

HSP: evolved and conserved proteins, structure and sequence studies

Desai N.S.*, Agarwal A.A. and Uplap S.S.

*Padmashree Dr. D. Y. Patil University, Department of Biotechnology and Bioinformatics,
Plot No.50, Sector-15, CBD Belapur, Navi Mumbai-400614, MS, India, neetindesai@gmail.com

ABSTRACT- Heat shock proteins (HSPs) are the proteins which are present normally in the cell but their expression level increases under stress condition and are mainly divided into five groups, low molecular weight HSP (LMW HSP), HSP 60, HSP 70, HSP 90 and high molecular weight HSP (HMW HSP). All these classes of HSPs are highly conserved and ubiquitous in nature and hence serve as a good model for phylogenetic analysis. For the first time in this study, the sequence and structural analysis has been carried out to predict evolution of HSPs. The results obtained clearly show higher degree of sequence and structural conservation. HSP 60 and HSP 70 are highly conserved in terms of both sequence and structure alignment in comparison to HSP 90. The minimum amino acid identity that has been observed between all the homologous sequences is 32.46%, 38%, 23.60% for HSP 60, HSP 70 and HSP 90 respectively, indicating HSP 70 as the most conserved protein family followed by HSP 60 and HSP 90 family. The structural analysis of these proteins showed dominance of beta sheets in HSP70 and helices in HSP 90. The detailed analysis of all the HSP homologues revealed high conservation of glycine residues and ATP binding pockets. Thus this study has revealed that HSPs are highly structurally and functionally conserved proteins and warrants further detailed analysis at organism level.

Key words: HSP 60, HSP70, HSP 90, sequence, structure, alignment, conserved and evolution.

INTRODUCTION

Heat shock proteins (HSP) are cellular proteins that are present under normal conditions but their expression level increases when the cell is under stress or shock [1]. These proteins were named as heat shock proteins because they were first thought to be induced by heat and are conserved and ubiquitous in nature that are essential for maintaining homeostasis in cell and helps in its survival during stress [2, 3, 4]. Ritossa (1962) observed unusual puffing pattern in *Drosophila* under temperature stress and also observed that the serendipitous heat shock was responsible for high transcription rate resulting in unusual puffing pattern [1, 5, 6]. Similar work was carried out in different organisms and the pattern of protein expression under heat stress was studied thus giving an overview that heat shock response is not restricted to few organisms but virtually in all organisms from *E. coli* to *Homo sapiens* [5]. It has also been demonstrated that not only heat stress but also many environmental and physical factors are responsible for the induction of HSPs [8]. These factors include biotic and abiotic stress. The later type of stress constitute temperature, osmotic stress, certain chemicals, presence of heavy metals, hypoxia, hyperoxia, changes in pH, free radicals, toxins, ischemia, ionization radiation etc whereas biotic stress includes injury to cell, infection, viral pathogenesis, various stages of development (gametogenesis, embryogenesis and metamorphosis), cell division and differentiation, tumor formation, aging and senescence, type of organelle etc [9, 10, 3]. Geographical distribution and climatic condition also play a major role in the observed variation in the stress induced expression of heat shock proteins in an organism [11].

Types of heat shock proteins:

HSPs range in size from 27 to 110 kDa and are divided into five groups based on molecular weight and function [12]. Major families of HSPs are low molecular weight (LMW) HSP, HSP 60, HSP70, HSP 90 and high molecular weight (HMW) HSP [17]. LMW HSPs are super family of chaperones characterized by a conserved C-terminal domain of ~90 amino acids referred to as the α -crystalline domain, flanked by a short C-terminal extension and an N-terminal arm of variable length and highly divergent sequence [18]. Upon heat stress, both prokaryotic and eukaryotic cells produce a group of proteins with molecular mass of 15 to 42 kDa, designated as small heat shock proteins (sHSP) [16].

HSP 60 are the first HSPs to be termed as 'molecular chaperons' and their role is supported by both the genetic and biochemical studies. HSP 60 homologues are mainly located in mitochondria and chloroplast but they are encoded by nucleus. Prokaryotic HSP 60 homolog is GroEL, expressed in the cell in normal condition but its expression level increases under heat stress [19]. HSP 60 participates in many cellular processes, mainly protein folding or assembly [19]. These chaperone proteins are also used as advanced marker sequences to study the evolution of mitochondrial genome [48]. HSP 70 expression is universal and is highly conserved across all the kingdoms [51, 52, 53]. HSP70 homolog plays an important role in protein folding of newly synthesized peptides and protection of cells under stress condition. Besides, it also performs a major role in the intracellular protein trafficking

and apoptosis [54, 56]. In all the cases it has been observed that these proteins operate by similar fundamental mechanisms of substrate binding/sequestration and release [56]. In eukaryotes there exist at least three types of HSP70s, each of which localizes to a different cellular compartment: cytoplasm, mitochondrion (MT), and endoplasmic reticulum (ER) [51]. Members of the HSP70 family are strongly upregulated by heat stress and toxic chemicals and heavy metals such as arsenic, cadmium, copper, mercury, etc [55]. Elevated cytosolic HSP70 levels mediate protection of tumor cells from stress-induced lethal damage by interfering with apoptotic pathways and thus, may induce resistance to chemoradiotherapy. Therefore, high cytosolic levels of HSP70 are frequently associated with a negative clinical outcome and a higher frequency of metastasis in human cancer patients. In contrast, membrane bound and extracellularly located HSP70 is known to elicit cancer immunity and, thus, might be beneficial for the clinical outcome [55]. HSP90 proteins are essential molecular chaperones which play an important role in the signal transduction, cell cycle control, stress management, and folding, degradation, and transport of proteins. These proteins have been found in a variety of organisms suggesting that they are ancient and conserved [58]. Heat shock protein 90 acts in concert with co-chaperones and plays an important role in the refolding and stabilization of denatured proteins after cellular stress conditions, such as heat or hypoxia. Examples of HSP90 client proteins include steroid receptors, mutated p53 and several protein kinases like ErbB2, Src, Abl, Raf, Akt and cyclin-dependent serine kinases. Inhibition of HSP90 prevents the association between the chaperone and various oncogenic proteins, which leads to their degradation by the proteasome. Therefore, HSP90 inhibitors have antiproliferative effects and, depending on the cellular context, they induce apoptosis, cell cycle arrest, or differentiation. Thus, HSP90 has been recognized as a valuable and attractive target for anticancer drug discovery. [59] In the present paper, both the sequence and structural aspects has been considered to deduce the evolutionary trend and to study the level of conservation in each of the major HSP families i.e. HSP60, HSP70 and HSP90.

MATERIALS AND METHODS

The sequences used from various species for each of the family were retrieved from National Centre of Biotechnology Information (NCBI). For each of the major HSP family, i.e. HSP 60, HSP 70, HSP 90, the homolog sequences were selected from following species: *Homo sapiens*, *Mus musculus*, *Drosophilla melanogaster*,

Arabidopsis thaliana, *Saccharomyces cerevisiae*, *Escherichia coli* (Table 1). The amino acid conservation in all the homologues selected was determined by performing multiple sequence alignment on all the sequences, for each of the family. Multiple sequence alignment (MSA) was carried out using CLUSTAL W application present in BioEdit (version 7.09) software [45]. To study the evolutionary relationship between each of the homolog belonging to a particular family, phylogenetic analysis was performed. For phylogenetic analysis phylogenetic tree was prepared by the maximum likelihood method using online tool Phylogeny.fr [46]. Further to study the structural conservation of the HSP homologs of the respective family, homology modeling was used to provide structure to the sequences whose structures are yet unavailable. To perform homology modeling on each of the sequences, online tertiary structure prediction tool SWISS-MODEL version 8.05 was used [40, 41, 42, 43, 44]. For structural analysis, the structures were superimposed using software Swiss-Pdb viewer [47] and further analysis was performed using licensed version of DS visualizer 1.5. The template used for each of the sequences and the percentage of sequence alignment between the selected template and each of the query sequences are given in Table 2. No templates with alignment score less than 50% were selected for the homology modeling purpose. The individual structures downloaded for each of the sequences were viewed and analyzed using Swiss-Pdb Viewer version 3.7 [47].

RESULT AND DISCUSSION

HSP 60

The 60 kD heat shock proteins are ubiquitous abundant proteins. More than 150 homologues of HSP60 sequences are currently available with pairwise similarity extending from 40 to 100% at the amino acid level [48]. Each bacterial homolog of HSP 60 or GroEL forming monomer is divided into three structural Domains: A ~Apical, I ~Intermediate and E ~Equatorial. Seven identical GroEL monomers are assembled in rings that dimerise in a 14-meric structure [49]. The HSP60 protein family plays an important role as chaperone facilitators of protein folding and in rescuing the cell from stress conditions [48, 49]. Unlike GroEL, very little is known about the structure and physiological functions of the mammalian chaperonin homologue. HSP60 was first reported as mitochondrial P1 proteins [50]. It was believed that HSP60 may have functions only in the mitochondria and that there was no chaperonin homologue in the cytoplasm of eukaryotic cells. But recently it has been found that a protein called as T - complex polypeptide 1 (CCT/TriC) composed of eight subunit of \approx 60 kDa each assist in folding of actin and tubulin

filament in presence of ATP [50]. Although CCT/TriC belongs to a chaperonin family that includes GroEL and HSP60, the sequence homology between CCT/TriC and GroEL is <40% [50]. In the present investigation emphasis is given on mitochondrial HSP 60 from prokaryotes and eukaryotes. The result of MSA (figure 1) on prokaryotic homologues of HSP 60 family shows approximately 33 % of sequence conservation whereas MSA of eukaryotic HSP60 homologues showed increase in the sequence conservation to 41% (Table 3). The percent alignment at organism level is illustrated in Table 4. MSA performed on HSP 60 sequence showed conservation of charged amino acid and glycine residue (approx 80%). Similar kind of results are reported by Brocchieri and Karlin (48) in their study on conservation of HSP 60 and concluded that G residues can act as a "filler" or "hinge" for backbone conformations which lack other side chains and the charged residue can provide strong substrate interaction which increases the efficiency of protein folding in the GroEL complex [48]. Since HSP 60 plays a major role in protein folding, conservation of such residues provides a favorable environment for HSP 60 for better substrate interaction and efficient protein folding. Phylogenetic tree generated (figure 2) showed that human and mouse HSP60 homologues formed one group which originated from the single node while *D. melanogaster* homolog showed the closest relationship to it followed by *Arabidopsis thaliana* homologue. *S. cerevisiae* and *E. coli* homologues formed completely different group, showing parallel evolution. The branch length of the tree shows that *E. coli* HSP 60 protein is highly evolved protein and justified the fact that HSP 60 homologue GroEL is a very specialized and the only protein which plays major role in protein folding and growth in bacteria [50].

The individual sequences of HSP 60 homologues from various organisms were modeled using homology-modeling software. Each modeled structure for individual organism is presented in figure 3. When each of the modeled structures was superimposed upon another, it showed high structural alignment. Though there was a sequence difference, the structural conservation suggests that the overall function of the protein may be conserved. The major regions that found to be conserved in each of the homologue modeled structures are: K-XXX-FG, GV-X-VA-X-, VIT-XX-G, EL-XXX-EGM-XX-, PL-X-I-XX-, -X-V-X-A-X-KA, VA-XXXX, G-XXX-G, I-X-DP-X-forming sheets and -XXX-R-XXX-L-x-GV-xx-LA-x-AV-xx-, K-XX-N-X-GA-XXXX-VA-XXX-N-XX, TT-X-ATVL-XXX-I-XX-E, -XXX-IAQV-X-TISAN, -X-AL-XXXX-N-XX-, -X-RK-XXXX-D-X-A-XX-, -X-Y-XX-EKL-X-ER-X-A-X-L-X-, KV-X-R-XXXX-A-XX-V-XX-L-XX-T forming helices. Where -X- is the symbol used for amino acid differing at one or

two places in all the organisms. The structures that are found to be conserved in all the homologues also lie in the region of the most conserved sequence domain. Thus both the structural and sequence analysis supports the fact that HSP 60 homologues are highly conserved.

HSP 70

Heat shock 70 proteins (HSP70s) are ubiquitous protein family, which are highly conserved across all domains of life [52]. The HSP70s are chaperones and crucial housekeeping proteins that play a major role in the transport of proteins across membranes into organelles, the folding of newly translated proteins and the refolding of denatured protein under heat stress [53]. The MSA results (Figure 4) clearly show high conservation of amino acid sequences at N-terminal domain, and the middle domain which consists of major ATP and substrate binding site respectively. And found very less conservation at C-terminal end. This suggests that the major substrate and ATP binding portion of HSP 70 is highly conserved across all the domains. The percentages of sequence conservation across all domains are illustrated in Table 5 and the percentage of similarity between pair of homologues is given in Table 6. When all the HSP70 homologues were compared, 247 amino acids were found to be conserved (approximately 38.5% of sequence conservation) whereas in case of eukaryotic homologues alone, 376 amino acids were found to be conserved showing approximately 58.6% of conservation. Phylogenetic tree generated (figure 5) showed that HSP 70 sequences are divided into two groups belonging to prokaryotes and eukaryotes. *E. coli* formed a complete out group showing parallel evolution to that of eukaryotic homologues. In eukaryotic sequences human and mouse homologues formed the same group, but a gene duplication event is observed after speciation process that gave rise to two different types namely HSP70 1a and HS0P70 1b in both the species. *A. thaliana* formed next closest group to that of human and mouse sequences, followed by *D. melanogaster* homolog, whereas *S. cerevisiae* formed a completely different group. Thus, evolution of HSP 70 homologues can be observed with an increase in the complexity of the organism.

All HSP70 homologous sequences were modeled by using homology modeling tool and the generated structures are shown in Figure 6. When all the modeled structures were superimposed upon each other it showed higher conservation of the structures. Structural analysis is supported by multiple sequence alignment performed and When all the homology modeled structures were compared, domains which were found to be conserved are -X-GIDLG, T-X-SCV-XXXX-, -X-

TPS-XX-A-X-T, E-X-L-X-G, NT-X-, T-X-AV-X-T, NT-X-, T-X-AV-X-T, X-E-X-NVLIFDLG, TFD-X-S-XXX-ID, -X-FEV-XX-T-X-GD, D-XX-LV, SLG-X-ET, G-X-MT-X-L forming sheets and DAAK, P-XX-IS-XX-VL-XX-KM-XXX-AE-XX-LG, D-X-QRQATKDAG-X-I, EPTAAA-X-AY, EDFD-X-RL-XXXXXXXX-EFK-K/R, RL-XX-A-X-E-X-AK-XX, DL-XXXXXXXX-P-XXXX-L-X-D, AVA-X-GAAVQ-XX forming helices. The symbol '-X-' is used to denote amino acids, which were not matching at one or two places but were conserved in most of the HSP70 homologues. Thus, the sequence alignment supports the hypothesis of major structural conservation in the protein from prokaryotes to eukaryotes. Glycine residues were found to be highly conserved mainly in the middle domain of the protein. Since middle domain is the substrate binding region, the sequence conservation in this domain signifies its importance in efficient interaction between HSP70 and substrate. It has also been observed that HSP 70 homologue structures are dominated by the beta sheets compared to helical structures: approximately 28 sheets were found in the structure whereas only 13 helices were observed in the homology modeled structures in human. This may be due to its major role in providing stability to the proteins and maintaining homeostasis in the cell during stress condition.

HSP 90

HSP 90 are highly conserved ubiquitous proteins that provide useful model for evolutionary studies. HSP 90 are those proteins whose function in the cell is least understood [57]. The result of multiple sequence alignment (Figure 7) performed, clearly shows the conservation of amino acids majorly at N-terminal region and ATP binding site. Comparison of all the HSP90 homologues showed that 173 amino acids were found to be conserved, giving rise to approximately 23.6% of sequence conservation, whereas in case of eukaryotic homologues alone, 394 amino acid were found to be conserved forming approximately 54% of sequence conservation (Table 7). The percentage of alignment between pair of protein is illustrated in Table 8. The region between 235-289 aa and 634-639 aa found in human HSP90 homologues is rich in charged amino acids. This particular region is found to be absent in *E.coli* HSP 90 homologue. The reason behind deletion of these regions is not yet known.

The phylogenetic tree (Figure 8) based on maximum likelihood method showed that *E. coli* formed an out group, whereas all the eukaryotic sequences belonged to a single group. In case of eukaryotic homologues, human and mouse showed the closest relationship but a gene duplication event is observed before speciation process occurred, in contrast to HSP70 homologues in which gene duplication event is

observed after speciation process. Because of this reason HSP90 alpha and HSP90 beta formed two different groups. In comparison to *A. thaliana*, *D. melanogaster* showed a closer relationship to human and mouse homologues, whereas *S. cerevisiae* formed a different group. For structural alignment, the sequences were modeled *in silico* by homology modelling and, generated structures (figure 9) were superimposed on each other. Two different templates have been used to model N terminal and middle domain with C- terminal. The suitable template for the region connecting N terminal and the middle domain could not be found, hence that region is not taken into consideration for structural analysis. The modeled structures were aligned by superimposing structures on each other. When all homologues were compared, the domains that were found to be conserved are ET-XX-F, VA-XX-V-X-V-XXX-, -XXX-WES, RGT-XXX-L-XXX-, KLYV-X-RVFI forming sheets and EIFLRELISN-X-SDA-X-DK-X-R-XXX-L, SGTK-X-F-X-E-X-L, FYS-XX, EY-XX-FYK-XXX- forming helices. As compared to other HSPs, HSP90 is highly evolved and least conserved protein, this may be because of its specialized and specific role in higher organism.

HSP 70 is the most conserved protein family compared to other HSP families. It has revealed that in all the families, glycine residues, charged amino acids and ATP binding sites are highly conserved. The results obtained shows dominance of beta sheets in the HSP 70 whereas dominance of helices in the HSP 90 structures. Since HSP 70 homologues have more stress related response, the HSP 90 homologues have more defined signal transduction role, which justifies the functional aspect of these proteins.

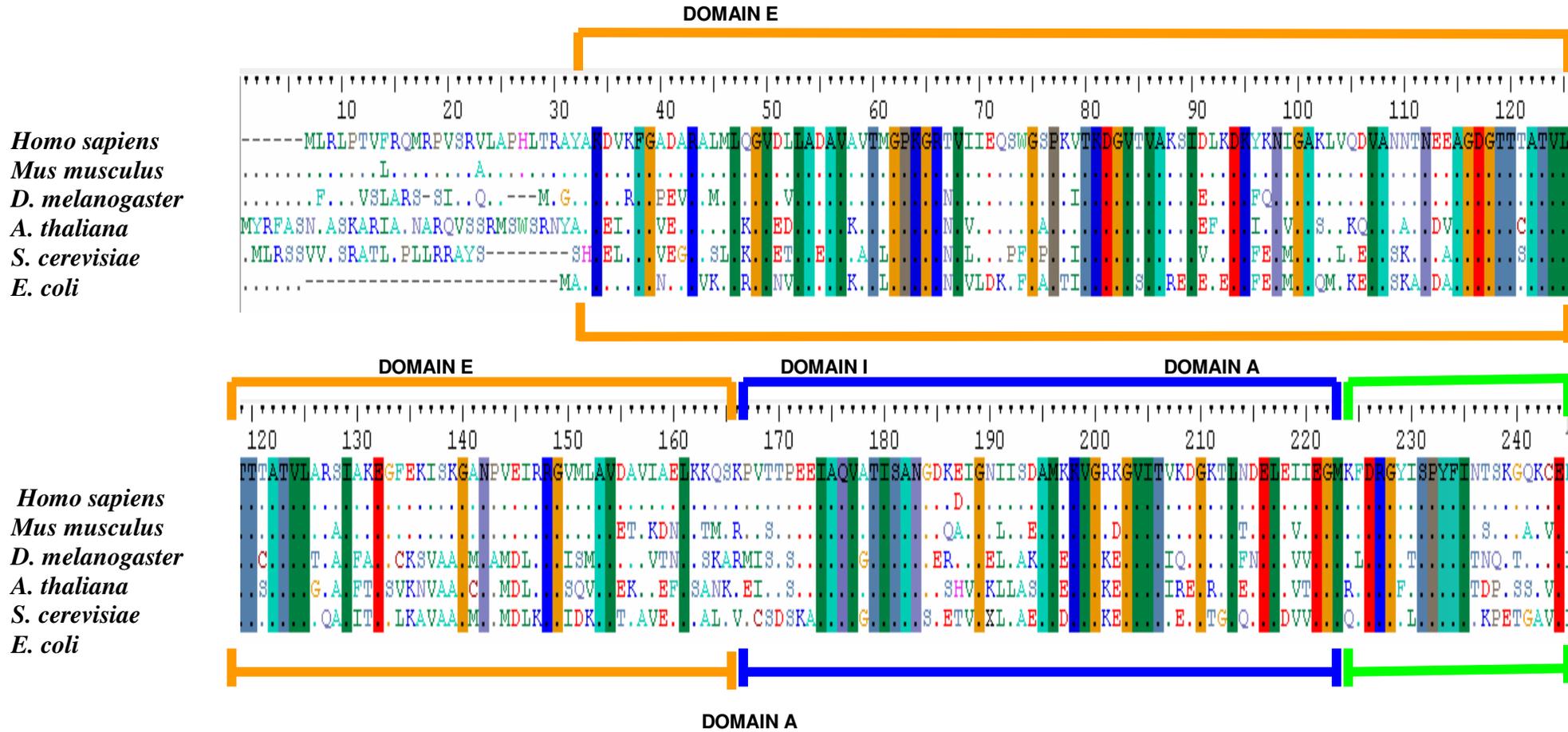
The idea behind conducting these studies is to examine the conservation pattern in each of the family so that a universal sequence could be designed for each of the family which is not organism specific and can be expressed in any organism using a proper promoter. Thus such universal sequence could be used for performing the family specific task and to solve many problems like preventing temperature specific protein degeneration which is the major problem in fermentation technology, anticancer drug development, developing heat tolerant transgenic varieties and many more.

References

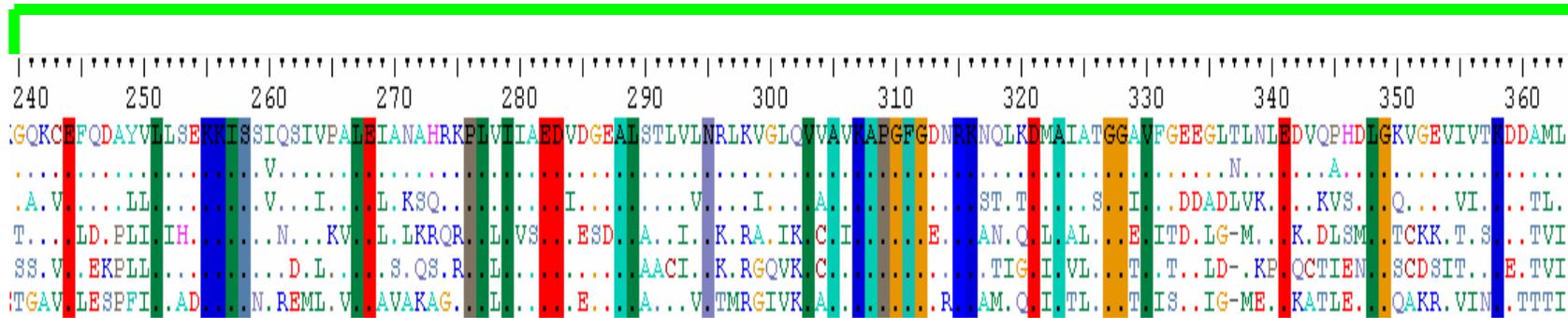
- [1] Robert J. (2003) *Developmental and Comparative Immunology*, 27: 449–464.
- [2] Sikora A. and Grzesiuk E. (2007) *Journal of physiology and pharmacology*, 58 (3):43-62.
- [3] Feder M. E. and Hofmann G.E (1999) *Annual Review Physiology*, 61:243-82

- [4] Serensen J. G., Kristensen T N and Loeschcke V, (2003) *Ecology Letters*, 6:1025–1037.
- [5] Schlesinger M. J. (1986) *The Journal of Cell Biology*, 103: 321-325.
- [6] Ritossa F. (1996) *Cell stress and chaperones* 1 (2):97-98.
- [7] Parcellier A., Gurbuxani S., Schmitt E., Solary E., and Garrido C. (2003) *Biochemical and Biophysical Research Communications*, 304:505–512.
- [8] Tomanek L. (2002) *Integ. And comp. Boil.* 42:797-807
- [9] Diamante S., Eliahu N., Rosenthal D. and Goloubinoff P. (2001) *The journal of biological chemistry*, 276 (43):39586-39591.
- [10] Feder M.E. *Society for experimental biology, seminar series: 59, Cambridge university press*
- [11] Feder M.E. (1999) *Amer. Zool.*, 39:857-864.
- [12] Moseley P.L. (1997) *J.Appl. Physiol.*, 83(5):1413–1417.
- [13] Osteryoung K. W. and Vierling E. (1994) *The journal of biological chemistry*, 269(46):28676-28682.
- [14] Hernandez L.D. and Vierling E. (1993) *Plant Physiol.*, 101:1209-1216.
- [15] Wu M.L., Lin T.P., Lin M.Y., Cheng Y.P. and Hwang S.P. (2007) *Annals of Botany*, 99:461–475.
- [16] Sun W., Montagu M.V. and Verbruggen N. (2002) *Biochimica et Biophysica Acta.*, 1577:1 – 9.
- [17] Joe M.K., Park S.M., Lee Y.S., Hwang D.S. and Hong C.B. (2000) *Mol Cells.*:10(5):519-24.
- [18] Basha E., Friedrich K.L., and Vierling E. (2006) *J. Biol. Chem.*, 281(52):39943-39952.
- [19] Vierling E. (1991) *Annu. Rev. Plant Physiology Plant Mol. Biol.*; 42:579-620.
- [20] Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grotbeck E.J., Davis N.W., Lim A., Dimalanta E., Potamouisis K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A. and Blattner F.R. (2001) *Nature* 409 (6819):529-533.
- [21] Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansorge W., Benes V., Bruckner M., Delius H., Dubois E., Dusterhoft A., Entian K.D., Floeth M., Goffeau A., Hebling U., Heumann K., Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Kotter P., Louis E.J., Messenguy F., Mewes H.W., Hoheisel J.D. et al. (1997) *Nature* 387 (6632):87-90.
- [22] Jindal S., Dudani A.K., Singh B., Harley C.B. and Gupta R.S. (1989) *Mol. Cell. Biol.* 9 (5): 2279-2283.
- [23] Kozlova T., Reynaud E., Perezgasga L. and Zurita M. *The D. Melanogasterhomologue of the hsp60 is an essential gene and is differentially expressed during fly development, Unpublished.*
- [24] Lotscher E. and Allison J.P. (1990) *Nucleic Acids Res.* 18 (23):7153.
- [25] Jindal S., Dudani A.K., Singh B., Harley C.B. and Gupta R.S. (1989) *Mol. Cell. Biol.* 9 (5):2279-2283.
- [26] Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grotbeck E.J., Davis N.W., Lim A., Dimalanta E., Potamouisis K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A. and Blattner F.R. (2001) *Nature* 409 (6819):529-533.
- [27] Wei W., Mccusker J.H., Hyman R.W., Jones T., Ning Y., Cao Z., Gu Z., Bruno D., Miranda M., Nguyen M., Wilhelmy J., Komp C., Tamse R., Wang X., Jia P., Luedi P., Oefner P.J., David L., Dietrich F.S., Li Y., Davis R.W. and Steinmetz L.M. (2007) *Proc. Natl. Acad. Sci. U.S.A.* 104 (31):12825-12830.
- [28] Karch F., Torok I. and Tissieres A.(1981) *J. Mol. Biol.* 148 (3):219-230.
- [29] Goto K., Kojima A., Morioka S., Naito T., Akema T., Matsuba Y., Fujiya H., Sugiura T., Ohira Y. and Yoshioka T. (2007) *Biochem. Biophys. Res. Commun.* 358 (1):331-335.
- [30] Belova G.I., Postnikov Y.V., Furusawa T., Birger Y. and Bustin M. (2008) *J. Biol. Chem.* 283 (13):8080-8088.
- [31] Wu B., Hunt C. and Morimoto R. (1985) *Mol. Cell. Biol.* 5 (2):330-341.
- [32] Bardwell J.C. and Craig E.A. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84 (15):5177-5181.
- [33] Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansorge W., Araujo R., Aparicio A., Barrell B., Badcock K., Benes V., Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M., Chung E., Churcher C., Coster F., Davis K., Davis R.W., Dietrich F.S., Delius H., Dipaolo T., Hani J. Et al. (1997) *Nature* 387 (6632):103-105.
- [34] Milioni D. and Hatzopoulos P. (1997) *Plant Mol. Biol.* 35 (6):955-961.
- [35] Hoskins R.A., Carlson J.W., Kennedy C., Acevedo D., Evans-Holm M., Frise E., Wan K.H., Park S., Mendez-Lago M., Rossi F., Villasante A., Dimitri P.,

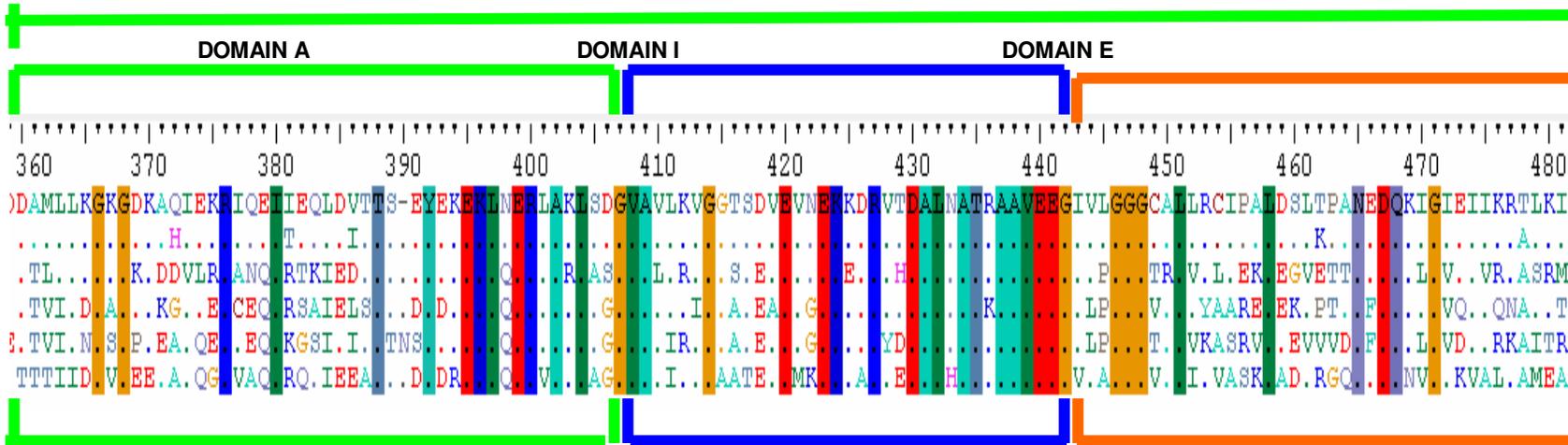
- Karpen G.H. and Celniker S.E. (2007) *Science* 316(5831):1625-1628.
- [36] Moore S.K., Appella E., Villar C.J. and Kozak C.A. (1991) *Genomics* 10 (4):1019-1029.
- [37] Moore S.K., Rijli F. and Appella E. (1990) *DNA Cell Biol.* 9 (6):387-400.
- [38] Yamazaki M., Akaogi K., Miwa T., Imai T., Soeda E. and Yokoyama K. (1989) *Nucleic Acids Res.* 17 (17):7108.
- [39] Rebbe N.F., Ware J., Bertina R.M., Modrich P. and Stafford D.W. (1987) *Gene* 53 (2-3):235-245.
- [40] Arnold K., Bordoli L., Kopp J., and Schwede T. (2006) *Bioinformatics*, 22:195-201.
- [41] Kopp J. and Schwede T. (2004) *Nucleic Acids Research*, 32:D230-D234.
- [42] Schwede T., Kopp J., Guex N., and Peitsch M.C. (2003) *Nucleic Acids Research*, 31:3381-3385.
- [43] Guex N. and Peitsch M. C. (1997) *Electrophoresis*, 18:2714-2723.
- [44] Peitsch M. C. (1995) *Bio/Technology*, 13:658-660.
- [45] Hall T.A. (1999) *Nucl. Acids. Symp. Ser.*, 41:95-98.
- [46] Higgins D., Thompson J., Gibson T., Thompson J.D., Higgins D.G. and Gibson T.J. (1994) *Nucleic Acids Res*; 22:4673-4680.
- [47] Guex N. and Peitsch M.C. (1997) *Electrophoresis*, 18:2714-2723.
- [48] Karlin S. and Brocchieri L. (2000) *PNAS* 10, 97 (21):11349.
- [49] Brocchieri L. and Karlin S. (2000) *Protein Science*, 9:476-486.
- [50] Itoh H., Komatsuda A., Ohtani H., Wakui H., Imai H., Sawada K., Otaka M., Ogura M., Suzuki A. and Hamada F. (2002) *Eur. J. Biochem.* 269:5931-5938.
- [51] Renner T. and Waters E. R. (2007) *Cell Stress & Chaperones* 12 (2):172-185.
- [52] Gupta R. S. and Singh B. (1994) *Current Biology*, 4:1104-1114.
- [53] Krebs A. R. and Bettencourt B.R. (1999) *Amer. Zool.*, 39:910-919.
- [54] Hohfeld J., Cyr M. D. and Patterson C. (2001) *EMBO reports*, 2(10):885-890.
- [55] Gehrman M., Radons J. and Molls M. (2008) *Cell Stress and Chaperones*; 13:1-10.
- [56] Shaner L. and Morano A. K. (2007) *Cell Stress and Chaperones*; 12(1):1-8.
- [57] Gupta R. S. (1995) *Mol. Bio. Evol.* 12(6):1063-1073.
- [58] Chen B., Zhong D. and Monteiro A. (2006) *BMC Genomics*, 7:156.
- [59] Sgobba M., Degliesposti G., Ferrari A. M. and Rastelli G. (2008) *Chem Biol Drug Des*; 71: 420-433.



Homo sapiens
Mus musculus
D. melanogaster
A. thaliana
S. cerevisiae
E. coli



Homo sapiens
Mus musculus
D. melanogaster
A. thaliana
S. cerevisiae
E.coli



HUMAN 1A

DOMAIN E

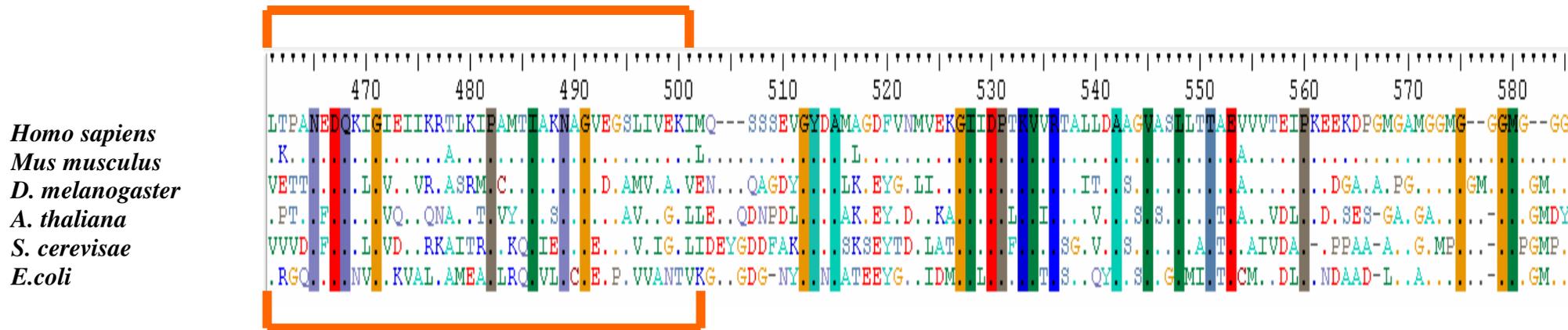
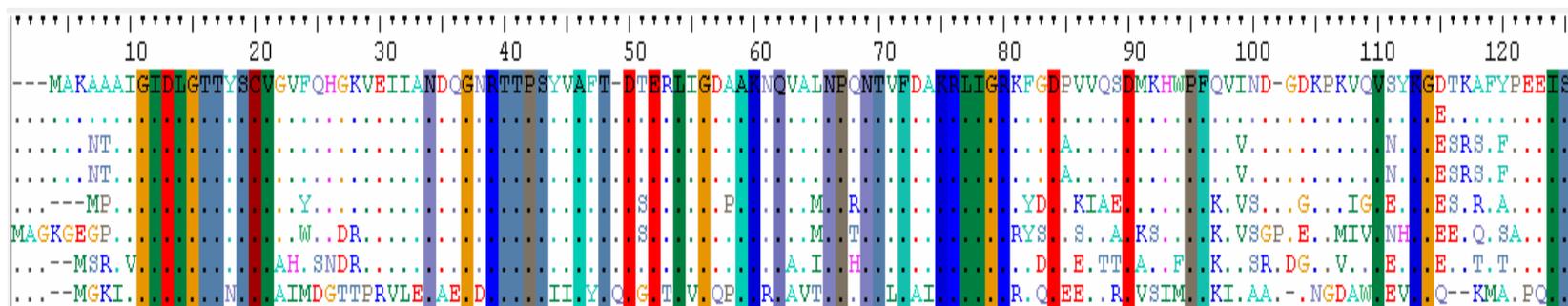
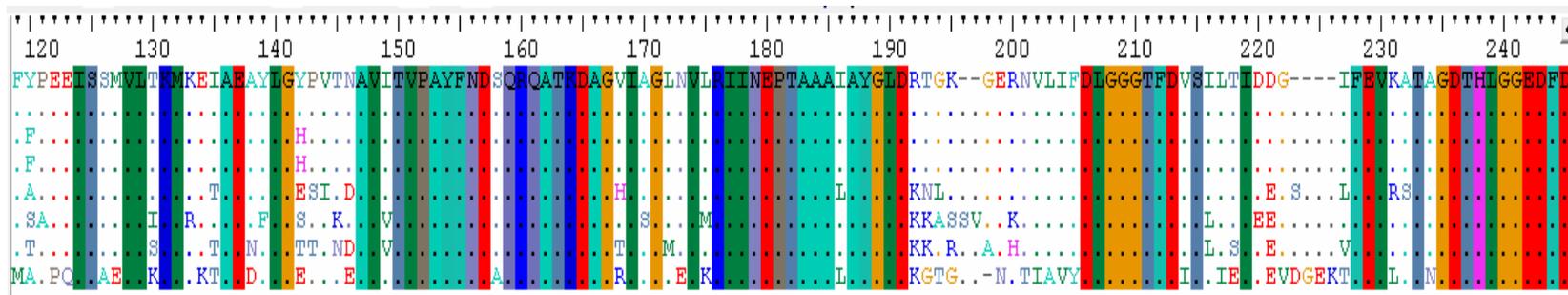


Fig. 1- Multiple sequence alignment of HSP 60 sequences. The colored columns in the alignment signifies complete conservation of the sequence. Each dot shows the similar amino acid to that of the amino acid present in the topmost sequence. Region enclosed in orange, blue and green box represents domain E, domain I and domain A respectively

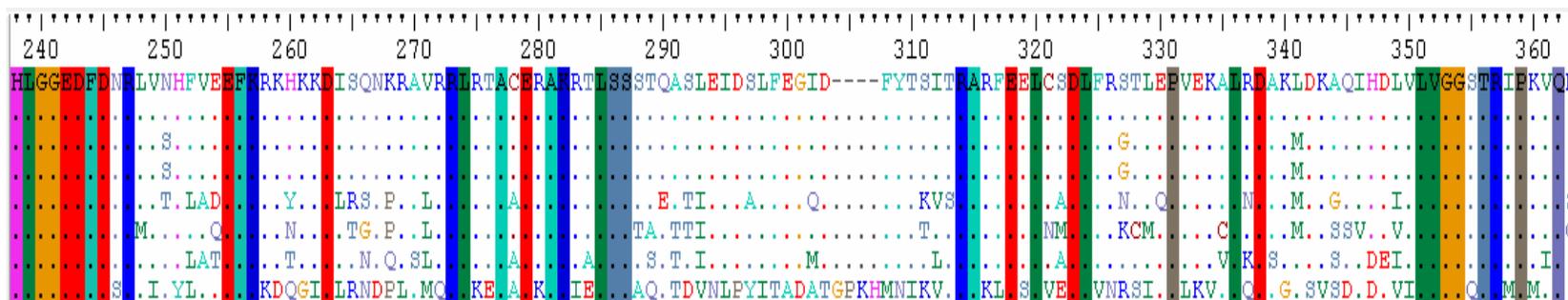
Homo sapiens HSP 70 1a
Homo sapiens HSP70 1b
Mus musculus HSP70 1b
Mus musculus HSP70 1a
D. melanogaster
A. thaliana
S. cerevisiae
E. coli



Homo sapiens HSP70 1a
Homo sapiens HSP70 1b
Mus musculus HSP70 1b
Mus musculus HSP70 1a
D. melanogaster
A. thaliana
S. cerevisiae
E. coli



Homo sapiens HSP70 1a
Homo sapiens HSP70 1b
Mus musculus HSP70 1b
Mus musculus HSP70 1a
D. melanogaster
A. thaliana
S. cerevisiae
E. coli



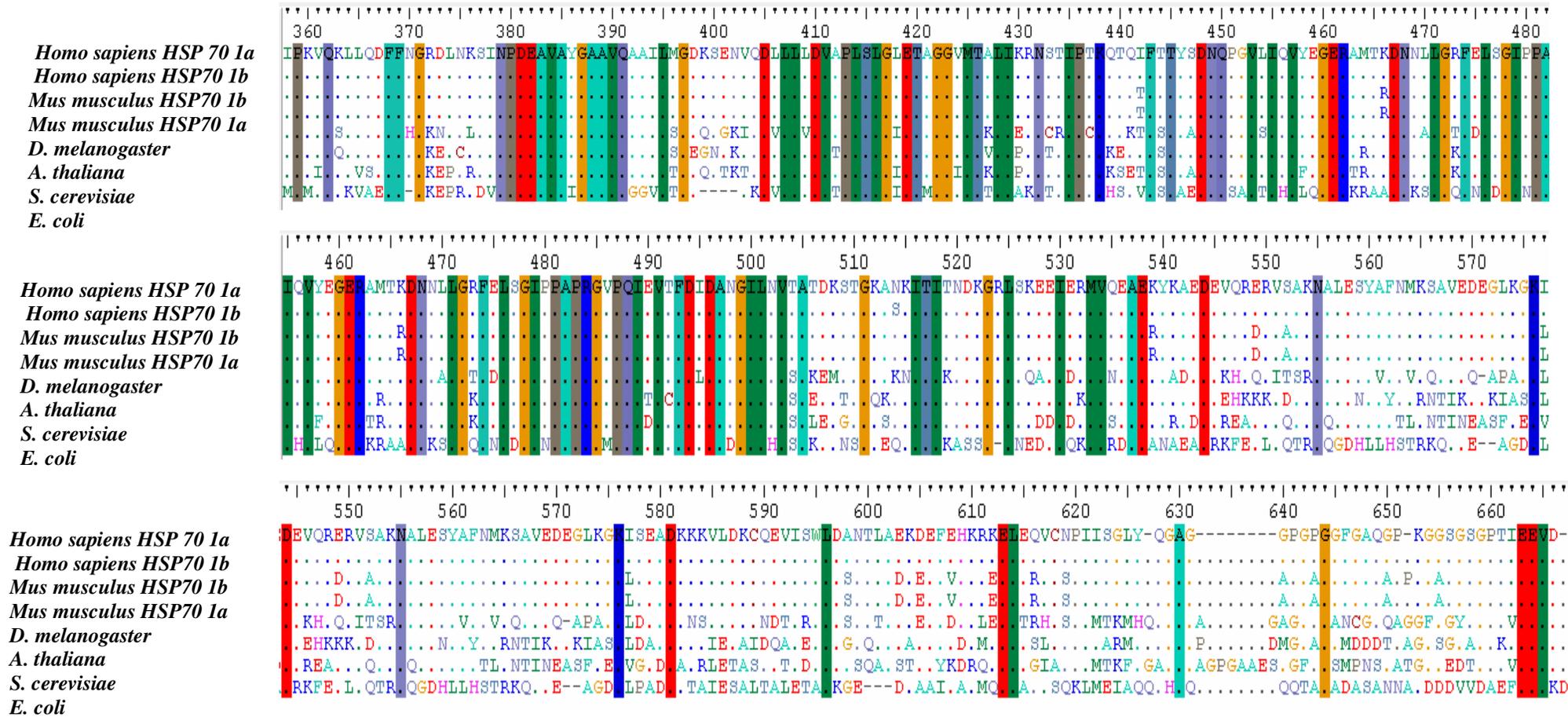


Fig. 4- Multiple sequence alignment (MSA) between all the sequences of HSP 70 family from different organism. Each dot in the alignment denotes the match with the top most sequence. And the colored regions signify complete conservation of amino acid in all the sequences

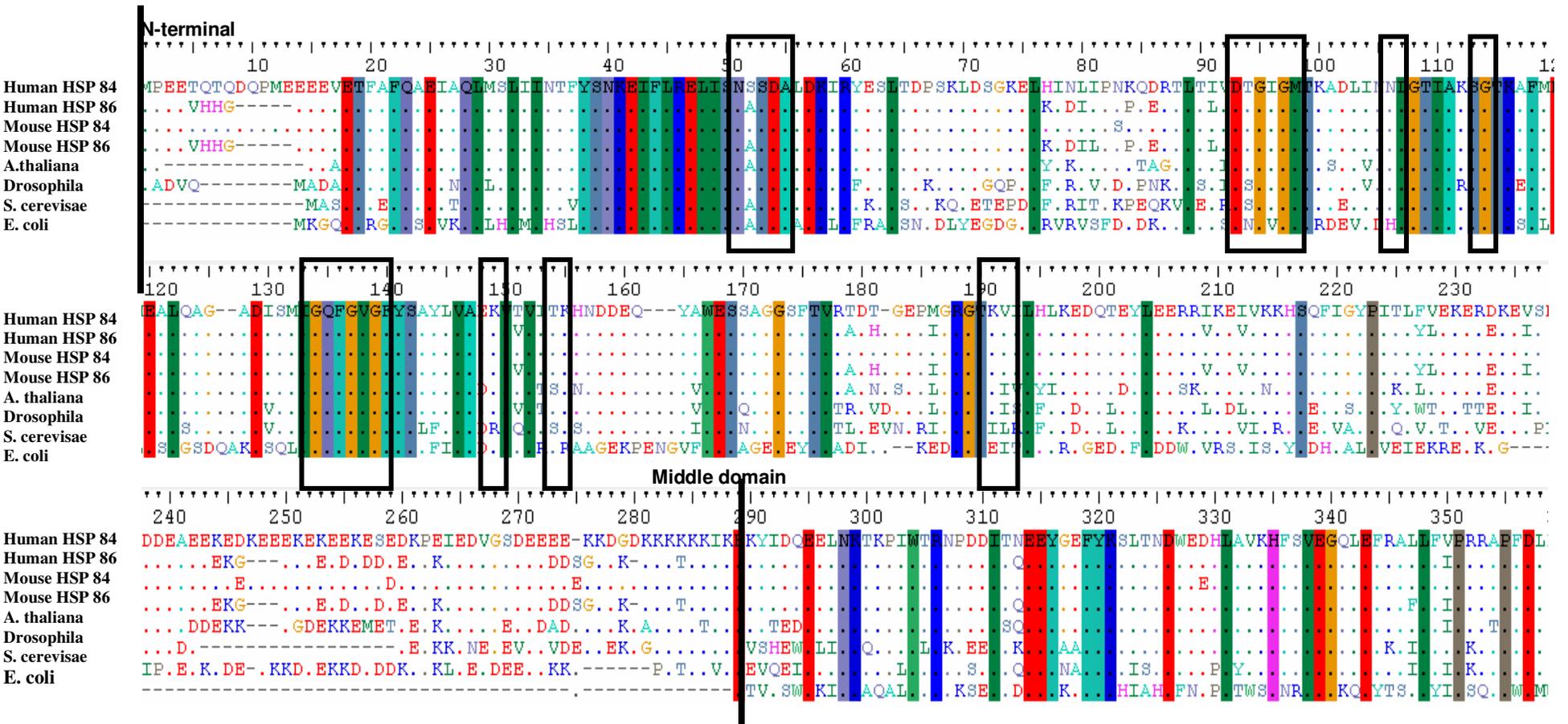


Table 1- Sequences selected for the analysis with their accession number and reference.

FAMILY	ORGANISM	ACCESSION NUMBER	LENGTH (Amino Acids)	REFERENCE
HSP 60	<i>E.coli</i>	NP_290776	548	Perna <i>et al</i> , 2001 [20]
	<i>S. cerevisiae</i>	NP_013360	572	Johnston, M. <i>et al</i> ;1997[21]
	<i>A. thaliana</i>	NP_189041	577	Jindal S <i>et al</i> , [22]
	<i>D. melanogaster</i>	CAA67720		Kozlova, T <i>et al</i> ;Unpublished [23]
	<i>Mus musculus</i>	P63038	573	Lotscher, E and Allison,J.P. 1990 [24]
	<i>Homo sapiens</i>	P10809	573	Jindal,S., <i>et al</i> ; 1989 [25]
HSP 70	<i>E. coli</i>	NP_285706	638	Perna, N.T., <i>et al</i> ;2001[26]
	<i>S. cerevisiae</i>	EDN64545	649	Wei, W., <i>et al</i> ;2007 [27]
	<i>A. thaliana</i>	NP_187864	650	Genpept*
	<i>D. melanogaster</i>	P02825	642	Karch,F., <i>et al</i> ; 1981 [28]
	<i>Mus musculus</i> HSP 70 1a HSP 70 1b	NP_034609 NP_034608	641 642	Goto, K., <i>et al</i> ; 2007 [29] Belova, G.I., 2008 [30]
	<i>Homo sapiens</i> HSP 70 1a HSP 70 1b	NP_005336 NP_005337	641 641	Wu,B., <i>et al</i> ; 1985 [31] Wu,B., <i>et al</i> ; 1985 [31]
	<i>E. coli</i>	P0A6Z3	624	Bardwell,J.C. and Craig,E.A. (1987)[32]
HSP 90	<i>S. cerevisiae</i>	NP_015084	709	Bussey,H. <i>et al</i> ;1997 [33]
	<i>A. thaliana</i>	CAA68885	704	Milioni,D. and Hatzopoulos,P.;1997 [34]
	<i>D. melanogaster</i>	NP_523899	717	Hoskins,R.A., 2007 [35]
	<i>Mus musculus</i> HSP 86 alpha HSP 84 beta	P07901 P11499	733 724	Moore,S.K., 1991 [36] Moore,S.K., 1990 [37]
	<i>Homo sapiens</i> HSP 86 alpha HSP 84 beta	P07900 P08238	732 724	Yamazaki,M. <i>et al</i> ,1989[38] Rebbe,N.F., <i>et al</i> ,1987 [39]

Table 2- Template selected for each of the species and the percentage of sequence identity observed

FAMILY	SEQUENCE	PROTEIN TEMPLATE(PDB ID)	Sequence identity (%)
HSP 60 family	<i>Homo sapiens</i>	1sx3E	50
	<i>Mus musculus</i>	1sx3E	50
	<i>D. melanogaster</i>	1sx3E	52
	<i>A. thaliana</i>	1kp8H	56
	<i>S. cerevisiae</i>	1kp8H	54
	<i>E. coli</i>	1sx3E	99
HSP 70 family	<i>Homo sapiens</i> HSP 70 1a	1yuwA	88.448
	<i>Homo sapiens</i> HSP 70 1b	1yuwA	88.448
	<i>Mus musculus</i> HSP 70 1a	1yuwA	88.267
	<i>Mus musculus</i> HSP 70 1b	1yuwA	88.267
	<i>D. melanogaster</i>	1yuwA	78.947
	<i>A. thaliana</i>	3c7nB	80.839
	<i>S. cerevisiae</i>	3c7nB	79.889
	<i>E. coli</i>	2v7yA	62.734

HSP 90 family	<i>Homo sapiens</i> HSP 90 alpha N Terminal C Terminal	1osfA 2cgeA	100 62.716
	<i>Homo sapiens</i> HSP 90 beta N Terminal C Terminal	1uymA 2cgeA	100
	<i>Mus musculus</i> HSP 90 alpha N Terminal C Terminal	1osfA 2cgeA	99.535 62.222
	<i>Mus musculus</i> HSP 90 beta N Terminal C Terminal	1uymA 2cgeA	99.522 62.222
	<i>D. melanogaster</i> N Terminal C Terminal	2bz5A 2cgeA	86.19 61.975
	<i>A. thaliana</i>	2cg9B	63.024
	<i>S. cerevisiae</i>	2cg9B	89.774
	<i>E. coli</i>	2iopA	99.038

Table 3-Percentage of conservation of amino acids in HSP 60

SEQUENCE CONSERVATION			
Organisms	Total aa	Conserved/total aa	Percentage of sequence conservation
<i>E. coli</i>	548	187/548	34.12%
<i>S. cerevisiae</i>	572	187/572	32.69%
<i>D. melanogaster</i>	573	187/573	32.63%
<i>A. thaliana</i>	576	187/576	32.46%
<i>Mus musculus</i>	573	187/573	32.63%
<i>Homo sapiens</i>	573	187/573	32.63%
EUKARYOTIC SEQUENCE CONSERVATION			
Organisms	Total aa	Conserved/total aa	Percentage of sequence conservation
<i>S. cerevisiae</i>	572	236/572	41.26
<i>D. melanogaster</i>	573	236/573	41.19
<i>A. thaliana</i>	576	236/576	40.97
<i>Mus musculus</i>	573	236/573	41.19
<i>Homo sapiens</i>	573	236/573	41.19

Table 4- Percent similarity of amino acid sequences belonging to HSP 60 family

ORGANISMS	A	B	C	D	E	F
<i>E. coli</i> (A)	100.00	52.5547	52.1898	55.4745	49.4526	49.4526
<i>S. cerevisiae</i> (B)		100.00	53.1469	56.6434	53.3217	53.1469
<i>D. melanogaster</i> (C)			100.00	54.1012	53.3217	71.0297
<i>A. thaliana</i> (D)				100.00	56.1955	56.0209
<i>Mus musculus</i> (E)					100.00	97.5567
<i>Homo sapiens</i> (F)						100.00

Table 5- Percentage of conservation of amino acid in all HSP 70 homologues

Eukaryotic sequence conservation			
Organism	Total aa	conserved/total aa	percentage of sequence conservation
<i>Homo sapiens HSP70 1a</i>	641	376/641	58.65%
<i>Homo sapiens HSP70 1b</i>	641	376/641	58.65%
<i>Mus musculus HSP70 1a</i>	641	376/641	58.65%
<i>Mus musculus HSP70 1b</i>	641	376/641	58.65%
<i>A. thaliana</i>	650	376/650	57.84%
<i>D. melanogaster</i>	642	376/642	58.56%
<i>S. cerevisiae</i>	649	376/649	57.93%
Sequence conservation across all domains			
Organism	Total aa	conserved/total aa	percentage of sequence conservation
<i>Homo sapiens HSP70 1a</i>	641	247/641	38.50%
<i>Homo sapiens HSP70 1b</i>	641	247/641	38.50%
<i>Mus musculus HSP70 1a</i>	641	247/641	38.50%
<i>Mus musculus HSP70 1b</i>	641	247/641	38.50%
<i>A. thaliana</i>	650	247/650	38.00%
<i>D. melanogaster</i>	642	247/642	38.47%
<i>S. cerevisiae</i>	649	247/649	38.05%
<i>E. coli</i>	638	247/638	38.71%

Table 6- Percentage of amino acid sequence similarity between pair of protein belonging to HSP 70 family.

TAXA	A	B	C	D	E	F	G	H
<i>Homo sapiens hsp 70 1a</i> (A)	100	99.69	95.16	99.17	75.04	73.947	72.69	43.7304
<i>Homo sapiens hsp 70 1b</i> (B)		100	95.16	94.69	7504.00%	74.1	72.07	43.7304
<i>Mus musculus hsp 70 1a</i> (C)			100	99.53	7410.00%	74.88	70.04	42.79
<i>Mus musculus hsp 70 1b</i> (D)				100	7429.00%	74.61	70.71	42.79
<i>A. thaliana</i> (E)					100	67.28	68.1	43.2602
<i>D. melanogaster</i> (F)						100	68.69	43.1034
<i>S. cerevisiae</i> (G)							100	44.2006
<i>E. coli</i> (H)								100

Table 7- Percentage of amino acid (aa) sequence conservation in HSP 90 homologues

Percentage of sequence conservation			
Organism	Total aa	Conserved/Total aa	Percentage of sequence conservation
<i>Homo sapiens HSP 84</i>	724	173/724	23.89%
<i>Homo sapiens HSP 86</i>	732	173/732	23.63%
<i>Mus musculus HSP 86</i>	733	173/733	23.60%
<i>Mus musculus HSP 84</i>	724	173/724	23.89%
<i>A. Thaliana</i>	704	173/704	24.57%
<i>D. melanogaster</i>	717	173/717	24.12%
<i>S. cerevisiae</i>	709	173/709	24.40%
<i>E. coli</i>	624	173/624	27.72%

Eukaryotic sequence conservation			
Organism	Total aa	Conserved/Total aa	Percentage of sequence conservation
<i>Homo sapiens HSP 84</i>	724	391/724	54.00%
<i>Homo sapiens HSP 86</i>	732	391/732	53.41%
<i>Mus musculus HSP 86</i>	733	391/733	53.34%
<i>Mus musculus HSP 84</i>	724	391/724	54.00%
<i>A. Thaliana</i>	704	391/704	55.54%
<i>D. melanogaster</i>	717	391/717	54.53%
<i>S. cerevisiae</i>	709	391/709	55.14%

Table 8- Percentage of similarity between pair of HSP90 homologues

TAXA	A	B	C	D	E	F	G	H
<i>Homo sapiens HSP 84 (A)</i>	100	85.91	85.08	99.17	68.47	78.24	60.65	34.78
<i>Homo sapiens HSP 86 (B)</i>		100	98.77	85.91	70.31	78.1	60.37	35.26
<i>Mus musculus HSP 86 (C)</i>			100	85.08	70.03	77.68	59.8	34.94
<i>Mus musculus HSP 84 (D)</i>				100	68.32	77.82	60.65	34.45
<i>A. Thaliana (E)</i>					100	67.33	62.64	34.61
<i>D. melanogaster (F)</i>						100	58.96	33.65
<i>S. cerevisiae (G)</i>							100	35.09
<i>E. coli (H)</i>								100

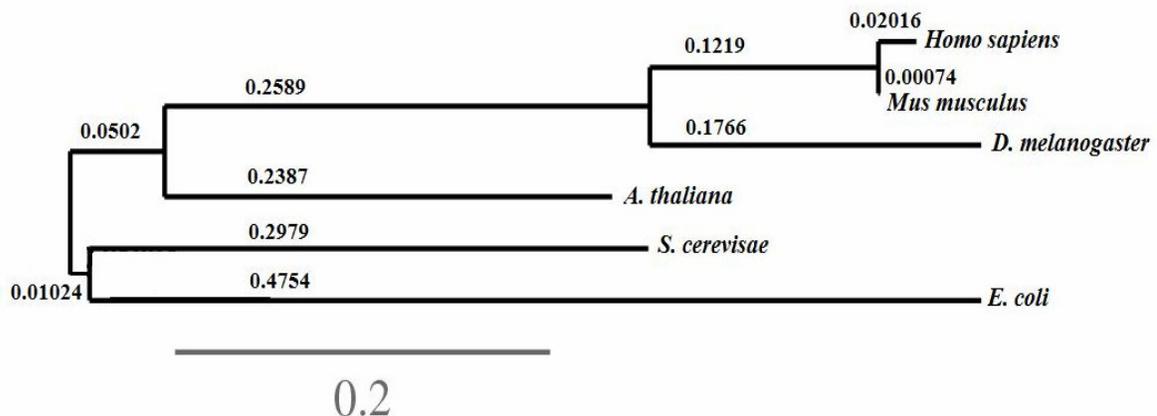


Fig. 2-Phylogenetic tree based on maximum likelihood method showing evolutionary relationship between all HSP60 homologues. The branch length shown is directly proportional to the number of character changes.

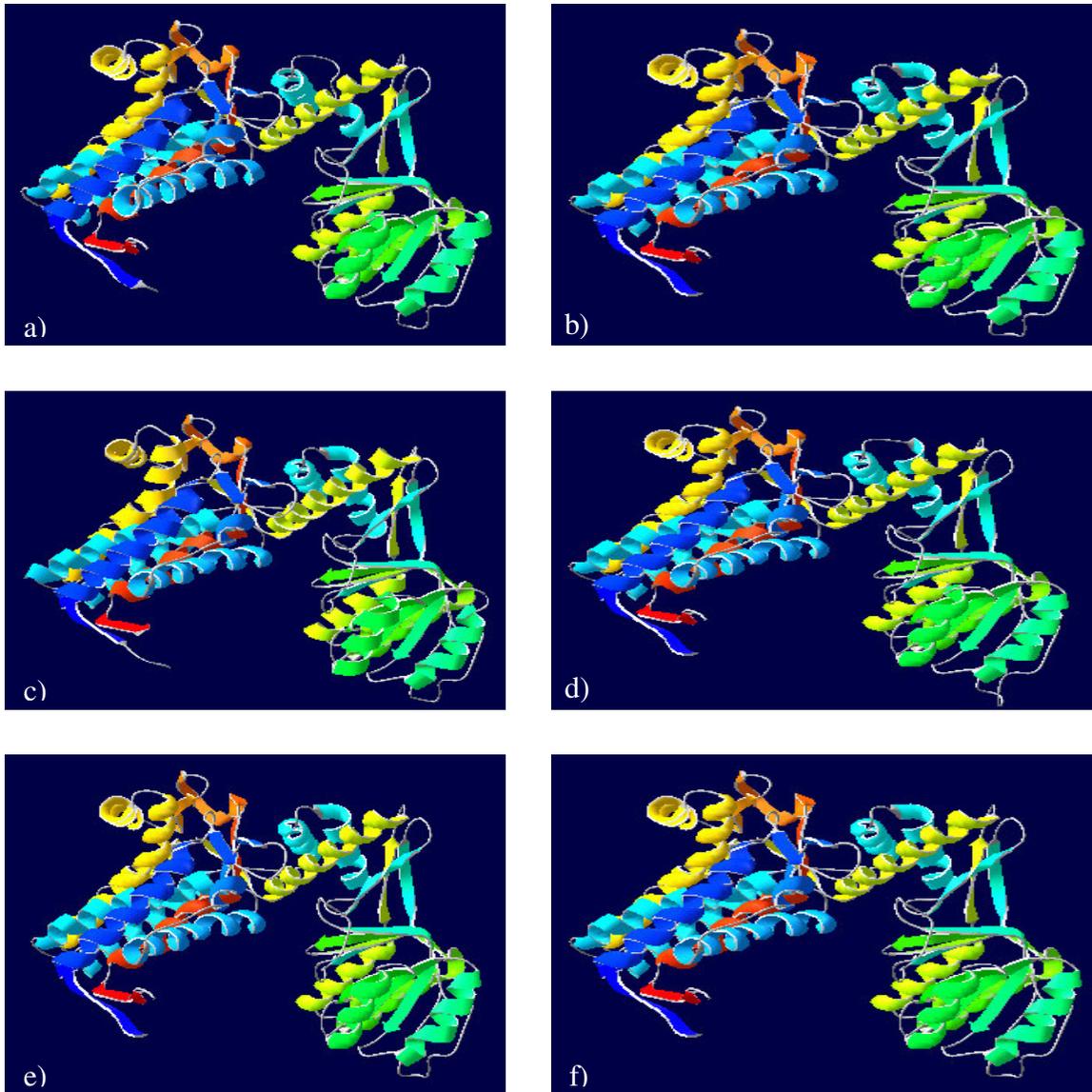


Fig. 3- Homology modeled structures for a) *E. coli* b) *S. cerevisiae* c) *Arabidopsis thaliana* d) *Drosophila melanogaster* e) *Mus musculus* f) *Homo sapiens*

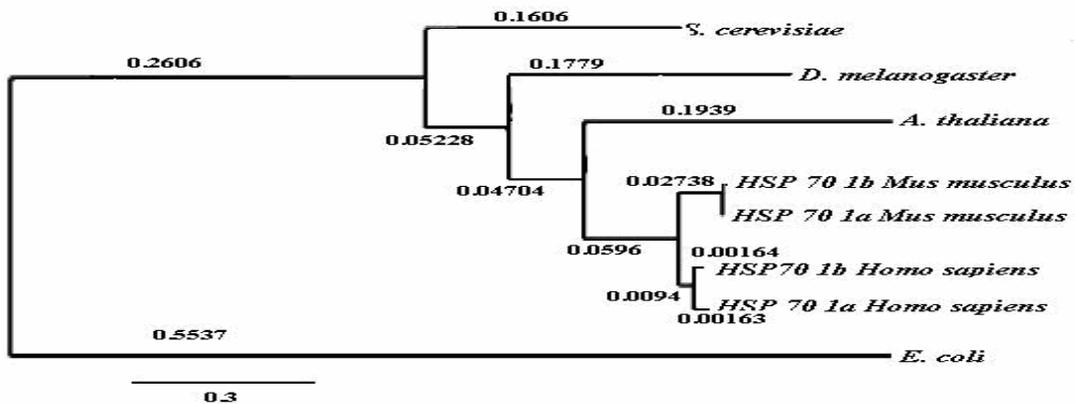


Fig. 5- Phylogenetic tree based on maximum likelihood method showing evolutionary relationship between all HSP70 homologues. The branch length shown is directly proportional to the number of character changes

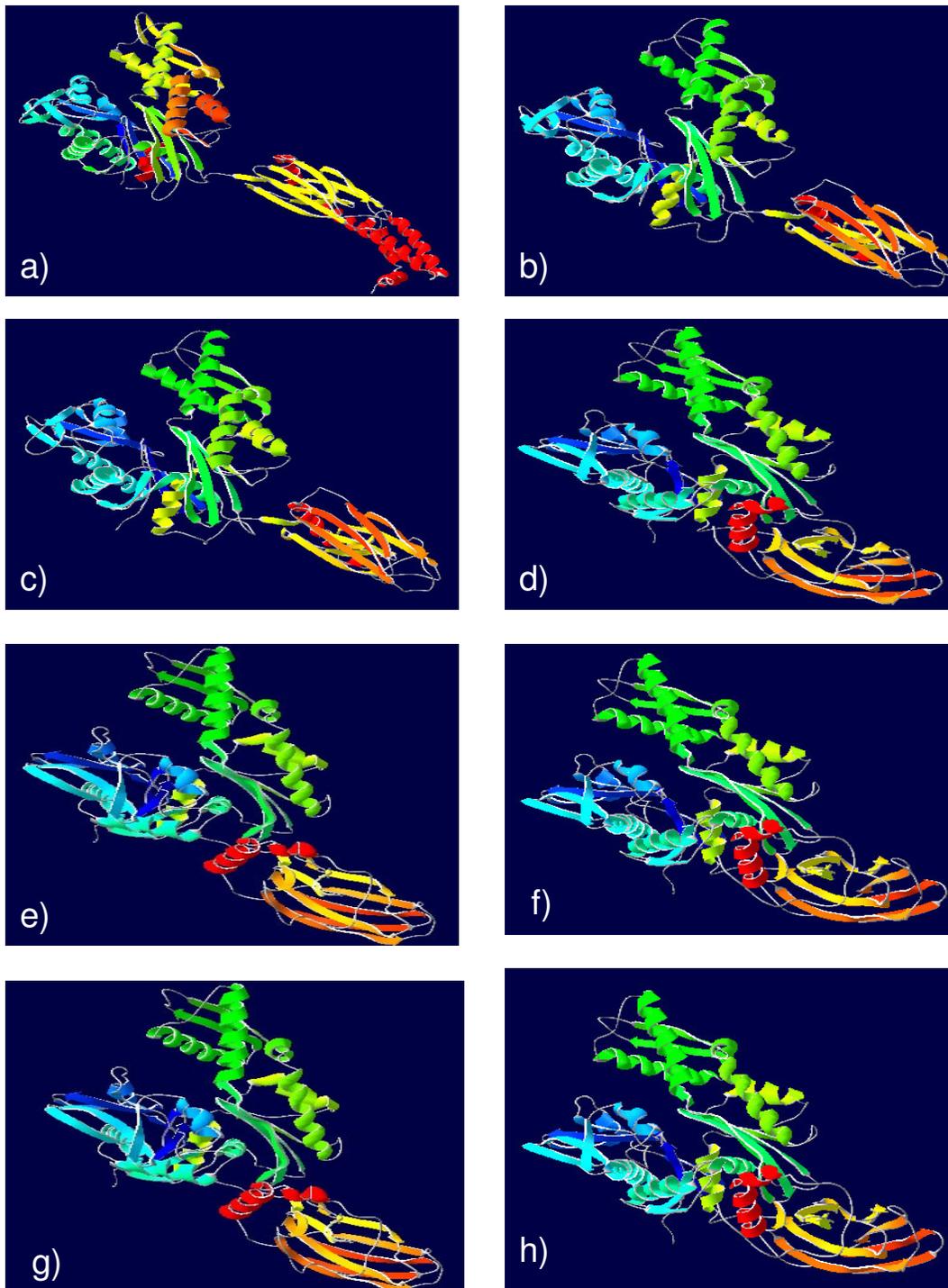


Fig. 6-Homology modeled structures for a) *E. coli* b) *S. cerevisiae* c) *Arabidopsis thaliana* d) *Drosophila melanogaster* e) *Mus musculus* HSP70 1a f) *Mus musculus* HSP70 1b g) *Homo sapiens* HSP70 1a h) *Homo sapiens* HSP70 1b

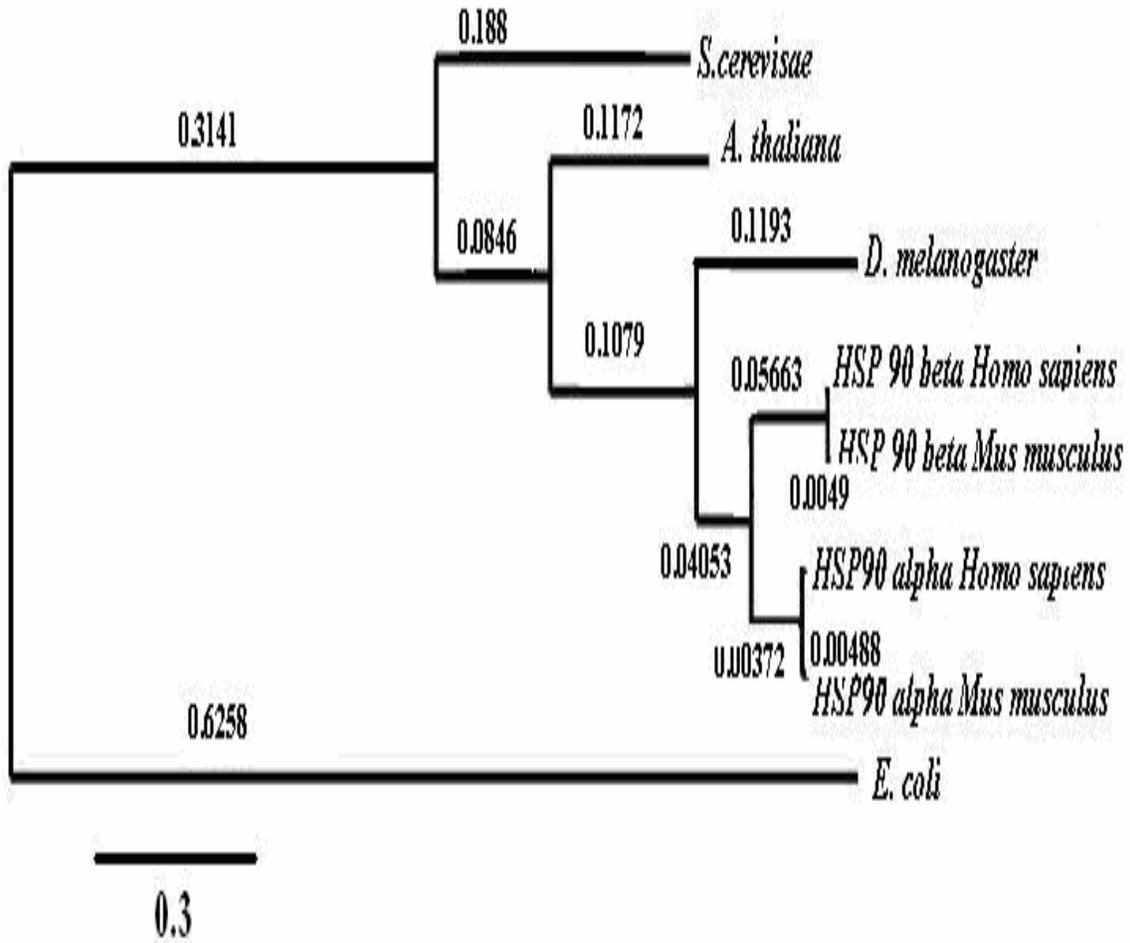


Fig. 8-Phylogenetic tree based on maximum likelihood method showing evolutionary relationship between all HSP60 homologues. The branch length shown is directly proportional to the number of character changes.

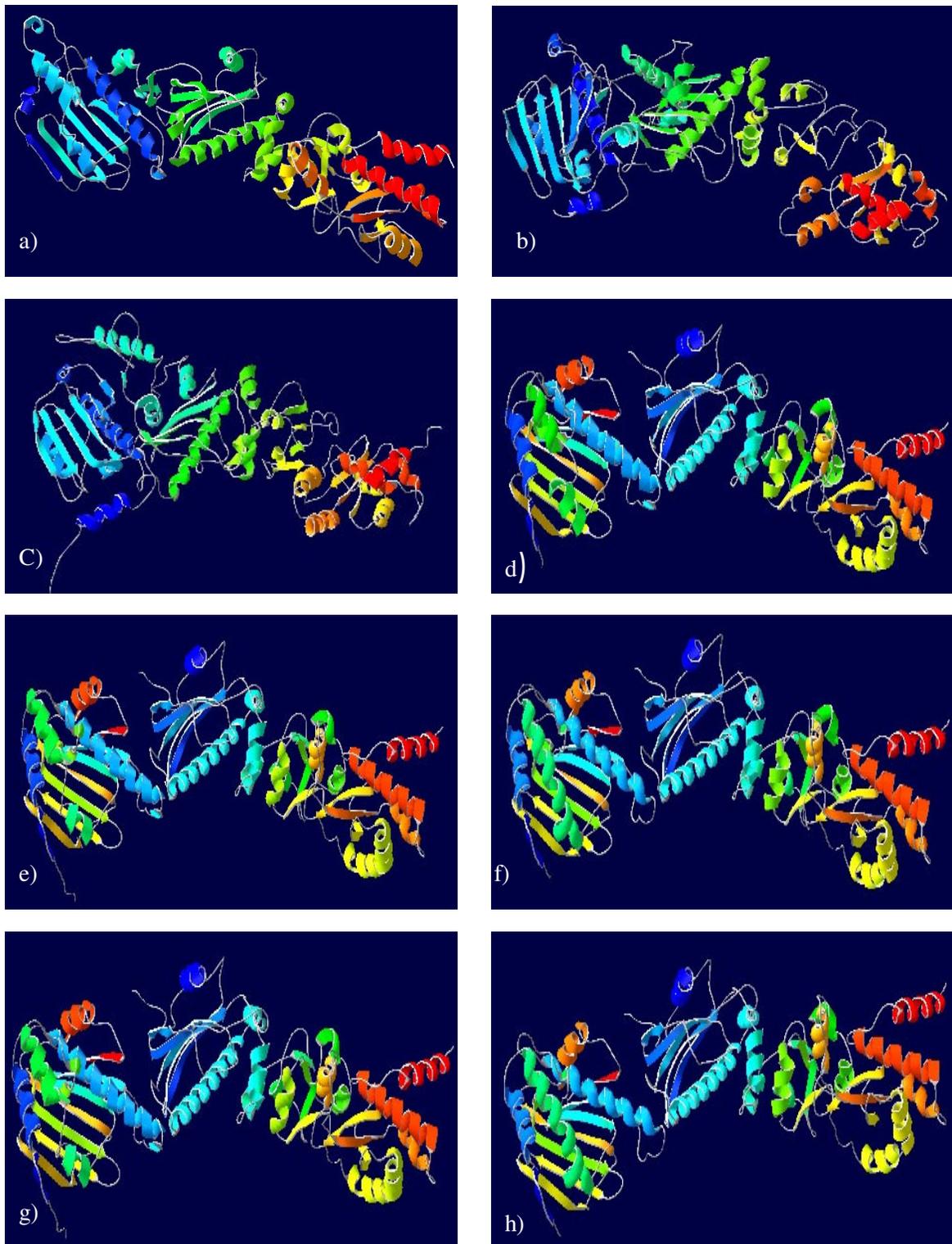


Fig. 9-Homology modeled structures for a) *E.coli* b) *S.cerevisiae* c) *Arabidopsis thaliana* d) *Drosophila melanogaster* e) *Musmusculus* HSP 90 alpha f) *Musmusculus* HSP 90 beta g) *Homo sapiens* HSP 90 alpha h) *Homo sapiens* HSP 90 beta.