

Phylogenetic Analysis of the 90 kD Heat Shock Family of Protein Sequences and an Examination of the Relationship among Animals, Plants, and Fungi Species

Radhey S. Gupta

Department of Biochemistry, McMaster University

The heat shock protein (Hsp) sequences, because of their ubiquity and high degree of conservation, provide useful models for phylogenetic analysis. In this paper I have carried out a global alignment of all available sequences (a total of 31) for the 90-kD heat shock protein (Hsp90) family. The minimum amino acid identity that is seen between presently known Hsp90 homologs is about 40% over the entire length, indicating that it is a highly conserved protein. Based on the alignment, a number of signature sequences that either are distinctive of the Hsp90 family or that distinguish between the cytosolic and the endoplasmic reticular forms of Hsp90 have been identified. Detailed phylogenetic analyses based on Hsp90 sequences reported here strongly indicate that the cytosolic and the endoplasmic reticulum (ER) resident forms of Hsp90 constitute paralogous gene families which arose by a gene duplication event that took place very early in the evolution of eukaryotic cells. A minimum of two additional gene duplication events, which took place at a later time, are required to explain the presence of two different forms of Hsp90 that are found in fungi and vertebrate species. In a consensus neighbor-joining bootstrap tree based on Hsp90 sequences, plants and animals species grouped together 989 times of 1,000 (a highly significant score), indicating a closer relationship between them as compared to fungi. A closer affiliation of plant and animal species was also observed in the maximum-parsimony tree, although the relationship was not significantly supported by this method. A survey of the recent literature on this subject indicates that depending on the protein sequence and the methods of phylogenetic analysis, the animal species are indicated as closer relatives to either plants or fungi with significant statistical support for both topologies. Thus the relationship among the animal, plant, and fungi kingdoms remains an unresolved issue at the present time.

The term *heat shock proteins* (Hsps) refers to a group of proteins whose synthesis is enhanced upon sudden increase in temperature or exposure to a variety of other stressors (Lindquist and Craig 1988; Morimoto et al. 1994). The major Hsps in cells have the approximate sizes of 60 kilodaltons (kD), 70 kD, and 90 kD and are referred to as Hsp60, Hsp70, and Hsp90, respectively. These proteins perform essential molecular chaperone functions in cells in facilitating proper interaction between various other proteins and in the intracellular transport of proteins to various destinations (see Lindquist and Craig 1988; Morimoto et al. 1994). Hsps, due to their ubiquitous presence in all species and high degree of sequence conservation, also provide very useful model systems for evolutionary studies (see Gupta and Golding 1993; Boorstein et al. 1994; Gupta 1995). We

have recently reported detailed phylogenetic analyses based on Hsp60 and Hsp70 families of sequences which has provided novel insight into the origin of eukaryotic cells (Gupta and Singh 1994; Gupta et al. 1994; Gupta 1995).

The members of the 90-kD Hsp family are the least understood of the major Hsps in terms of their cellular function. The known functions of these proteins include modulating the activities of the steroid hormone receptors, tyrosine kinases and serine-threonine kinases (Lindquist and Craig 1988; Pratt 1993; Bohen and Yamamoto 1994). In addition, Hsp90 is also found associated with cytoskeletal proteins such as actin and tubulin (see Bohen and Yamamoto 1994). In eukaryotic cells, specific Hsp90 homologs exist in the cytosol and the ER. Hsp90 homologs have also been cloned and sequenced from bacterial species (referred to as HptG) (Bardwell and Craig 1987). In the present paper, I have carried out a global alignment of all available Hsp90 sequences and used this alignment to deduce the evolutionary relationships between different homologs and species. These studies indicate that the cytosolic and ER Hsp90 homologs constitute paralogous families which diverged from each other at a very early stage in the

Key words: Hsp90, protein phylogeny, gene duplication, endoplasmic reticulum, phylogenetic methods.

Address for correspondence and reprints: Radhey S. Gupta, Department of Biochemistry, McMaster, Hamilton, Ontario L8N 3Z5, Canada; E-mail: GuptaR@mcmaster.cis.mcmaster.ca.

Mol. Biol. Evol. 12(6):1063–1073. 1995.
© 1995 by The University of Chicago. All rights reserved.
0737-4038/95/1206-0009\$02.00

Table 1
Hsp90 Sequences Analyzed in This Study

Species	Accession Number	Reference
Cytosolic:		
Human α	SP/Po7900	Yamazaki et al. (1989)
Mouse α	SP/P07901	Moore et al. (1989)
Chicken α	SP/P11501	Binart et al. (1989)
Human β	SP/P08238	Rebbe et al. (1987)
Mouse β	GB/M36829	Hoffmann and Hovemann (1988)
Rat β	GB/S45392	McGuire et al (1992)
Chicken β	SP/Q04619	Meng et al. (1993)
<i>Drosophila melanogaster</i>	SP/P02828	Blackman and Meselson (1986)
Maize	GB/S59780	Marrs et al. (1993)
<i>Arabidopsis thaliana</i>	SP/P27323	Conner et al. (1990)
<i>Pharbitis nil</i>	GB/M99431	Genbank*
Rice	SP/P33126	Swiss prot.*
Tomato	GB/M96549	Genbank*
<i>Trypanosoma brucei</i>	SP/P12861	Mottram et al. (1989)
<i>Trypanosoma cruzi</i>	SP/P06660	Dragon et al. (1987)
<i>Leishmania amazonensis</i>	SP/P27741	Shapria and Pedraza (1990)
<i>Plasmodium falciparum</i>	GB/L34027	GenBank*
<i>Theileria parva</i>	SP/P24724	GenBank*
<i>Saccharomyces cerevisiae</i> (c)	SP/P15108	Borkovich et al (1989)
<i>S. cerevisiae</i> (h)	SP/P02829	Farrelly and Finkenstein (1984)
<i>Ajellomyces capsulata</i>	GB/S21764	Minchoitti et al. (1992)
<i>Histoplasma capsulatum</i>	GB/M55629	Minchoitti et al. (1991)
ER Homologs:		
Human (e)	SP/P24625	Maki et al. (1990)
Mouse (e)	SP/P08113	Mazzarella and Green (1987)
Dog (e)	GB/U01153	Cala and Jones (1994)
Pig (e)	GB/X76301	GenBank*
Chicken (e)	SP/P08110	Kulomaa et al. (1986)
Barley (e)	EM/S31862	EMBL*
<i>C. roseus</i> (e)	GB/L14594	Schroder et al. (1993)
<i>Secale cereale</i>	GB/Z30243	GenBank*
Bacteria:		
<i>Escherichia coli</i>	SP/P10413	Bardwell and Craig (1987)

NOTE.—The abbreviations SP, GB, and EM indicate Swissprot, GenBank, and EMBL databases, respectively. The superscript asterisk (*) denotes that the sequence deposited in the databank is not published as yet. The letters c, h, and e in parentheses indicate cognate, heat-induced, and ER-resident forms of Hsp90, respectively.

evolution of eukaryotic cells. A number of other gene duplication events which took place at later times in specific branches of the eukaryotic tree are also identified. In the evolutionary tree based on Hsp90 sequences, the animal and plant species are indicated as closer relatives in comparison to fungi, which is contrary to the inference reached in some of the recent studies (Baldauf and Palmer 1993; Hasegawa et al. 1993; Wainright et al. 1993). The evolutionary relationship among animals, plant, and fungi species is discussed in light of the available results.

Material and Methods

Sequence Data and Phylogenetic Analysis

The Hsp90 sequences from various species were obtained from the GenBank, Swissprot, and Protein Identification Resource databases. The species names,

database accession numbers, and the literature reference to the cloning and sequencing of various sequences are given in table 1. The amino acid identity between pairs of protein sequences was determined using the PALIGN program (Myers and Miller 1988) of the PCGene sequence analysis package (Intelligenetics Inc., Mountain View, California) using the structure gene matrix, and the unit gap and open gap costs of two and seven, respectively. The multiple alignment of sequences was initially carried out using the CLUSTAL program (Higgins and Sharp 1988) of the PCGene package. This alignment was then modified manually to correct for any misalignment as determined by the results of pairwise alignments and by visual inspection. Phylogenetic analysis was carried out on the sequence regions which could be aligned without ambiguity in all homologs. For such purposes

a small part of the sequence from the C- and N-terminal ends (outside of the arrows in fig. 1) and a variable region from the middle (boxed in fig. 1) were omitted from consideration. The remainder of the sequence alignment which consisted of 620 aligned amino acid positions was employed for phylogenetic analysis. The phylogenetic analysis on the sequence data was carried out using the various programs (namely, PROTDIST, NEIGHBOR, BOOT, CONSENSE, PROTPARS, PROTML, etc.) from the PHYLIP version 3.5 program package (Saitou and Nei 1987; Kishino et al. 1990; Felsenstein 1985, 1991).

Results

Table 1 lists the various Hsp90 homologs for which the complete sequence information is currently available in the databases. These homologs were identified by carrying out blast searches using sequence data for *Escherichia coli*, yeast, and human Hsp90 sequences. In addition to these sequences, partial sequence information is available for a number of other homologs (not shown), which were not considered in the present investigation.

Because of the slight variation in the relative molecular masses (Mr) of these proteins in different studies (between 80 and 90 kDa), these homologs have been referred to by different names in the literature (e.g., Hsp80, Hsp82, Hsp83, Hsp84, Hsp85, Hsp86, Hsp90, etc.; see table 1). The bacterial homologs are referred to as HptG, and the endoplasmic reticular forms are often designated as Grp90 or Grp94. However, in the present study the common term Hsp90 will be used to describe all of these homologs. In eukaryotic species distinct Hsp90 homologs are found in the cytosol and in the ER, but thus far no homolog that is specific for the organelles (namely mitochondria or chloroplasts) has been identified. In vertebrate species two different cytosolic forms of Hsp90 (referred to as α and β in this study) have been reported. Both these forms are moderately heat inducible; however, how these two forms differ in terms of their physiological function is presently not clear (see Lindquist and Craig 1988; Hardesty and Kramer 1989). The yeast *Saccharomyces cerevisiae* also contains two cytosolic Hsp90 homologs; of these one is constitutively expressed, whereas the other is highly heat inducible (Lindquist and Craig 1988).

Table 2 presents a pairwise similarity matrix for Hsp90 sequences from representative cytosolic and ER homologs and the sole bacterial species (*E. coli*) for which the complete sequence is known. The minimum amino acid similarity that is observed between any two Hsp90 homologs over their entire length (e.g., between *E. coli* and various eukaryotic homologs) is about 40% identical residues plus an additional 16%–20% conservative amino acid replacement, indicating the highly

conserved nature of this protein. The α and β group of cytosolic homologs in vertebrate species show about 87% identity to each other. The eukaryotic cytosolic homologs show much higher identity to each other (minimum identity between any two members \approx 61%) than to the homologs from the same species that are found in the ER (maximum amino acid identity seen between a cytosolic and an ER homolog is about 50%; table 2). For the ER Hsp90, the homologs from the plant and animal species showed much lower identity to each other (\approx 50% identity) as compared to that seen between the cytosolic homologs (67%–70% identity).

A global alignment of Hsp90 sequences, carried out as described in Material and Methods, is presented in figure 1. The sequences from some species that are closely related to the others presented here are not shown in the alignment due to space consideration. From the alignment presented in figure 1, a number of inferences regarding Hsp90 sequences could be drawn. First, several sequence regions are highly conserved in all Hsp90 homologs. A few of the highly conserved segments—NKEIFLRELISN(S/A)SDALDKIR, LGTIA(K/R)SGT, IGQFGVGFYSA(Y/F)LVA(E/D), IKLY-VRRVFI, and GVVDS(E/D)DLPLN(I/V)SRE (marked I–V in fig. 1)—could be used as signature sequences for identifying this family of proteins.

Second, all of the ER Hsp90 homologs contain an N-terminal signal sequence characteristic of the ER-targeted proteins. The various ER homologs also contain the ER retention sequence KDEL at their extreme C-terminal end (Pelham 1989). In contrast to the ER homologs, all of the cytosolic Hsp90 proteins end with the sequence MEEVD on their C terminus. A similar sequence (V/I)EEVD is also found at the C-terminal ends of the majority of the cytosolic Hsp70 homologs (Boorstein et al. 1994; Gupta et al. 1994). The functional significance of this conserved sequence feature is not known at present.

Third, the Hsp90 homolog from *Secale cereale* appears different from other eukaryotic cytosolic or ER homologs. This particular protein lacks sequence features that are present in both these groups, and it also shows a much lower degree of similarity to the other Hsp90 homologs (identity values in the range of 42%–46%). Although an N-terminal leader sequence is present in this protein, no similarity is observed between this sequence and the signal sequence of other plants' ER Hsp90 sequences. It is possible that the protein product of this homolog may be targeted to a different cellular compartment (chloroplast or mitochondria).

Finally, the eukaryotic Hsp90 contains a region of variable length (boxed in fig. 1) which is rich in acidic residues. This region is lacking in the *E. coli* homolog.

Human α PEETQTQDPQMEEEE
 Chicken α M--AV-----
 Human β ---VHH G---
 Chicken β M--QV-H G-D-
 D.melanogaster M---
 Maize MASADVHMAGGAAE
 A.thaliana MADVQMAD
 Tomato MSD
 Pl.falciparum MS
 L.amazonensis M
 T.brucei M
 S.cerevisiae(h) MA
 H.capsulatum MS
 Chicken (e) MKSAWALACTLLLAASVTAE EVDVDATVEEDLGKSGREGSRDTDEVVQREEEAIQLDGLNASQIKEIREK
 Human (e) MRALWVLGCVLLTFGVSRADEVDVDTVEEDLGKSGREGSRDTDEVVQREEEAIQLDGLNASQIRELREK
 Dog (e) MRALWVLGCVLLTFGVSRADEVDVDTVEEDLGKSGREGSRDTDEVVQREEEAIQLDGLNASQIRELREK
 Barley (e) MRK WALSCALLVLLLTTPDPAKKLQVNAEESSEVDGFPKVEEKLGAHPGLSTDEVVQRESEISRKTLRNSA K-E---VSR-DI---SL---D
 Se.cereale MAPALSRTLGPSSVAALRPSPSRGLPLAALLPQGRSSSARGVRWEAGRGLVGARCASAVAETKAGEEEAA G-K-EY---VSR-D-D-VHSL-H--
 E.coli MKG Q-RG--S-VK--LH-M-HSL

I II III

Human α IFRLRELISNSSDALDKIRYETLTDPSKLDGK LHLINLPNKQDRTLITVDTGIMTKADLINNLGTIAKSGTKAFMEAL QAGADISMIQGFGVGF
 Chicken α -----S-----D K-----H-----V-----S
 Human β -----A-----S -K-DI---P-E---L-----
 Chicken β -----A-----S -K-DIV--PR-P--LL-----V
 D.melanogaster -----A-----S -Y-K---TAG--I---S-V-----
 Maize -----A-----F-S-K---AQP -F-R-V-D-ASK-S-I-S-V-S-V-R-E---A---V
 A.thaliana -----A-----F-S-K---GQP -F-R-V-D-ANK-S-I-S-V-R-E---V
 Tomato -----A-----F-S-K---GQP -F-HI--D-ANN--I-S-V-R-E---A---V
 Pl.falciparum -----A-----SI--TQ-SAEP FF-RI--D-TNN--E-S-N-----I -SG
 L.amazonensis -----DV--A-C-QS--V-GDATR -CVRV-D-ENK--VE-N-----V-R-E---E---M
 T.brucei -----A-----C-QS--NQ-V-GDEPH -R-RV--DRVNK--VE-S-----V-R-S---E-G-M
 S.cerevisiae(h) -----A-----KS-S-KQ-ETEPD -F-RIT-KPEQKV-E-R-S---E-----S-V
 H.capsulatum -----F-----KA-S-S-V-E-D-D -R-DIT-D-ENK--R-----V-R-Q---T
 Chicken (e) -----A-----LIS--ENA-AGNE -TVKIKCD-EKNM-HVT---EE--K---SE-LNKMTMQDQSQSTSEL
 Human (e) -----A-----LIS--ENA-SGNE -TVKIKCD-EKNL-HVT--V-REE-VK---SE-LNKMTAEAGDQGSTSEL
 Dog (e) -----A-----LIS--ENA-AGNE -TVKIKCD-EKNL-HVT--V-REE-VK---SE-LNKMTAEAGDQGSTSEL
 Barley (e) -----A-----FLA--KEVMGE-DTAK -E-QIKLD-ENKI-S-R-R-V-E-K-K---S-V-KM -T-G-LNL
 Se.cereale V---V-A-A-L-FLSV-S-V-AD-G -E-RIK-DPDAG-I-T-S---DE-KDC---Q--SK-LK--DL--NGL
 E.coli -----A--A-L-FRA-SN-DLYEGDG -RVRVSFD-DK---S-N-V--RDEV-DH---S-L-S-GS QKAK-SQL

IIII

Human α YSAYLVAEKVTVITK HNDDEQY AWESSAGGSFTV RTDTGEPHGRGTVKVLHLKEDQTEYLEERRIKVKKHSQFIGYPTLFEKERDKEVSDDEAE
 Chicken α -----V-R----- -L-N--L-----R-----
 Human β -----V----- -A-H--I-M-----V-V-----YL--E-I
 Chicken β -----D--TS-N----- -H--I-----Y-----Y-E
 D.melanogaster -----D--TS-N----- -A-NS-L-----IV-YI--D--SK--N-----K-L--E-----DD
 Maize -----DR-M-T----- V--Q-----THDT--QL-----IT-F-D-L-----L-DL--E-S--Y-WT--TTE-I--E-E
 A.thaliana -----V-T----- V--Q-----T-DVD--L-----IS-F-D-L-----L-DL--E-S--Y-WT--TTE-I--D-D
 Tomato -----V-T----- V--Q-----T-DTS-NL-----MV-Y--L-----L-DLI--E-S--S-W--TIE-I--E-E
 Pl.falciparum -----DH-V-S-N----- V--A-----TKDE-N-KL-----I-----L-K--DL--E-SF-K-YC-RQNE-ITAS-E-E
 L.amazonensis -----DR--TS-N-S-V----- V-----T-ITSAPED--KLPARIT--L-A--L-LI--E--D-E-M--TTE--T-EDE
 T.brucei -----DR--VS-N-E-DA----- T-----T-TS-PCDC LK--RIV--Q-----L-DLI--E--D-E-M--NTTE--T-EEDK
 S.cerevisiae(h) -----LF--DR-Q-S----- I--N-----TLDEVN-RI--ILR-F-D-L--K-----VI-R-E-VA--Q-V-T-VE--PIP-E-K
 H.capsulatum -----D--S----- I--N--T-K-TQD-D-RAI--M--DE--N-SK--V--Q-E-F--Y-H-L-NE--P-ED
 Chicken (e) -----F--DR-I-TS--N-T-H I--DSNE -S-IDDR-NL-----TIT-V--EASD--LDTV-NL--Y--NF--YVWSS-TETV-EPVE-E-A
 Human (e) -----F--D-I-TS--N-T-H I--DSNE -S-IADPR-NL-----TIT-V--EASD--LDT-NL--Y--NF--YVWSS-TETV-EPME-E-A
 Dog (e) -----F--D-I-TS--N-T-H I--DSNE -S-IADPR-NL-----TIT-V--EASD--LDT-NL--Y--NF--YVWSS-TETV-EPME-E-A
 Barley (e) -----V--DY-E-VS-----K V--K-D--AISED-TW--L-EIK--RDEAK--GKL-DL--Y-E-NF--Y-WAT-V-V-P-A-E-S
 Se.cereale -----F--V-S-SPT-K-I--AE-NS-YVIREETDP-K-T--QIT-F-R-DK-FADPA--QGL-NY--VSF--FTWQ--S-TV--E-E-S
 E.coli -----FI--D--R-RAAGEKPENGVF--AGE-EY--ADITKD ---EIT--R-GED-F-DDW-VRS-IS-Y-DH-AL-VEIE

KEDKEEKEKEKESEDKPEIEDVGSDEEEEEKKDG DKKKKKKIKKYIDQEELNKTKPIWTRNPDDITNEEYGEFYKSLTNDW EDHLAVKHFSVEGQLE
 Chicken α ---EEK---EKTEDKPEIEDVGSDEEE---
 Human β EKGEK--ED--DDE-KPKIEDVGSDEE-D SGKD K---T-----Q-----
 Chicken β EKVEK--E-SKDE-KPKIEDVGSDEEE--- GEKS K---T-----Q-----
 D.melanogaster EKKEGD--KEM-TDEPKIEDVGEDEADK -DKD A---T-----TED---SQ
 Maize DNK--GDV--VDD--DTDK KS-----V-VSHEWVQI--Q---L-K-EE--RD--AS
 A.thaliana EPK--N-G-V--VDE-KEK ---G-----VSHEW-LI--Q---L-K-EE--K-SAA
 Tomato EKKD--GKV--VDE-KEK EE-----V-VSHEWVQI--Q---M-K-EE--K-AA -E
 Pl.falciparum *E-EE-KTG-KDKA-ESKE-NEDEEKK-DEE-D KE-R---HTVEHEW--L--L-M-K-EEV--AS
 L.amazonensis AKKAD-DG-EPKV-EVTEG-EG K---T--V--VTKEV-VQ--H--L--D-K-V-K--AA--AIS-- --PP-T
 T.brucei EAA-KA-EGE-P-VE-V-DGVDADA K---T--V--VKGEFVWQ--H--L--D-K-V-K--AA--AIS-- --EQ-ST
 S.cerevisiae(h) -DEEKKDE--KDDEKKPKLE-VDEEE K-P-T--V--EVQEI-----L-S--Q--NA--IS-- --P-Y
 H.capsulatum VKEDGDD-APKVE-VDEE-DKTKAKK K-A-T--NK-EE-----A--Q--AS--
 Chicken (e) -E---TDDN-AAVE-EEE K-P-T--VEKTVW-W-LM-DI---Q-PSKEVED--KA---TFSKEH D-PM-YI--TA--EVT
 Human (e) AKEEK--SDD-AAVE-EEE-E K-P-T--VEKTVW-W-LM-DI---Q-PSKEVED--KA---FSKES D-PM-YI--TA--EVT
 Dog (e) AKEEK-DSDD-AAVE-EEE-E K-P-T--VEKTVW-W-LM-DI---Q-PSKEVED--KA---FSKES D-PM-YI--TA--EVT
 Barley (e) N-EE-STT-TT-E-ETEDD-E K-P-T-TV--TTTEW-L--DM-AV-L-S-KEV-E--AK--H-AK-FGDDKPMWS--A--DV
 Se.cereale --GE-TAEGEK-E -T-T--W-W-LA-E--M--KEVEET--N--KAF-EF L-P--HA--TT--EV
 E.coli -REE-DG-TV-SW-KI--AAQAL--KSF--D--K--NIAH-F N-P--TWS--NR--KO

FIG. 1.—Alignment of Hsp90 sequences from representative species and homologs. Amino acid residues that are identical to that present in the top row are denoted by dashes. The Roman numerals I–V identify highly conserved regions which could be used as signature sequences for Hsp90 homologs. The shaded regions identify the distinctive C-terminii sequences for the cytosolic and ER Hsp90 homologs. The arrows mark the outside boundaries of the sequence region which was employed in phylogenetic analysis. The box identifies an acidic variable region

	IV	V	
Human α	FRALLFVPRAPFDLFEN	RKKKNKIKLYVRRVIMDNCEELIPEYLNFRGVDSDELPLNISREMLQQSKILKVRKNLVKKCLELFTLAED	KENYKFFY
Chicken α	-----I-----	-----M-----	-----S-----
Human β	-----I-----	-----S-D-----	-----I-----S-----
Chicken β	-----I-----	-----S-D-----	-----I-----S-----
D.melanogaster	-----I---T-----	Q-R-----D-----MK-----N-V-----TM-IE-T--	-----
Maize	-K-I-----DT-----	-----W-G-VK-----D-----T-----N-----I-M-F-I-N	-DD-A--
A.thaliana	-K-I-----K-----DT	-----S-VK-----D-----T-----N-----I-M-N-I-N	-D-T--
Tomato	-K-V---K-----DT	K-P-----D-----S-VK-I-----T-----N-----V---F-I-N	-D-N--
Pl.falciparum	-K---I-K---M---	-----R-----D---I---W---VK-----S---N---K---I---DM-S---N	-----
L.amazonensis	---IM---K---ML-P	N-R-----D-C-DW-G-VK-----N---N-----I---M-E-V-N	-D-Q--
T.brucei	---I---L-K---M-P	N-R-----D-C-W-G-L-----N---N-----I---A-A-E---GN	-D-----
S.cerevisiae(h)	---I---I-K---S	K-----T-EA-D---W-S-VK-----L-----N-M---I---LI-A-N-I---	S-QFE--
H.capsulatum	---I---K---T	K-T-----T-DATD---W-S-K-----GL-T---N-M---I---T---N-I---	R-QFD--
Chicken (e)	-KSI---NS-RG-DEYGS-SDF	-T-DFHMM-K---VK-----D---V---T---H-L---K-R-T-DMIKKI-E	-Y- DT-W
Human (e)	-KSI---TS-RG-DEYGS-SDY	-T-DFHMM-K---VK-----D---V---T---H-L---K-R-T-DMIKKI-D	-Y- DT-W
Dog (e)	-KSI---TS-RG-DEYGS-SDY	-T-DFHMM-K---VK-----D---V---T---H-L---K-R-T-DMIKKI-DE	-Y- DT-W
Barley (e)	-K---K---H---Y-SYMAN-S-L	-S-EFDD-L-K-S-LM-I---DT---V---HSS-T-K-K-K-IR-A-DMIKK-ED*K-GQ-A-W	-----
Se.cereale	--SV-YI-QM-LSN-EIMNP-TK-R-K	-S-DFD-F-R-S-VK-----N---V---I---E-R-VRIM-R-R-TFDMQDI-DKD N--D---G	-----
E.coli	YTS--YI-SQ-W-MWN	-DH-HGL---Q---DA-QFM-N-R-V-LI-S---V---I---D-TVTRNL-NA-T-RV-QMLEK-K-D	A-K-QT-W
Human α	EQFSKNIKLGIHDSQNRKLSSELLRYTTSASGD	EMVSLKDYCTRMENKQHIYYITGETKDQVANSFAVERLRKNGLEVIYMIPIDEYCVQQLKEFEGKTLVSV	
Chicken α	-----A-----	-----V-----	-----
Human β	-A---L-----T---RR-----H-Q--	-T-SE-VS---T-S-----S-E-----V-R-F-V-T-----D-S--	-----
Chicken β	-A---L-----T---R-----H-Q--	-T-SE-VS---S-S-----S-E-----V-R-F-V-T-----D-S--	-----
D.melanogaster	D-----L---V---N---A-ADF-FH---	DFC-A-VS---D---V-F-S---S---S---VKAR-F-V-T-----VI-H---YK-Q--	-----
Maize	DA-----L---A-AD-HSTK---	-TT---V---G-D---SRKA-E-P-L---K-K-Y-LF-VDA---A-G---YD-K-A	-----
A.thaliana	-A---L-----G-IAD-HSTK---	-T-F-V---G-D-F---S-KA-E-P-L-K-R-Y-L-VDA---A-G---YD-K-A	-----
Tomato	-A---L-----A-FA-HSTK---	-T---V---G-ND---S-KA-E-P-L-K-K-K-Y-L-VDA---SIG---K-A	-----
Pl.falciparum	-----L---NA-T-I-T-FQ-K---	-IG-E-VD---D---SINA-S-P-L-A-T-K-F---VD---A---D-D-K-KCC	-----
L.amazonensis	---G-----TA---M---F-S-TE-E---	V-TT---V---AE-NS---DS-KLES-P-I-QAKRR-F-V-LF-T-Y---VM-V-D-D-KFACL	-----
T.brucei	-----V---T---M---FMS-E-E---	-TT---V---G-C---V-D-S-KLET-P-I-QA-RR-M-LF-TD---VM-V-D-D-KFACL	-----
S.cerevisiae(h)	SA-----V---T---AA-AK-NSTK-V-	LLT-T-V---P-H-N---SLKA-EK-P-LDA-KAKNF-LFLTD---AFT---DI	-----
H.capsulatum	SA-----V---A---PA-AK-NSTK---	-TT-T-V---P-HH-T---SLKA-QK-P-LDT-KEKNF-LFLVD-N---AMT---D-K-DI	-----
Chicken (e)	KE-GT-V---VI---HS-TR-AK-FQS-HNES	NLT-DQ-VE---K-DK-FMA-ASRKEAES-P---L-K-Y---LT-V---I-A-P-D-RFQN	-----
Human (e)	KE-GT-V---VI---HS-TR-AK-FQS-HHPT	DIT-DQ-VE---K-DK-FMA-SSRKEAES-P---L-K-Y---LT-V---I-A-P-D-RFQN	-----
Dog (e)	KE-GT-V---VI---HS-TR-AK-FQS-HHPS	DIT-DQ-VE---K-DK-FMA-SSRKEAES-P---L-K-Y---LT-V---I-A-P-D-RFQN	-----
Barley (e)	NE-G-SV---I---AT-NR-AK-FES-K-DG	KL--DE-IS---SG-D-F-L---SS-E-LEK-P-L-Q-T-KNY---FTD-V---LM-Y-MDY-D-KFQN	-----
Se.cereale	-S-G-FM---CI---G-Q-R-AP---F-S-KNET	DLI-DQ-VEN-P-T-A---ATDSLOSAKTAP-L-K-LQKDI-L-L---VAI-N-QTYKE-KF-DI	-----
E.coli	Q---GLVL-E-PA-FA-QEATIAK---FASTHTDSSAQT---E-VS---G-EK---ADSYAAKS-PHL-L---K-I---LLSDR---WMNY-T-D-PFQ--		-----
Human α	TKEGLELPE	DEEEKKQKEEKTKFENLCKIMKD	ILEKKVKEVWVSNRLVTPSPCCIVTSTYGWTANMERIMKAQAL
Chicken α	-----A-----	-----A-----	RDNSTMGYMAAKKHLEINPDHSIIETL
Human β	-----M-S-A-----L-E	-D---TI---S-----	-----M-----P-V--
Chicken β	-----NM-S-A---T---L-E	-D---TI---S-----	-----M-----P-V--
D.melanogaster	-----S---R-D-A---S-L-S	-DN-----D-----QF-S---	---TA---G-Q---P-V--
Maize	-K-DEED	-A-RR-R-KR-E-VI---	-GDR---D-I-D---L-GE-----
A.thaliana	-K-ED ET	-R-KS---T-I-E	-GD---D-I-D---L-GE-----
Tomato	-K-D	-S-D---L-E---G-V--	V-GD---I-D-V-D---L-GE-----
Pl.falciparum	-----DIDD	-S-A-DF-TL-AEY-G-VI---	V-HE---GQ-ITD---VL---EF-S---
L.amazonensis	---VHFE	-S-KT-RQR-E-ATC-K---T-E	V-GD---T-E-A---IL---EF-S-H-QM-RN---
T.brucei	---VHFE	-T---QR-E-ASY-R-A-E	V-GD---D-A---IL---EF-S-H-Q-RN---
S.cerevisiae(h)	-D F-E	TD---ADR-KEIKEY-P-T-AL-E	-GDQ---YK-LDA-AA-R-GQF-S---
H.capsulatum	-D F-E	T---DR-AEEKEY-G-A-SLEN	-GD---HK-IG---A-R-GQF-S---
Chicken (e)	A---VKFE	-S-KS-ESR-ALEKE-P-LNW-KA-KD-I-A-L-Q-TQ---AL-A-Q---SG---	YQTGK-I---NY-ASQ-TF---R-PL-KDM
Human (e)	A---VKFD	-S-KT-ESR-AVEKE-P-LNW-KA-KD-I-A-L-Q-TQ---AL-A-Q---SG---	YQTGK-I---NY-ASQ-TF---R-PL-RDM
Dog (e)	A---VKFD	-S-KT-ESR-AIEKE-P-LNW-KA-KD-I-A-L-Q-TQ---AL-A-Q---SG---	YQTGK-I---NY-ASQ-TF---R-PL-KDM
Barley (e)	S---K-GK	EDL EDK---KE-TDWW-KALDTGEGDS-KI---HNT-VV---K---SS-K-Q-T	S-A-KQA-RG-RV---R-P-K
Se.cereale	S-D---GD	SKL EDK---T-QEYTL-DWI-QQ-GD---A-QI-K-SS---VL-SGKF-S---	L---T---G-T-SLEF-RGRRIF---P-VKD
E.coli	S-VDES-EK	LAD-VDES-AK-AKALTPFIDRV-A	L-GER-KD-RLTH---TDT-ATVS-DADEMSTQ-AKLFA-AG
Human α	RQKAEA	KDNKDSVKDLVILLYETALLSSGFSLEDQTHANRYRMIKLGLGIDEDDPTADDTSAAVTEEMPPLEGDDTSRMEEYD	
Chicken α	-----A-----	-----A-----	TA-EA-P-
Human β	-----A-----V-F-----S-----	-----EVA-EEPN---PD-I---E-A---	-----
Chicken β	-D N-----A-----V-F-----S-----	-----EVI-EES-I-PPD-I---E---	-----
D.melanogaster	-D N-----A-----F-S-----DS-V-----S-----	-----E-MTT-DAQSAGDAPSLV-DTE-A-H---	-----
Maize	-KR---R-----L-F-----T-----D-N-F-A-H-L-N-----AAADE-ADMPALD	---AAEE-K---	-----
A.thaliana	-KR---R-----M-----T-----DE-N-F-A-H-L-S-----ENVEE-GDMPEL	---DAEE-K---	-----
Tomato	-KR-D-----L-F-----T-----E-N-FG-H-L-S-----ESGDA-ADMPAL-DP	---A-AEG-K---	-----
Pl.falciparum	K---D---S-T---IW-FD-S-T---A-E-T-FSK-H-----S---EEN-N-IDLPPL-T	V-ATD-K---	-----
L.amazonensis	-RRV---E---A---F-FD-S-T---Q---T Y-E-H-----SL-EE EEEAVE-AVA-TA-A-VTAG-S	L---	-----
T.brucei	KRRV---E---AA---IF-FD-S-T---T-D-TAY-D-H-----SL-D-A	E EE E-QAPVAAAANSSTGA-G	-----
S.cerevisiae(h)	KKRVDEGAQ-T---TK-----T---DE-TSF-S-N-L-S---N---EE-ETAPE	STAA-V-EVPADT---	-----
H.capsulatum	EG---RT---SITQ-F-S-S-V---TI-E-SGF-G-HKLV-S---N---AETSEK	ADTVAAEAPAE-D	-----
Chicken (e)	LRRVKE	NED-T-S-AVV-F---T-R-YM-P-TKEYGD-E-LR-S-N-L-AKVEEPEE	EDAAEEAEQDE EVDADA EDSETQKESTDVKEE
Human (e)	LRRRIKE	-ED-T-L-AVV-F---T-R-YL-P-TKAYGD-E-LR-S-N-P-AKVEEPEE	P-TAEDTTE-TEQDE-MAQVGDDEEETAKESTAKDEE
Dog (e)	LRRVKE	-ED-T-S-AVV-F---T-R-YL-P-TKAYGD-E-LR-S-N-P-AKVEEPEE	P-TTETDTE-TEQDE-M-AGTDDDEEETVKKSTAEKDEE
Barley (e)	-D-VAQ	-SDSEGL-QTAR-V-Q---ME---N-P-KDF-SS---SVQKS-DLSP-AAVEE	EEVEPEEV-EKESAKQAEPEKEQYD KEE
Se.cereale	SAACKN	EPDSTE-RA-E---I---YTP-S-AELGGK-E-MTIA--GRWGRG-EAETN-DGDSSEGV	VPEVI-PSEVRATENENDPWQD
E.coli	R-AD	TEDEAKFSEV-E-LDQ---AERGT---NLFIR-MNQLVS	

in the Hsp90 sequence. This region was also omitted from phylogenetic analysis. In the places marked with asterisks (*) in *Plasmodium falciparum* and barley (e), sequences small insertions are present which are not shown here. The abbreviations used in the species names are as follows: A., *Arabidopsis*; Aj., *Ajellomyces*; C., *Catharanthus*; D., *Drosophila*; E., *Escherichia*; H., *Histoplasma*; L., *Leishmania*; P., *Pharbitis*; Pl., *Plasmodium*; S., *Saccharomyces*; Se., *Secale*; T., *Trypanosoma*; Th., *Theileria*; c, h, and e in parentheses indicate cognate, heat-induced, and ER-resident forms of Hsp90, respectively.

Table 2
Similarity Matrix of Hsp90 Sequences

Species → ↓	A	B	C	D	E	F	G	H	I	J	K
(A) Human α		86.3	78.4	69.0	62.5	66.5	63.0	47.6	47.5	44.2	41.2
(B) Human β	93.8		78.3	67.6	63.0	67.1	65.2	45.9	47.2	42.5	41.2
(C) <i>Drosophila melanogaster</i>	88.3	87.0		66.7	61.6	65.6	62.4	48.3	46.9	45.5	39.7
(D) Maize	81.4	80.7	81.0		64.9	69.5	66.4	50.5	49.2	45.3	42.7
(E) <i>Saccharomyces cerevisiae</i>	75.9	76.4	76.6	77.7		61.6	61.0	48.0	48.8	45.0	41.0
(F) <i>Plasmodium falciparum</i>	78.9	79.8	79.1	82.1	76.8		65.2	46.3	47.4	43.1	40.9
(G) <i>Trypanosoma brucei</i>	77.2	79.6	77.6	80.3	73.8	79.6		45.9	47.7	42.0	40.2
(H) Human (e)	62.5	63.1	62.4	65.9	64.8	62.5	63.1		49.2	43.4	40.5
(I) Barley (e)	62.8	64.6	63.6	65.4	67.3	62.4	65.1	64.4		43.5	41.0
(J) <i>Secale cereale</i>	60.5	60.6	62.7	62.1	61.1	58.9	60.1	59.4	60.0		42.1
(K) <i>Escherichia coli</i>	60.1	60.1	60.9	60.2	60.2	59.7	59.8	58.4	59.9	58.4	

NOTE.—Pairwise alignment of sequences was carried out as described in Material and Methods. The upper and lower triangles indicate the percentage amino acid identities and percentage amino acid similarities (i.e., identical plus conservative changes) between pairs of protein sequences. The amino acid residues that are defined as similar in this program are A, S, and T; D and E; N and Q; R and K; I, L, M, and V; F, Y, and W. The abbreviations used are as described in table

It has been suggested that this region which mimics DNA configuration is involved in shielding the DNA-binding domain of steroid hormone receptors (see Binart et al. 1989).

Phylogenetic Analysis

The global alignment of Hsp90 sequences was utilized to examine the evolutionary relationship between Hsp90 family members. The phylogenetic analysis was carried out on the sequence region which could be aligned without ambiguity in all homologs (see Material and Methods). A consensus neighbor-joining tree based on these sequences that was obtained after 975 bootstrap replicates is shown in figure 2. The tree was rooted using the Hsp90 homolog from *Escherichia coli*.

From the branching order of Hsp90 sequences in figure 2, a number of inferences can be drawn. First, it is clear that the cytosolic and ER Hsp90 homologs from various animal and plant species form paralogous gene families which are very distantly related to each other. Second, the α and β Hsp90 sequences in vertebrate species also formed separate but closely related clusters with a high degree of reliability (i.e., in 100% of bootstraps), indicating that they arose by a gene duplication event in the common ancestor to this group. Third, the two Hsp90 homologs that are present in the fungal species *Saccharomyces cerevisiae* showed greater affinity for each other than to the homolog from other fungi (*Histoplasma capsulatum* and *Ajellomyces capsulata*), indicating that they evolved by a separate gene duplication event.

In the phylogenetic tree shown in figure 2, while the cytosolic Hsp90 sequences from the animals, plants, and fungi species formed separate monophyletic groups

in 100% of the bootstraps, the relationship between these groups was not statistically resolved. For the branch leading to the animal and plant species, a bootstrap score of 81% was observed. The relatively low bootstrap score in this case was mainly due to the uncertainty in the branching position of plasmodia species (*Plasmodium falciparum* and *Theileria parva*). These species often branched with the plant group (64% of the time; see fig. 2), while in the other bootstrap sets not shown, they branched with the kinetoplastid protists (*Trypanosoma* and *Leishmania*). One other branch which was not well resolved corresponded to that leading to the ER homologs. The lower bootstrap score of this branch (76%) was due to the uncertainty in the branching position of the homolog from *Secale cereale*, which sometimes branched with the ER homologs (in 18% of the bootstrap) but most often showed deeper branching than either the cytosolic or ER homologs. As indicated earlier, the Hsp90 homolog from *S. cereale* is unusual from other Hsp90s, and its grouping or relationship to other Hsp90 homologs is unclear.

The evolutionary relationship based on Hsp90 sequences could be more clearly shown if one omits the three species (*P. falciparum*, *T. parva*, and *S. cereale*) which show unstable branching patterns. A neighbor-joining distance tree based on the remaining sequences is shown in figure 3. The bootstrap scores for various branches (of 1,000 replicates) as obtained from a separate consensus tree are shown at the nodes. As seen, all of the main branches in this tree are well resolved. In addition to confirming the relationships mentioned earlier, the phylogenetic tree now reveals a closer relationship between animal and plant species. A plant-animal clade, exclusive of other species, was observed in

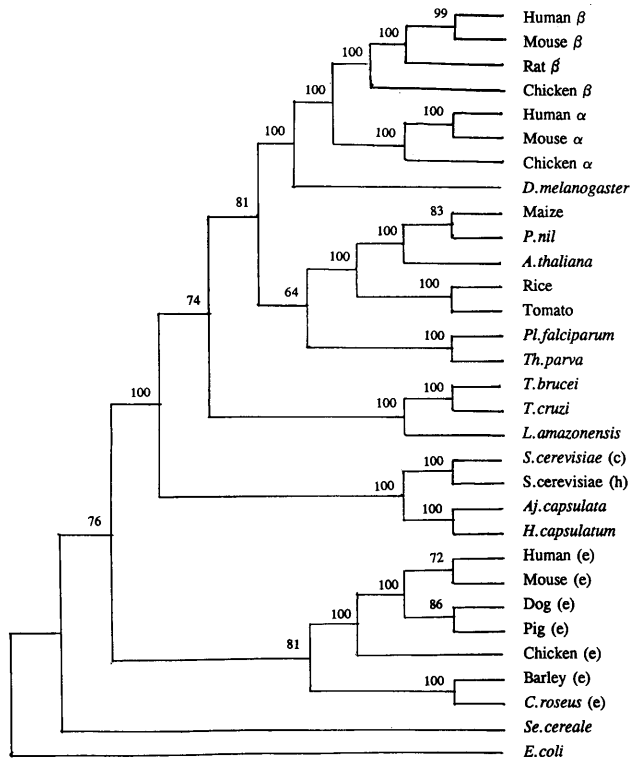


FIG. 2.—A consensus bootstrap NJ tree (Saitou and Nei 1987) based on all available Hsp90 sequences obtained after 975 bootstrap replicates. The numbers above the branches indicate the percentage of times the species which are to the right grouped together in the bootstrap trees. The abbreviations used are as indicated in fig. 1.

989 of 1,000 bootstrap replicates, which is highly significant. For the cytosolic homologs, the deepest branching was observed for the fungi species, and the kinetoplastid protist species (*Trypanosoma* (*T.*) *brucei*, *T. cruzi*, and *Leishmania amazonensis*) branched in between the fungi and the animal-plant clade. However, the bootstrap score of the node leading to the protist species was low (i.e., 588/1,000), indicating that the branching position of protists was unstable. In 40% of the bootstrap replicates that are not shown in figure 3, the protist species branched earlier than the fungi, indicating that their placement in the tree is unreliable.

The evolutionary relationship between Hsp90 sequences was also examined by means of parsimony analysis. A consensus bootstrap parsimony tree based on these sequences is presented in figure 4. The branching pattern of various species in the parsimony tree is virtually identical to that observed in the neighbor-joining trees. However, a number of branches which were well resolved in the neighbor-joining tree were not significantly supported by this method. The animal-plant clade which was significantly supported by the neighbor-joining method was observed in only 49% of the boot-

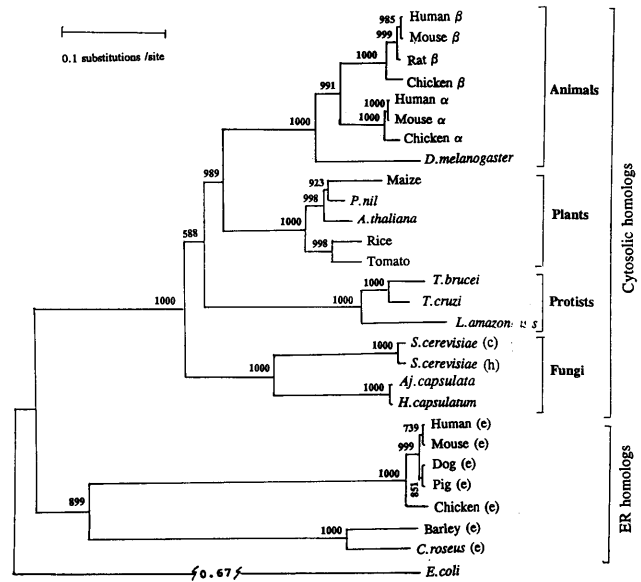


FIG. 3.—An NJ distance tree based on Hsp90 sequences. A few of the species (*Plasmodium falciparum*, *Theileria parva*, and *Secale cereale*) which showed unstable branching in fig. 2 were omitted in this tree. The distances between the proteins were determined using the Dayhoff's PAM matrix option from the PROTDIST program. The bootstrap scores obtained from a consensus NJ tree after 1,000 replicates are indicated at the nodes.

strap replicates (majority) in the parsimony tree. In the bootstrap sets that are not included in the consensus tree, a fungi-animal clade was observed in about 14% of the bootstrap samples. Thus, although the maximum-

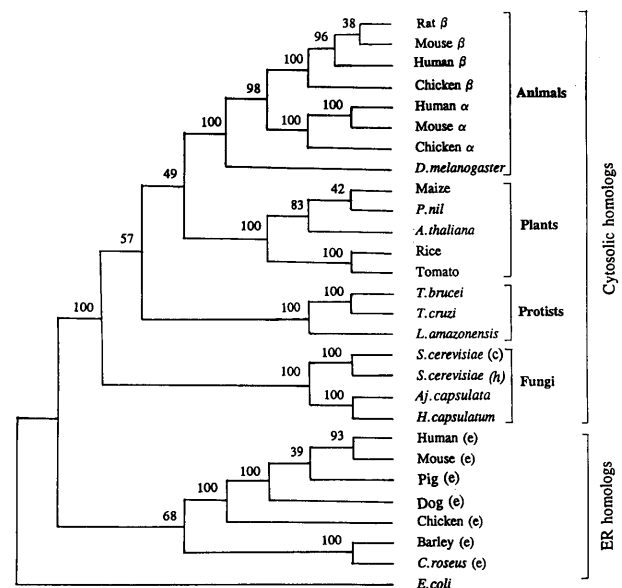


FIG. 4.—A parsimony consensus bootstrap tree based on Hsp90 sequences. The tree was bootstrapped 100 times.

parsimony analysis of Hsp90 sequences also favors a plant-animal grouping, the observed preference is not statistically significant.

Discussion

Gene Duplications in the Hsp90 Family

The phylogenetic analysis based on Hsp90 sequences presented here points to a number of gene duplication events that have taken place in this gene family. From the branching pattern of different homologs from various species, it is clear that the cytosolic and ER Hsp90 sequences form paralogous protein families which diverged from each other at a very early stage in the evolution of the eukaryotic cell. In our recent work based on the Hsp70 family of protein sequences, the cytosolic and ER homologs were again shown to comprise paralogous gene families (Gupta et al. 1994). In this case, since the cytosolic and ER homologs were present in one of the earliest branching eukaryotic species (*Giardia lamblia*), and since they both contained unique sequence signatures not found in any prokaryotic or organellar homologs, it was postulated that they evolved by a gene duplication event that accompanied the evolution of ER or nucleus in the eukaryotic cell ancestor (Gupta et al. 1994). Although for Hsp90 the cytosolic and ER homologs have not yet been identified in species such as *G. lamblia*, the observed branching pattern of the two groups of sequences in the Hsp90 phylogenetic trees (figs. 3 and 4) is consistent with and supports the above proposal.

The branching pattern of the cytosolic homologs in both the neighbor-joining and parsimony trees identifies a minimum of two additional gene duplication events that took place at a later time (i.e., after divergence of cytosolic and ER sequences). One event gave rise to the cognate and heat-inducible forms of Hsp90 found in the yeast *Saccharomyces cerevisiae*. Two Hsp90 sequences have also been reported in the fungus *Histoplasma capsulatum*. However, these two sequences differ only in a few positions and at present it is not clear whether they correspond to distinct forms of Hsp90. (If *H. capsulatum* contains two distinct Hsp90s, then one would have to postulate one more gene duplication event in this or a closely related species.) A second gene duplication event in the cytosolic branch is necessary to explain the presence of α and β subfamilies of sequences that are found in all vertebrate species. A similar inference regarding gene duplication within the Hsp90 family of sequences has been reached in earlier studies by Moore et al. (1989), based upon phylogenetic analysis of a limited number of sequences.

Evolutionary Relationship among Animals, Plants, and Fungi

The phylogenetic relationship among plants, animals and fungi has been a subject of continued debate

in recent years (see Cavalier-Smith 1987; Gunderson et al. 1987; Gouy and Li 1989; Loomis and Smith 1990; Baldauf and Palmer 1993; Hasegawa et al. 1993; Sidow and Thomas 1994). Phylogenetic analysis based on rRNA sequences indicated an animal-plant (Gunderson et al. 1987; Vossbrinck et al. 1987; Douglas et al. 1991; Hendriks et al. 1991), animal-fungi (Hasegawa et al. 1985, 1993; Cavalier-Smith 1987; Wainright et al. 1993; Cavalier-Smith et al. 1994), or a plant-fungi clade (Nairn and Ferl 1988) in different studies. Gouy and Li (1989) carried out detailed phylogenetic studies based on rRNA and six other conserved protein sequences and concluded that the plants and animals were sibling kingdoms that diverged more recently than the fungi. Sidow and Thomas (1994) have recently reported detailed analysis based on the two largest subunits of RNA polymerase II. Their analysis also strongly indicates that the plant and animal species shared a last common ancestor that excludes fungi.

However, several recent studies based on SSU rRNA, elongation factor-1 α , and a number of other conserved proteins (α -tubulin, β -tubulin, and actin) have made a strong case for a closer evolutionary relationship between animals and fungi as compared to the plant species (Baldauf and Palmer 1993; Hasegawa et al. 1993; Wainright et al. 1993; Nikoh et al. 1994). In the study by Baldauf and Palmer (1993), the inference from phylogenetic studies was supported by the fact that the animals and fungi shared a few insertions/deletions (including a 12 amino acid insertion in EF-1 α) that were not found in the plant species. However, the presence of these insertions or deletions was shown only in a few species of each kind.

In a recent detailed investigation on this subject, Nikoh et al. (1994) have evaluated the sequences of 23 different proteins from various species by the maximum-parsimony (MP), maximum-likelihood (ML), and neighbor-joining (NJ) methods to determine which of the possible relationships (i.e., [A,F] P or [A,P] F, or [P,F] A) between Animalia (A), Plantae (P), and Fungi (F) was supported. Although the inferred tree topologies were found to differ for different protein sequences, and also based on the phylogenetic method of analysis, the authors concluded that their results overall supported an (A,F) P topology. However, an examination of the results obtained by these authors show that the main support for the above inference was based on the ML method of analysis, which favored the (A,F) P topology significantly over the others. In the other two methods that were employed (i.e., MP and NJ), the (P,F) A topology was indicated only in a small number (2 or 3 of 23) of cases. However, both these methods supported the (A,P) F or (A,F) P topologies in about an equal

number of cases. The NJ method in fact slightly preferred the (A,P) F clade over the (A,F) P grouping.

In the phylogenetic analysis based on Hsp90 sequences that is reported here, a closer relationship between the Animalia and Plantae group was strongly suggested by the NJ method. The high bootstrap score (989 of 1,000) of the branch point leading to the animal-plant clade indicates that the affinities between these groups is robust and highly significant. The parsimony tree also favors a plant-animal clade; however, the support for this was not statistically significant. The ML analysis of Hsp90 sequences, likewise, did not significantly favor either of the above tree topologies (results not shown).

Thus depending on the protein sequences and the method of analysis, either fungi or plants are indicated as closer relatives to the animals. (The plant-fungi clade which is significant only for the large subunit of the RNA polymerase II [see Nikoh et al. 1994] seems unlikely to be the true topology.) Some of these differences may be related to the methods of phylogenetic analysis, as seen in the present study where significant support for the (A,P) F grouping was obtained only by the NJ method. Of the three commonly used methods for phylogenetic analysis, the NJ and ML methods are generally considered superior to the MP method, which is sensitive to the violation of constant evolutionary rate among different lineages (see Sourdis and Nei 1989; Hasegawa and Fujiwara 1993; Tateno et al. 1994). The ML method is also robust only when all of the assumptions of the ML model are satisfied (Hasegawa and Fujiwara 1993; Tateno et al. 1994). The main requirement for the NJ method to obtain correct tree topology is that the distances between the species need to be correctly estimated (Hasegawa and Fujiwara 1993; Tateno et al. 1994). However, since, depending on the protein species, all three of these methods have suggested either (A,P) F or (A,F) P clades with statistical significant support (Nikoh et al. 1994), it is difficult to explain the observed differences or to determine which of these truly represent the species topology. Recently, although a number of studies have suggested that the (A,F) P grouping represents the true tree topology in the eukaryotic lineage (Baldauf and Palmer 1993; Hasegawa et al. 1993; Wainright et al. 1993; Nikoh et al. 1994), no satisfactory explanation has been provided for the equally large number of protein species which support the alternative (A,P) F grouping. Unless all of these observations can be satisfactorily explained, the question of whether the animals are more closely related to plants or fungi should be considered largely unresolved.

Acknowledgments

The work from the author's laboratory was supported by research grants from the Medical Research

Council of Canada. I thank B. Golding for access to his Sun Workstop for phylogenetic analysis which was performed using the PHYLIP program package made available by Joseph Felsenstein. I thank Barbara Sweet for secretarial assistance.

LITERATURE CITED

- BARDWELL, J. C. A., and E. A. CRAIG. 1987. Eukaryotic Mr 83,000 heat shock protein has a homologue in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **84**:5177–5181.
- BALDAUF, S. L., and J. D. PALMER. 1993. Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc. Natl. Acad. Sci. USA* **90**:11558–11562.
- BINART, N., B. CHAMBRAUD, B. DUMAS, D. A. ROWLAND, C. BIGOGNE, J. M. LEVIN, J. GARNIER, E. E. BAULIEU, and M. G. CATELLI. 1989. The cDNA-derived amino acid sequence of chick heat shock protein Mr 90,000 (HSP90) reveals a "DNA like" structure: potential site of interaction with steroid receptors. *Biochem. Biophys. Res. Commun.* **159**:140–147.
- BLACKMAN, R. K., and M. MESELSON. 1986. Interspecific nucleotide sequence comparisons used to identify regulatory and structural features of *Drosophila* hsp82 gene. *J. Mol. Biol.* **188**:499–515.
- BOHEN, S. P., and K. R. YAMAMOTO. 1994. Modulation of steroid receptor signal transduction by heat shock proteins. Pp. 313–314 in R. I. MORIMOTO, A. TISSIERES, and C. GEORGOPOULOS, eds. *The biology of heat shock proteins and molecular chaperones*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- BOORSTEIN, W. R., T. ZIEGELHOFFER, and E. A. CRAIG. 1994. Molecular evolution of the HSP70 multigene family. *J. Mol. Evol.* **38**:1–17.
- BORKOVICH, K. A., F. W. I. FARRELLY, D. B. FINKELSTEIN, J. TAULIEU, and S. LINDQUIST. 1989. Hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures. *Mol. Cell. Biol.* **9**:3919–3930.
- CALA, S. E., and L. R. JONES. 1994. Grp94 resides within cardiac sarcoplasmic reticulum vesicles and phosphorylated by casein kinase II. *J. Biol. Chem.* **269**:5926–5931.
- CAVALIER-SMITH, T. 1987. The origin of eukaryotes and archaebacterial cells. *Ann. N.Y. Acad. Sci.* **503**:17–54.
- CAVALIER-SMITH, T., M. T. E. P. ALLSOPP, and E. E. CHAO. 1994. Chimeric conundra: are nucleomorphs and chromists monophyletic or polyphyletic? *Proc. Natl. Acad. Sci. USA* **91**:11368–11372.
- CONNER, T. W., P. R. LAFAYETTE, R. T. NAGAO, and J. L. KEY. 1990. Sequence and expression of a hsp83 from *Arabidopsis thaliana*. *Plant Physiol.* **94**:1689–1695.
- DOUGLAS, S. E., C. A. MURPHY, F. F. SPENCER, and M. W. GRAY. 1991. Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. *Nature* **350**:148–151.
- DRAGON, E. A., S. R. SIAS, E. A. KATO, and J. D. GABE. 1987. The genome of *Trypanosoma cruzi* contains a constitutively

- expressed tandemly arranged multicopy gene homologous to major heat shock protein. *Mol. Cell. Biol.* **7**:1271-1275.
- FARRELLY, F. W., and D. B. FINKENSTEIN. 1984. Complete sequence of the heat shock inducible hsp90 gene of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **259**:5745-5751.
- FELSENSTEIN, J. 1985. Confidence limits in phylogenies: an approach using the bootstrap. *Evolution* **39**:783-791.
- . 1991. PHYLIP manual, version 3.5. University of California, Berkeley.
- GOUY, M., and W.-H. LI. 1989. Molecular phylogeny of the kingdoms animalia, planta and fungi. *Mol. Biol. Evol.* **6**: 109-122.
- GUNDERSON, J. H., H. ELWOOD, A. INGOLD, K. KINDLE, and M. L. SOGIN. 1987. Phylogenetic relationships between chlorophytes, chrysophytes and oomycetes. *Proc. Natl. Acad. Sci. USA* **84**:5823-5827.
- GUPTA, R. S. 1995. Evolution of the chaperonin families (HSP60, HSP10 and Tcp-1) of proteins and the origin of eukaryotic cells. *Mol. Microbiol.* **15**:1-11.
- GUPTA, R. S., K. AITKEN, M. FALAH, and B. SINGH. 1994. Cloning of *Giardia lamblia* HSP70 homologs: implications regarding origin of eukaryotic cells and of endoplasmic reticulum. *Proc. Natl. Acad. Sci. USA* **91**:2895-2899.
- GUPTA, R. S., and G. B. GOLDING. 1993. Evolution of HSP70 gene and its implications regarding relationships between archaeobacteria, eubacteria and eukaryotes. *J. Mol. Evol.* **37**:573-582.
- GUPTA, R. S., and B. SINGH. 1994. Evolutionary analysis of HSP70 protein sequences and a chimeric model for the origin of eukaryotic cell nucleus. *Current Biol.* **4**:1104-1114.
- HARDESTY, B., and G. KRAMER. 1989. The 90,000 dalton heat shock protein, a lot of smoke but no function as yet. *Biochem. Cell Biol.* **67**:749-750.
- HASEGAWA, M., and M. FUJIWARA. 1993. Relative efficiencies of the maximum likelihood, maximum parsimony, and neighbor-joining methods for estimating protein phylogeny. *Mol. Phylogenet. Evol.* **2**:1-5.
- HASEGAWA, M., T. HASHIMOTO, J. ADACHI, N. IWABE, and T. MIYATA. 1993. Early branchings in the evolution of eukaryotes: ancient divergence of entamoeba that lacks mitochondria revealed by protein sequence data. *J. Mol. Evol.* **36**:380-388.
- HASEGAWA, M. Y., Y. IIDA, T. YANO, F. TAKAIWA, and M. IWABUCHI. 1985. Phylogenetic relationships among eukaryotic kingdoms inferred from ribosomal RNA sequences. *J. Mol. Evol.* **22**:32-38.
- HENDRIKS, L., R. DE BAERE, Y. VAN DE PEER, J. NEEFS, and A. GORIS. 1991. The evolutionary position of rhodophyte *Porphyra umbilicalis* and the basidiomycete *Leucosporidium scottii* among other eukaryotes as deduced from complete sequences of small ribosomal subunit RNA. *J. Mol. Evol.* **32**:167-177.
- HIGGINS, D. G., and P. M. SHARP. 1988. CLUSTAL: a package for performing multiple sequence alignments on a micro-computer. *Gene* **73**:237-244.
- HOFFMANN, T., and B. HOVMANN. 1988. Heat shock proteins hsp84 and hsp86 of mice and men: two related genes encode formerly identified tumor-specific transplantation antigens. *Gene* **74**:491-501.
- KISHINO, H., T. MIYATA, and H. HASEGAWA. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J. Mol. Evol.* **30**:151-160.
- KULOMAA, M. S., N. L. WEIGEL, D. A. KLEINSEK, W. G. BEATTIE, O. M. CONNELLY, C. MARCH, T. ZARUCKI-SCHULTZ, W. T. SHRADER, and B. W. O'MALLEY. 1986. Amino acid sequence of a chicken heat shock protein derived from the complementary DNA nucleotide sequence. *Biochemistry* **25**:6244-6251.
- LINDQUIST, S., and E. A. CRAIG. 1988. The heat shock proteins. *Annu. Rev. Genet.* **22**:631-677.
- LOOMIS, W. F., and D. W. SMITH. 1990. Molecular phylogeny of *Dictyostelium discoideum* by protein sequence comparison. *Proc. Natl. Acad. Sci. USA* **87**:9093-9097.
- MAKI, R. G., L. J. OLD, and P. K. SRIVASTAVA. 1990. Human homologue of murine tumor rejection antigen gp 96: 5'-regulatory and coding regions and relationship to stress-induced proteins. *Proc. Natl. Acad. Sci. USA* **87**:5658-5662.
- MARRS, K. A., E. S. CASEY, S. A. CAPITANT, S. A. BOUCHARD, P. S. DIETRICH, I. J. METTLER, and R. M. SINIBALDI. 1993. Characterization of two maize HSP90 heat shock protein genes: expression during heat shock, embryogenesis, and pollen development. *Dev. Genet.* **14**:27-41.
- MAZZARELLA, R. A., and M. GREEN. 1987. Erp99, an abundant, conserved glycoprotein of the endoplasmic reticulum is homologous to the 90-kDa heat shock protein (hsp90) and the 94-kDa glucose regulated protein (GRP94). *J. Biol. Chem.* **262**:8875-8883.
- MCGUIRE, J. A., L. POELLINGER, A. C. WIKSTROM, and J. A. GUSTAFSSON. 1992. Cloning and regulation by glucocorticoid receptor ligands of a rat hsp90. *J. Steroid Biochem. Mol. Biol.* **42**:813-822.
- MENG, X., V. JEROME, J. DEVIN, E. E. BAULIEU, and M. G. CATELLI. 1993. Cloning of chicken hsp90B: the only vertebrate hsp90 insensitive to heat shock. *Biochem. Biophys. Res. Commun.* **190**:630-636.
- MINCHOITTI, G., S. GARGANO, and B. MARESCA. 1991. The intron-containing hsp82 gene of the dimorphic pathogenic fungus *Histoplasma capsulatum* is properly spliced in severe heat shock conditions. *Mol. Cell Biol.* **11**:5624-5630.
- . 1992. Molecular cloning and expression of hsp82 gene of the dimorphic pathogenic fungus *Histoplasma capsulatum*. *Biochim. Biophys. Acta* **1131**:103-107.
- MOORE, S. K., C. KOZAK, E. A. ROBINSON, S. J. ULLRICH, and E. APPELLA. 1989. Murine 86- and 84-kDa heat shock proteins, cDNA sequences, chromosome assignments and evolutionary origins. *J. Biol. Chem.* **264**:5343-5351.
- MORIMOTO, E. I., A. TISSIERES, and C. GEORGOPOULOS, eds. 1994. The biology of heat shock proteins and molecular chaperones. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- MOTTRAM, J., W. MURPHY, and N. AGABIAN. 1989. A transcriptional analysis of the *Trypanosoma brucei* hsp83 gene cluster. *Mol. Biochem. Parasitol.* **37**:115-128.
- MYERS, E. W., and W. MILLER. 1988. Optimal alignments in linear space. *Comput. Appl. Biosci.* **4**:11-17.

- NAIRN, C. J., and R. J. FERL. 1988. The complete nucleotide sequence of the small subunit ribosomal RNA coding region for the cycad *Zamia pumila*: phylogenetic implications. *J. Mol. Evol.* **27**:133–141.
- NIKOH, N., N. HAYASE, N. IWABE, K. KUMA, and T. MIYATA. 1994. Phylogenetic relationship of the kingdoms Animalia, Plantae and Fungi, inferred from 23 different protein sequences. *Mol. Biol. Evol.* **11**:762–768.
- PELHAM, H. R. B. 1989. Control of protein exit from endoplasmic reticulum. *Annu. Rev. Cell Biol.* **5**:1–23.
- PRATT, W. B. 1993. The role of heat shock proteins in regulating the function, folding and trafficking of the glucocorticoid receptor. *J. Biol. Chem.* **268**:21455–21458.
- REBBE, N. F., J. WARE, R. M. BERTINA, P. MODRICH, and D. W. STAFFORD. 1987. Nucleotide sequence of a cDNA for a member of the human 90-kDa heat shock protein family. *Gene* **53**:235–245.
- SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method of reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- SCHRODER, G., M. BECK, J. EICHEL, H. P. VETTER, and J. SCHRODER. 1993. HSP90 homologue from Madagascar periwinkle (*Catharanthus roseus*): cDNA sequence, regulation of protein expression and location in the endoplasmic reticulum. *Plant Mol. Biol.* **23**:583–594.
- SHAPRIA, M., and G. PEDRAZA. 1990. Sequence analysis and transcriptional activation of heat shock protein 83 of *Leishmania amazonensis*. *Mol. Biochem. Parasitol.* **42**: 247–255.
- SIDOW, A., and W. K. THOMAS. 1994. A molecular evolutionary framework for eukaryotic model organisms. *Current Biol.* **4**:596–603.
- SOURDIS, J., and M. NEI. 1989. Relative efficiencies of the maximum parsimony and distance matrix methods in obtaining the correct phylogenetic tree. *Mol. Biol. Evol.* **5**: 298–311.
- TATENO, Y., N. TAKEZAKI, and M. NEI. 1994. Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum-parsimony methods when substitution rate varies with site. *Mol. Biol. Evol.* **11**:261–277.
- VOSSBRINCK, C. R., J. V. MADDOX, S. FRIEDMAN, B. A. DEBRUNNER-VOSSBRINCK, and C. R. WOESE. 1987. Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. *Nature* **326**:411–414.
- WAINRIGHT, P. O., G. HINKLE, M. L. SOGIN, and S. K. STICKEL. 1993. Monophyletic origins of the metazoa: an evolutionary link with fungi. *Science* **260**:340–342.
- YAMAZAKI, M., K. AKAAGI, T. MIWA, T. IMAI, E. SOLDA, and K. YOKOYAMA. 1989. Nucleotide sequence of a full length cDNA for 90 kDa heat shock protein from human peripheral blood lymphocytes. *Nucleic Acids Res.* **17**:7108–7108.

NAOYUKI TAKAHATA, reviewing editor

Received January 18, 1995

Accepted May 19, 1995