEDYGLFMREGIV 453

HSP90AA1-1 ED--SQNRKKLSELLRYYTSASGD-EMVSLKDYCTRMKENQKHIYYITGETKDQVANSAFVERLRKHGLEVIYMIEPIDEYCVQQLKEFEGKTLVSVTKE 547
HSP90AB1-1 ED--STNRRRLSELLRYHTSQSGD-EMTSLSEYVSRMKETQKSIYYITGESKEQVANSAFVERVRKRGFEVVYMTEPIDEYCVQQLKEFDGKSLVSVTKE 539
HSP90B1-1  ED--HSNRTRLAKLLRFQSSHHPT-DITSLDQYVERMKEKQDKIYFMAGSSRKEAESSPFVERLLKKGYEVIYLTEPVDEYCIQALPEFDGKRFQNVAKE 594
TRAP1-1    TATEQEVKEDIAKLLRYESSALPSGQLTSLSEYASRMRAGTRNIYYLCAPNRHLAEHSPYYEAMKKKDTEVLFCFEQFQFDELTLLHLREEDKKKLISVETD 553

                                            middle domain                      charged dom 2
HSP90AA1-1 GLELPEDEEEKKKQE-----EKKTKFENLCKIMKD-ILEKKVEKVVVSNRLVTSPCCIVTSTYGWTANMERIMKAQALR---DNSTMGYMAAKKHLEI 636
HSP90AB1-1 GLELPEDEEEKKKME-----ESKAKFENLCKLMKE-ILDKKVEKVTISNRLVSSPCCIVTSTYGWTANMERIMKAQALR---DNSTMGYMMAKKHLEI 628
HSP90B1-1  GVKFDESEKTKESRE-----AVEKEFEPLLNWMKDKALKDKIEKAVVSQRLTESPCALVASQYGWSGNMERIMKAQAYQTGKDISTNYYASQKKTFEI 687
TRAP1-1    IVVDHYKEEKFEDRSPAAECLSEKETRELMAWMRN-VLGSRVTNVKVTLRLDTHPAMVTVLEMGAARHFLRMQQLAKTQ------EERAQLLQPTLEI 644

                                            C-terminal domain
HSP90AA1-1 NFDHSIIETLRQKAEADKNDKSVKDLVILLYYTALLSSGFSLEDPQTHANRIYRMIKLGLGIDEDDPTADDTSAAVTEEMPPLEGDDDTSRMEEVD--- 732
HSP90AB1-1 NPDHPIVETLRQKAEADKNDKAVKDLVVLLFYTALLSSGFSLEDPQTHSNRIYRMIKLGLGIDEDEVAAEEPNAAVPDEIPPLEGDEDASRMEEVD--- 724
HSP90B1-1  NPRHPLIRDMLRRIKEDEDDKTVLDLAVVLFYTYTLRSGYLLPDTKAYGDRIERMLRLSLNIDPDAKVEEEPEEEPEETAEDTTEDTEQDEDEEMD-20 803
TRAP1-1    NPRHALIKKLNQLRASEPG--LAQLLVDQIYENAMIAAG-LVDDPRAMVGRLNELLVKALERH---------------------------------- 704
```
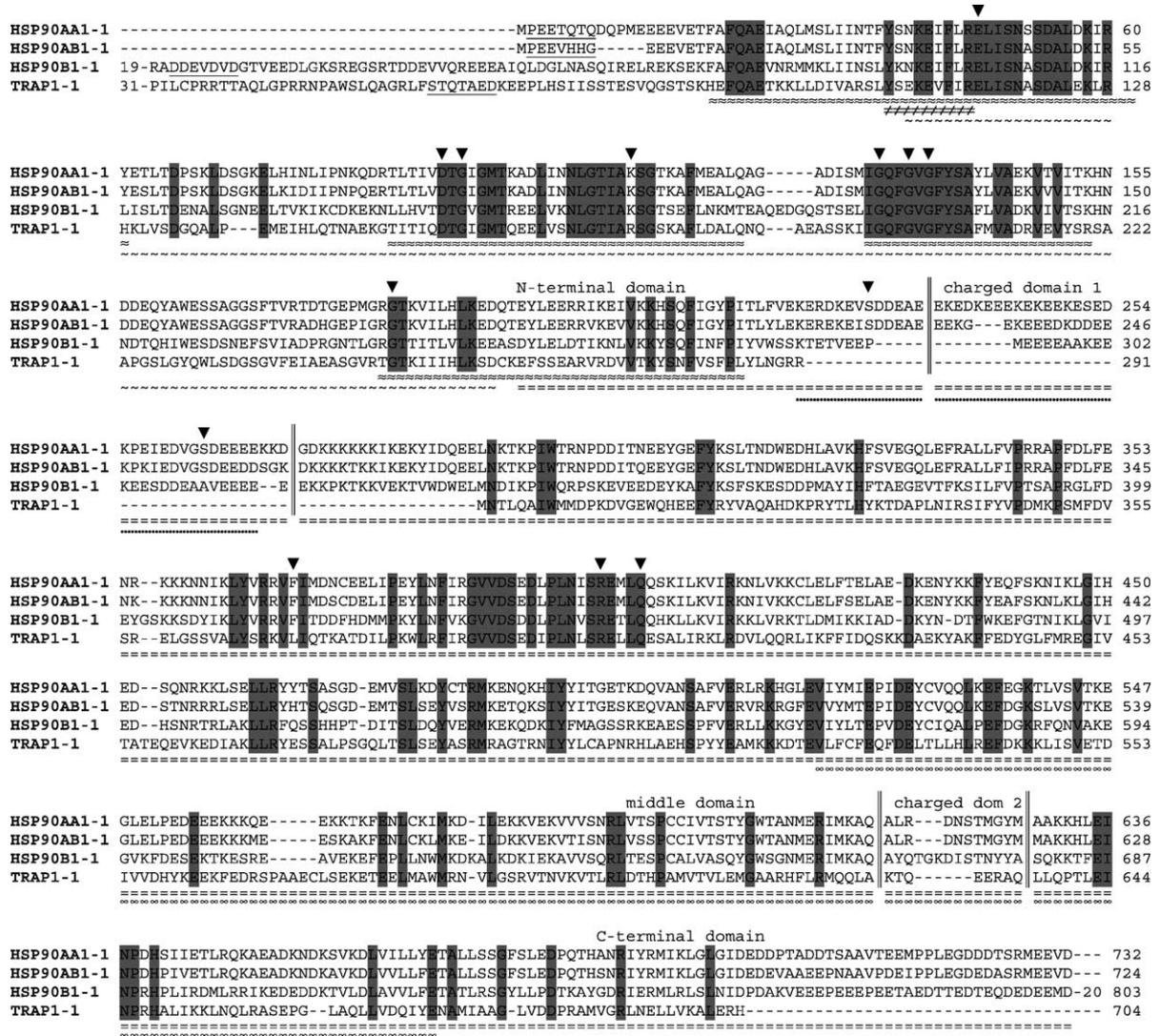
Fig. 3. Alignment of representative amino acid sequences in the human HSP90 family of genes, showing their domains, functionally important residues, and functional regions. Completely conserved amino acid residues are highlighted in gray. Gaps are marked with "-", and the last residue in each line is assigned a number. The domains are separated by thin black bars with the domain names above the alignment, and the cleavage sites are located just before the underlined residues. "▼" stands for the functionally important residues experimentally identified: E47 (refers to HSP90AA1−1) for ATP hydrolysis; D93 for ATP binding [22]; G95, G132, G135, G137, and G183 for both GA and p23 binding; K112 for GA binding [50]; R400 and Q404 for ATPase activity; F369 for interdomain interaction [24]; and S231 and S263 for phosphorylation by casein kinase II [51]. The conservative and functional regions are indicated by: "≈" and "≠" for HSP90 protein family signature; "=" for HSP90 protein; "~" for histidine kinase–like ATPases; "∞" for four–helical cytokine; and "..." for glutamic acid–rich region.

longer protein than HSP90AA1-1, whereas both HSP90AA1-3 and HSP90AA1-5 are shorter. HSP90N was found to be a chimeric gene with a role in the functional activation of *Raf* in transiently transfected cells [12]. HSP90AB1-2 does not terminate with MEEVD, whereas HSP90AB1-4 is shorter than HSP90AB1-1 in the N-terminus. There are 2 a.a. replacements between HSP90AB1-3 (3 mRNA only) and HSP90AB1-1.

HSP90B1-1 functions in the endoplasmic reticulum [31] and is known to participate in protein folding and assembly, in protein secretion, in protecting cells from undergoing apoptosis, and in mediating immunogenicity in tumor and virus-infected cells [1,32]. The highly conserved C-terminus sequence KDEL (residues 800–803) facilitates HSP90B1-1 retention in the ER [33]. HSP90B1-3 and HSP90B1-4 have the same splicing pattern as HSP90B1-1 but have 3 and 1 a.a.

differences, respectively. HSP90B1-2, however, has a completely different splicing pattern and a.a. sequence.

TRAP1-1 is found in the mitochondria, which is supported by the existence of a mitochondrial localization sequence in the N-terminus [6]. TRAP1-1 is also an ATP-binding protein and exhibits ATPase activity that is inhibited by both geldanamycin and radicicol, thus securing its place among the HSP90-like proteins [1,6]. For these reasons, we propose the alias of HSP90L for this gene. However, TRAP1-1 has functions that are distinct from those of other HSP90s [6]. These functional differences could be due to the lack of the MEEDV sequence in TRAP1-1 at its C-terminus [34,35]. TRAP1-5 and TRAP1-6 have the same splicing pattern as TRAP1-1, but they differ from it by a single a.a.. TRAP1-2 has a distinct splicing pattern, and a deletion relative to TRAP1-1, whereas TRAP1-3 has 5 a.a. replacements
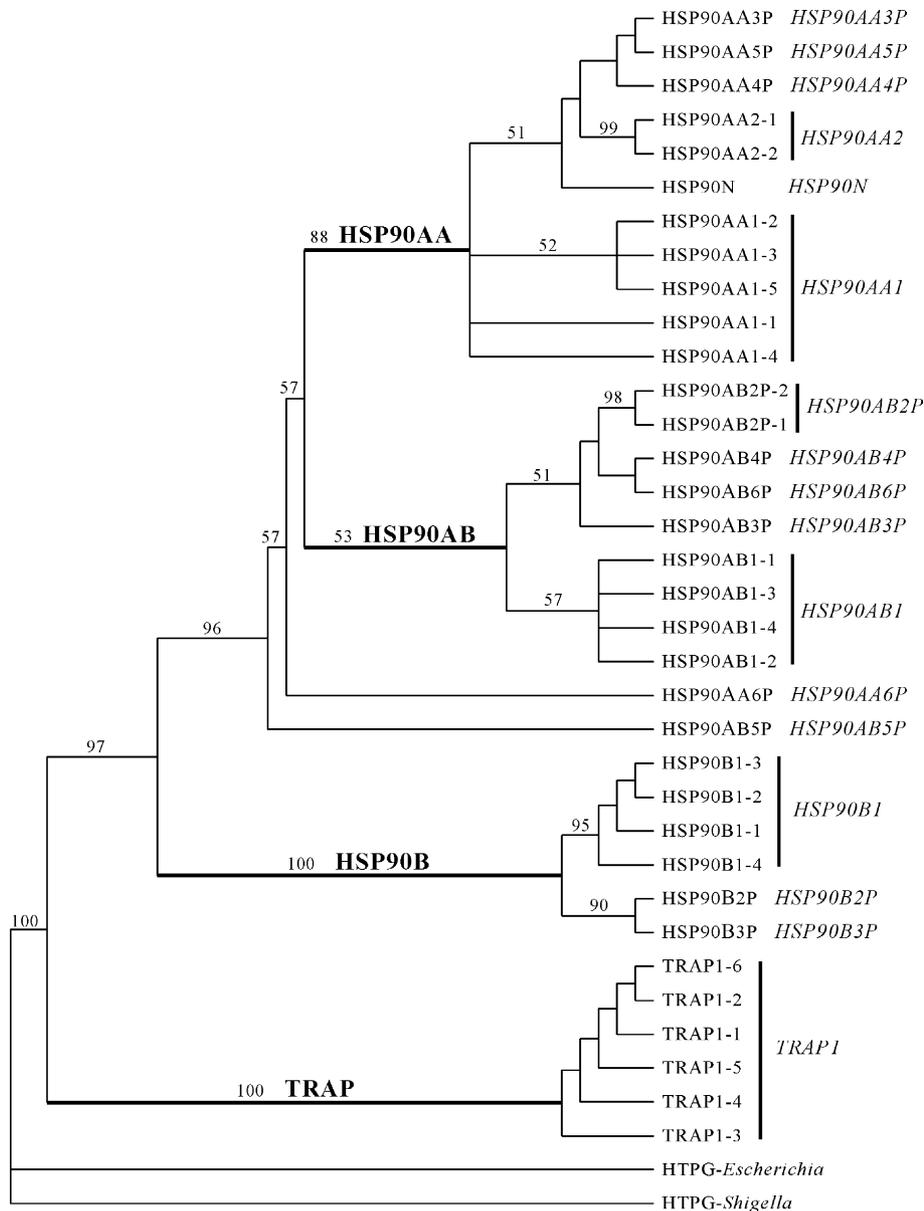
Fig. 4. Strict consensus of 618 equally most-parsimonious trees (2154 steps; see Results for parameter values) inferred from protein sequences. Bootstrap percentages of 1000 replicates are shown above the branches where they exceed 50%.

and 3 a.a. deletions. TRAP1-4 lacks exons 1–5. The function of these polymorphic and splice variants remains unknown.

Our phylogenetic analysis indicates that HSP90(AA+AB) and HSP90B (Grp94) are paralogous gene groups. A similar inference has been reached based on an analysis of 30 protein sequences of the HSP90 family from animals, plants, and fungi [14], in which the ancestral sequence to both HSP90(AA+AB) and HSP90B duplicated at a very early stage in the evolution of the eukaryotic cells. HSP90AA1-1 (HSP90-α) and HSP90AB1-1 (HSP90-β) sequences of human, chicken, and mouse group into two classes within each species, but the difference between them was markedly smaller than the difference among cytosolic HSP90 sequences across animals, plant, protists, and fungi [14]. Our results show that cytosolic HSP90 can be divided into two classes, HSP90AA and HSP90AB, but the monophyly of each class was not

significantly supported. Obviously the evolutionary relationship of these sequences remains an unresolved issue at the present time and a comparative genomic study based on different species would be necessary to elucidate it.

The present study shows that there is only one gene in the TRAP class in the human genome. Through GenBank searches, we found that *TRAP1* is also present in mouse, rat, chicken, fly, nematode, and the slime mold *Dictyostelium*, with all sequences possessing the characteristic mitochondrial localization sequence at their amino terminal. Curiously, TRAP1 has been localized also within the cytosol and the nucleus [36]. TRAP1 was proposed to be a distant eukaryotic relative of HSP90, resembling it both in size and in structural organization [34]. Bacterial HtpGs split into a group of mainly long-branching sequences (including TRAP1) and another comprising exclusively short-branching HtpG proteins, from
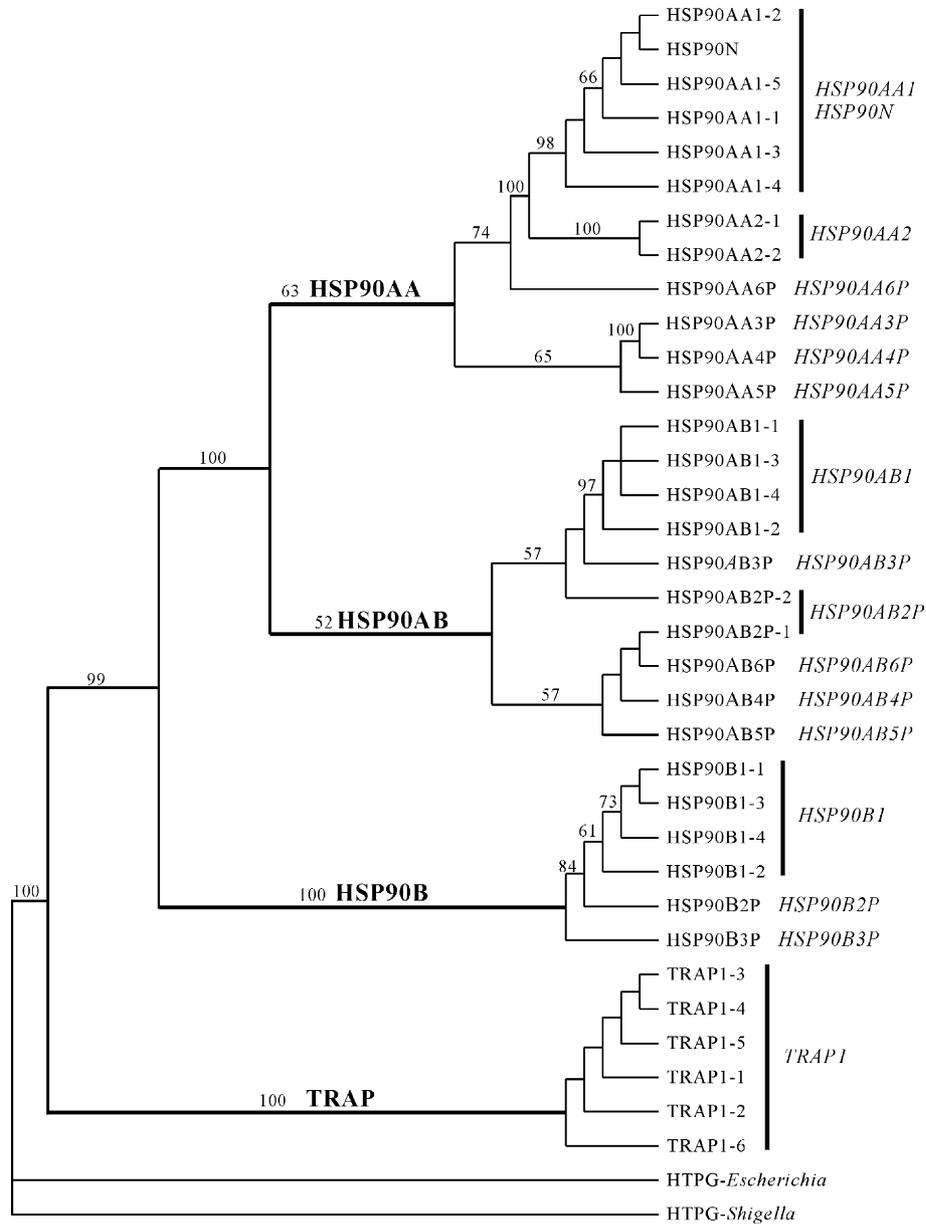
Fig. 5. Strict consensus of 512 equally most–parsimonious trees (5853 steps; see Results for parameter values) inferred from protein–coding nucleotide sequences. Bootstrap percentages of 1000 replicates are shown above the branches where they exceed 50%.

which the cytosolic/ER isoforms probably arose [8]. Our results indicate that mitochondrial TRAP1 and cytosolic/ER isoforms are well supported paralogous groups. Since there are only two HtpGs included in our study, we cannot speculate as to the mechanism of their origin.

Gene duplication is the main mechanism for the expansion of a gene family [37]. Eleven predicted HSP90 genes in the present study lack one or more exonic sequences from either end of a more complete member of that gene class, which possibly implies that the gene duplications for these genes are only partial. Interestingly, all 17 HSP90 genes are widely spread out in the genome, and no gene pair is immediately adjacent on the same chromosome. This indicates that most copies are relatively old, as young gene duplicates tend to produce genes that are adjacent on the same chromosome. *HSP90AA4P* and *HSP90AA6P*, and *HSP90AB2P* and *HSP90AB3P*, are all located

on chromosome 4 and could have duplicated via unequal sister chromatid exchange [37]. For other genes, both unequal crossing-over during meiosis and transposition (movement of genetic material from one chromosomal location to another via an RNA intermediate or DNA transposon) could be the most important mechanism for the gene duplication events [38]. However, further investigation is necessary to understand the mechanisms for the duplication events.

Materials and methods

Database searching and sequence retrieving

To find all putative HSP90 family members, we performed PSI-BLAST [39] searches of the human protein database at NCBI (http://www.ncbi.nlm. nih.gov) using the proteins of *Homo sapiens* HSP90AA1-1, HSP90AB1-1, HSP90B1-1 and TRAP1-1 (accession numbers in Table 1), *Arabidopsis*

thaliana chloroplast homologue AtHSP90-5 (AAD32922), and *Escherichia coli* HtpG (AAA23460) as queries, respectively. Each search resulted in a list of similar sequences, which was added to the next round of PSI-BLAST iteration searches, and each search continued until no new sequence with an alignment score above the default threshold was retrieved. The sequences returned by these queries were combined and all redundant sequence accession numbers were discarded. Nucleotide–nucleotide BLAST (BLASTN) searches using mRNA sequences corresponding to the query protein sequences mentioned above were carried out on the human nucleotide database at NCBI, the human gene database H-Inv DB (http://www.jbirc.aist.go.jp), and the genome database at TIGR (http://www.tigr.org) to obtain mRNA and new protein sequences of the gene family. All sequences were examined individually and aligned using Clustal X [40]. Protein and correspondent nucleotide sequences were matched by checking their sources in the databases and by inferred nucleotide translation. Any sequence with sequence identity lower than a predefined threshold for assigning homology [37], a curve that varied with sequence length was excluded from the study. Sequence identity and similarity were calculated using BioEdit v5.0 [41].

Gene identification and chromosomal location

To localize the members of the HSP90 gene family on the human genome assembly Build 35 version 1, we used MapViewer at NCBI. A TBLASTN search against the human genome assembly was applied to each protein sequence of the queries mentioned above and the BLASTN search was performed with each nucleotide sequence of these queries. The chromosomal location was recorded according to corresponding Map Elements obtained from search results. Subsequently, a 200-kb genome sequence flanking each hit or close hits was downloaded from the corresponding strand of the chromosome. The GENSCAN software [42] was used to identify the genes in each 200-kb sequence. This software uses both DNA and protein sequence homology information and predicts genes with greater accuracy than other gene prediction programs tested [43]. DNA and protein sequences of the predicted genes were separately aligned with the sequences of the six queries mentioned above, to remove genes other than those in the HSP90 family using the same threshold for assigning homology [37].

In order to map HSP90 gene family members onto chromosomes we used the BLAT software at UCSC [44] (http://genome.ucsc.edu). All genomic DNA sequences obtained from GENSCAN and all mRNA sequences retrieved from the databases were run on BLAT. All different mRNA sequences (or predicted coding sequences) that mapped to the same chromosomal site were considered to be different transcripts of a single gene. The nucleotide starting and ending locations of introns and exons of different transcripts were recorded for each gene in order to identify exon splicing patterns and total exon numbers. mRNA and amino acid sequences for each gene were compared to determine the number of silent and replacement polymorphisms.

Protein sequence properties

The size of the mature proteins, after removal of the predicted signal peptide, was calculated using both neural network (NN) and hidden Markov model (HMM) methods with the program SignalP v3.0 [45,46]. The cleavage site that separates the signal peptide from the rest of the protein (mature peptide) was determined by HMM, when both NN and HMM indicate the existence of a signal peptide with a probability larger than 50%. The molecular mass of the mature protein was calculated using BioEdit.

In order to identify biologically significant motifs and domains for each divergent protein sequence we used two different software programs. ScanProsite software (http://us.expasy.org/prosite) was used to search against the PROSITE database of protein families and domains (Release 18.40 of 22-Nov-2004) [47], whereas InterProScan [48] was used against InterPro, a database of protein domains and functional sites (http://www.ebi.ac.uk/interpro).

Phylogenetic analysis

In order to understand the evolutionary relationships of all the HSP90 copies in the human genome, we performed a phylogenetic analysis using PAUP* v4.0b8 [49]. Both protein and nucleotide alignments were produced using the "Profile Alignment Mode" in Clustal X, using the alignment of the six query sequences as a template. Maximum parsimony analysis was used for phylogenetic inference for both protein and nucleotide data sets. We performed a heuristic search employing stepwise addition with 200 random taxon addition sequence replicates and 10 trees held at each step. All characters were given equal weight and gaps were treated as "missing". The HtpG gene sequences from *E. coli* (nucleotide Accession No. M38777) and *Shigella flexneri* (protein Accession No. AE016979; nucleotide Accession No. AAP15950) were used as outgroups. The node support of a strict consensus tree of all equally most-parsimonious trees was assessed using 1000 bootstrap pseudo-replicates.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ygeno.2005.08.012.

References

[1] P. Csermely, T. Schnaider, C. Sõti, Z. Prohászka, G. Nardai, The 90-kDa molecular chaperone family: Structure, function, and clinical applications. A comprehensive review, Pharmacol. Ther. 79 (1998) 129–168.

[2] K. Nadeau, A. Das, C.T. Walsh, Hsp90 chaperonins possess ATPase activity and bind heat-shock transcription factors and peptidylprolyl isomerases, J. Biol. Chem. 268 (1993) 1479–1487.

[3] U. Jakob, J. Buchner, Assisting spontaneity—The role of Hsp90 and small HSPs as molecular chaperones, Trends Biochem. Sci. 19 (1994) 205–211.

[4] S. Rutherford, S. Lindquist, Hsp90 as a capacitor for morphological evolution, Nature 396 (1998) 336–342.

[5] A.S. Sreedhar, E. Kalmar, P. Csermely, Y.F. Shen, Hsp90 isoforms: Functions, expression and clinical importance, FEBS Lett. 562 (2004) 11–15.

[6] S.J. Felts, et al., The hsp90-related protein TRAP1 is a mitochondrial protein with distinct functional properties, J. Biol. Chem. 275 (2000) 3305–3312.

[7] P. Krishna, G. Gloor, The Hsp90 family of proteins in *Arabidopsis thaliana*, Cell Stress Chaperon. 6 (2001) 238–246.

[8] A. Stechmann, T. Cavalier-Smith, Evolutionary origins of Hsp90 chaperones and a deep paralogy in their bacterial ancestors, J. Eukaryot. Microbiol. 51 (2004) 364–373.

[9] N.F. Rebbe, J. Ware, R.M. Bertina, P. Modrich, D.W. Stafford, Nucleotide sequence of a cDNA for a member of the human 90-kDa heat-shock protein family, Gene 53 (1987) 235–245.

[10] T. Hoffmann, B. Hovemann, Heat-shock proteins, Hsp84 and Hsp86, of mice and men: Two related genes encode formerly identified tumour-specific transplantation antigens, Gene 74 (1988) 491–501.

[11] P.H. Krone, J.B. Sass, Hsp90α and Hsp90β genes are present in the zebrafish and are differentially regulated on developing embryos, Biochem. Biophys. Res. Commun. 204 (1994) 746–752.

[12] N. Grammatikakis, et al., The role of HSP90N, a new member of the Hsp90 family, in signal transduction and neoplastic transformation, J. Biol. Chem. 277 (2002) 8312–8320.

[13] S. Zolnierowicz, C. Work, K. Hutchison, I.H. Fox, Partial separation of platelet and placental adenosine receptors from adenosine A2-like binding protein, Mol. Pharmacol. 37 (1990) 554–559.

[14] R.S. Gupta, Phylogenetic analysis of the 90 kD heat shock family of

protein sequences and an examination of the relationship among animals, plants, and fungi species, Mol. Biol. Evol. 12 (1995) 1063–1073.

[15] V.V. Emelyanov, Phylogenetic relationship of organellar Hsp90 homologs reveal functional differences to organellar Hsp70 and Hsp60 evolution, Gene 299 (2002) 125–133.

[16] H.Y. Song, J.D. Dunbar, Y.X. Zhang, D. Guo, D.B. Donner, Identification of a protein with homology to hsp90 that binds the type 1 tumor necrosis factor receptor, J. Biol. Chem. 270 (1995) 3574–3581.

[17] C.F. Chen, et al., A new member of the Hsp90 family of molecular chaperones, interacts with the retinoblastoma protein during mitosis and after heat shock, Mol. Cell. Biol. 16 (1996) 4691–4699.

[18] K. Ozawa, Y. Murakami, T. Eki, E. Soeda, K. Yokoyama, Mapping of the gene family for human heat-shock protein 90α to chromosomes 1, 4, 11, and 14, Genomics 12 (1992) 214–220.

[19] A.S. Durkin, D.R. Maglott, N.C. Vamvakopoulos, H.Y. Zoghbi, W.C. Nierman, Assignment of an intron-containing human heat-shock protein gene (hsp90 β, *HSPCB*) to chromosome 6 near *TCTE1* (6p21) and two intronless pseudogenes to chromosomes 4 and 15 by polymerase chain reaction amplification from a panel of hybrid cell lines, Genomics 18 (1993) 452–454.

[20] R.G. Maki, R.L. Eddy, M. Byers, T.B. Shows, P.K. Srivastava, Mapping of the genes for human endoplasmic reticular heat shock protein gp96/grp94, Somat. Cell Mol. Genet. 19 (1993) 73–81.

[21] C.W. Schweinfest, et al., Cloning and sequence analysis of Hsp89-alpha-delta-N, a new member of the Hsp90 gene family, Biochem. Biophys. Acta 1398 (1998) 18–24.

[22] W.M.J. Obermann, H. Sondermann, A.A. Russo, N.P. Pavletich, F.U. Hartl, In vivo function of Hsp90 is dependent on ATP binding and ATP hydrolysis, J. Cell Biol. 143 (1998) 901–910.

[23] C.E. Stebbins, et al., Crystal structure of an hsp90-geldanamycin complex: Targeting of a protein chaperone by an antitumor agent, Cell 89 (1997) 239–250.

[24] P. Meyer, et al., Structural and functional analysis of the middle segment of Hsp90: Implications for ATP hydrolysis and client protein and cochaperone interactions, Mol. Cell 11 (2003) 647–658.

[25] D. Johnson, et al., Hsp90 chaperone activity requires the full-length protein and interaction among its multiple domains, J. Biol. Chem. 275 (2000) 32499–32507.

[26] S. Lindquist, The heat-shock response, Annu. Rev. Biochem. 55 (1986) 1151–1191.

[27] L.A. Sonna, J. Fujita, S.L. Gaffin, C.M. Lilly, Molecular biology of thermoregulation: Effects of heat and cold stress on mammalian gene expression, J. Appl. Physiol. 92 (2002) 1725–1742.

[28] H.J. Yost, S. Lindaquist, NA splicing is interrupted by heat shock and is rescued by heat shock protein synthesis, Cell 45 (1986) 185–193.

[29] T. Imanishi, et al., Integrative annotation of 21,037 human genes validated by full-length cDNA clones, PloS Biol. 2 (2004) 856–875.

[30] International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome, Nature 409 (2001) 860–921.

[31] C.V. Nicchitta, Biochemical, cell biological and immunological issues surrounding the endoplasmic reticulum chaperone GRP94/gp96, Curr. Opin. Immunol. 10 (1998) 103–109.

[32] Y. Tamura, P. Peng, K. Liu, M. Daou, P.K. Srivastava, Immunotherapy of tumors with autologous tumor-derived heat shock protein preparations, Science 278 (1997) 117–120.

[33] S.C. Chang, A.E. Erwin, A.S. Lee, Glucose-regulated protein (Grp94 and Grp78) genes share common regulatory domains and are coordinately regulated by common trans-acting factors, Mol. Cell. Biol. 9 (1989) 2153–2162.

[34] S.Y. Chen, W.P. Sullivan, D.O. Toft, D.F. Smith, Differential interactions of p23 and the TPR-containing proteins Hop, Cyp40, FKBP52 and FKBP51 with Hsp90 mutants, Cell Stress Chaperon. 3 (1998) 118–129.

[35] J.C. Young, W.M.J. Obermann, F.U. Hartl, Specific binding of tetratricopeptide repeat proteins to the C-terminal 12-kDa domain of hsp90, J. Biol. Chem. 273 (1998) 18007–18010.

[36] T. Morita, A. Amagai, Y. Maeda, Unique behavior of a *Dictyostelium* homolog of TRAP1, coupling with differentiation of *D. discoideum* cells, Exp. Cell Res. 280 (2002) 45–54.

[37] C.A. Orengo, D.T. Jones, J.M. Thornton, Bioinformatics: Genes, proteins & computers, BIOS Scientific Publisher Ltd, Oxford, UK, 2003.

[38] J.Z. Zhang, Evolution by gene duplication: An update, Trends Ecol. Evol. 18 (2003) 292–298.

[39] S.F. Altschul, et al., Gapped BLAST and PSI-BLAST: A new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.

[40] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res. 25 (1997) 4876–4882.

[41] T.A. Hall, BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucleic Acids Symp. Ser. 41 (1999) 95–98.

[42] C.B. Burge, S. Karlin, Finding the genes in genomic DNA, Curr. Opin. Struct. Biol. 8 (1998) 346–354.

[43] M. Burset, R. Guigó, Evaluation of gene structure prediction programs, Genomics 34 (1996) 353–367.

[44] W.J. Kent, BLAT—The BLAST-like alignment tool, Genome Res. 12 (2002) 656–664.

[45] H. Nielsen, J. Engelbrecht, S. Brunak, G. von Heijne, Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites, Protein Eng. 10 (1997) 1–6.

[46] J.D. Bendtsen, H. Nielsen, G. von Heijne, S. Brunak, Improved prediction of signal peptides: SignalP 3.0, J. Mol. Biol. 340 (2004) 783–795.

[47] N. Hulo, et al., Recent improvements to the PROSITE database, Nucleic Acids Res. 32 (2004) D134–D137.

[48] E.M. Zdobnov, R. Apweiler, InterProScan—An integration platform for the signature-recognition methods in InterPro, Bioinformatics 17 (2001) 847–848.

[49] D. Swofford, PAUP*: Phylogenetic analysis using parsimony (and other methods), version 4.0b8, Sinauer Associates, Inc., Sunderland, MA, 2001.

[50] J.P. Grenert, et al., The amino-terminal domain of heat shock protein 90 (hsp90) that binds geldanamycin is an ATP/ADP switch domain that regulates hsp90 conformation, J. Biol. Chem. 272 (1997) 23843–23850.

[51] S.P. Lees-Miller, C.W. Anderson, Two human 90-kDa heat shock proteins are phosphorylated in vivo at conserved serines that are phosphorylated in vitro by casein kinase II, J. Biol. Chem. 264 (1989) 2431–2437.