A COMPARATIVE *IN SILICO* STUDY FINDS A FUNCTIONAL CO-RELATION BETWEEN HUMAN HSP7C AND PLANT PROTEIN.

Animesh Sarker[1], and Al Amin[2]

[1]Department of Biotechnology and Genetic Engineering, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902.Bangladesh.

Telephone:+088-092162407

animesh_du_geb@yahoo.com

[2]Department of Biotechnology and Genetic Engineering, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902.Bangladesh.

Telephone:+088-092162407
niloygenetics@gmail.com

ABSTRACT

A bioinformatics finding shows that numerous human heat shock proteins (HSPs) homologues are common in plants. Human heat shock proteins (HSPs), which are expressed to higher temperature or other stress, have chaperone activity belong to four conserved classes: HSP60, HSP70, HSP90 and HSP100. Bioinformatics blast search reveals that each of the human HSP classes possess a number of plant homologues. The closest identified plant homologue of human HSP_7C is a protein of unknown function (NCBI Accession XP_002332067) derived from Populus trichocarpa. In silico comparative studies have showed invigorating similarity of human HSP_7C and the designated protein of Populus trichocarpa. Secondary and three-dimensional (3D) structure analysis of the predicted plant protein strongly supports its functional relationship to the class of human HSP70.

KEY WORDS

HSPs, Plant homologue, Comparative studies, Functional relationship

1. INTRODUCTION

Heat shock proteins (HSPs) are a class of functionally related proteins whose expression is increased when cells are exposed to elevated temperature or other stresses [1]. HSPs are found in virtually all living organisms, from bacteria to human. Heat shock proteins, as a class, are among the most highly expressed cellular proteins across all species. As their name implies, heat shock proteins protect cells when stressed by elevated temperatures. They account for 1–2% of total protein in unstressed cells. However when cells are heated, the fraction of heat shock proteins increases to 4–6% of cellular proteins [2, 3].

In eukaryotic organisms, the principal class of HSPs in human are HSP60, HSP70, HSP90 and HSP100 [4]. Another novel class HSP33 is also found exclusively, which interestingly absence in archaea or other eukaryotes [5]. In plants, Hsps are greatly varied in level of expression as well as their type [6]. The most prominent types are HSP20, HSP70, HSP90 and HSP100 according to well studied plant species Arabidopsis thaliana. Higher plants are characterized by the presence of at least 20 types of sHsps, but one species could contain 40 types of these sHsps[7].

In plants, most common class protein is Hsp70 has been characterised in diverse species [7,8]. The simplest flowering plant genome Arabidopsis thaliana includes at least 18 genes encoding members of the Hsp70 family, of which 14 belong to the DnaK subfamily and four to the Hsp110/SSE subfamily [3,9,10]. At least 12 Hsp70 members have been found in the spinach genome [11].Expression profile analysis of the Arabidopsis and spinach Hsp70 genes demonstrated that members of Hsp70 chaperones are expressed in response to environmental stress conditions such as heat, cold and drought, as well as osmotic, salinity, oxidative, desiccation, high intensity irradiations, wounding, and heavy metals stresses [9-11,12].

On the basis of sequence and structure similarities in silico comparative studies have been performed between human HSPs and homologous plant protein. Most of the computer aided bioinformatics findings reveal that proteins for HSP70 family are mostly analogous to its plant homologues. BLASTp search of the entire Angiosperm plantae kingdom against all of the human HSP classes discover a protein of unknown function from *Populus trichocarpa* closely similar to human HSP70. By using different computational tools and software the sequence and structure of this unknown protein was intensely compared with human HSP70 which discovered a functional co-relation between these two proteins.

The goal of the present work was to identify plant homologues of human HSPs that could play significant role in protein folding, assembly, translocation and degradation in many normal cellular processes, stabilize proteins and membranes, and could assist in protein refolding under stress conditions. They could also play a crucial role in protecting plants against stress by re-establishing normal protein conformation and thus cellular homeostasis[13]. It seems that the synthesis of these proteins require enormous energy that impact on the yield of the organism [14].

2. MATERIALS AND METHODS

The full-length protein sequences of all the four types heat shock proteins (HSPs) were retrieved from NCBI protein Database (http://www.ncbi.nlm.nih.gov/), with accession numbers, HSP60 (P10809), HSP70 (P11142), HSP90 (P07900), HSP100 (Q92598). The homologous sequences of human HSPs were identified from position specific iteration BLASTp (NCBI). Percent similarities and identities were computed by NCBI blast2seq program. Global alignment of the coding sequences was performed with the program AVID using a window size of 100 bp and a conservation level of 70%. The results were viewed with the program VISTA. To map sequence conservation, human HSP7c were aligned with several other plant proteins using multiple sequence alignment tool ClustalW2 (http://www.ebi.ac.uk/tools/clustalw2). This program uses a progressive method to build its alignments using the BLOSUM 62 substitution matrix for proteins. Phylogenetic tree was also constructed for HSP7C_HUMAN and its plant homologues by using ClastalW. Secondary structure of the predicted HSP_Populus1 and HSP7C_Human were computed using the program Hierarchical Neural Network (HNN: http://www.expasy.org/tools/). The 3D structure of predicted protein and HSP7C were constructed by DeepView/Swiss-PdbViewer v4.0.1(spdbv) (http://www.expasy.org/spdbv/). Swiss-PdbViewer is tightly linked to SWISS-MODEL, an automated homology-modelling server developed by the Structural Bioinformatics Group of the Swiss Institute of Bioinformatics (SIB) at the Biozentrum in Basel. To comprehend the functional relationship between these two proteins, SVMProt was used to predict protein function.

3. RESULTS AND DISCUSSION

3.1 Homology study:

The primary knowledge of Human heat shock proteins, whose expression rises when cells are unprotected to higher temperatures or other stress, were collected from public databases and published research papers in order to homology study [10]. NCBI protein blast (BLASTp) was

carried out intended for scanning Human Heat Shock Proteins arranged in **(Table 1),** against the kingdom of flowering plants. A number of plant homologues (Angiosperm) of human heat shock protein 60, 70, 90 and 100 were identified on the basis of better E-value and higher percentage of similarity-identity **(**T**able 2)**.

Most significant members of each family are tabulated here, however there are some species which may express additional chaperones, co-chaperones, and heat shock proteins. In addition, many of these proteins may have multiple splice variants (Hsp90α and Hsp90β, for instance) or conflicts of nomenclature (Hsp72 is sometimes called Hsp70) [17].

Since UniProt and NCBI database carries enormous sequence of human heats hock proteins, so prior to blast search, one candidate from each of the four classes were selected (**Table 1**). These are: Heat shock protein 105 kDa (GI:97536358), Heat shock protein 90A kDa (GI:92090606), Heat shock protein 70C kDa (GI: 123648), Heat shock protein 60 kDa (GI:41399285).

Table 1. Selected candidates of human heat shock protein (HSPs).

HSPs	NCBI (GI)	UniProt (Accession)
HSP60	GI: 41399285	P10809 (HSP60_Human)
HSP70	GI: 123648	P11142 (HSP7C_HUMAN)
HSP90	GI: 92090606	P07900 (HS90A_HUMAN)
HSP100	GI:97536358	Q92598 (HSP105_Human)

Position specific iterated BlAST scanning of each the candidate HSPs hits a lot of predicted plant homologous proteins (those function is not identified yet). Two of those hypothetical members from each class were selected by filtering proteins with better E-values and higher percentage of similarities-identities **(Table 2)**.

3.2 Phylogenetic study

According to BLAST result it is clear that two members of *Populus trichocarpa* have higher sequence similarity to human HSP7C as one (Gb|EEE71404.1|) hits with 88% similarity and 78-79% identity and other (Gb|EEE71404.1|) hits with 88% similarity and 78% identity. In order to identify the closest homologues, these two potential candidates are allowed for pair wise sequence alignments with several other flowering plant proteins (**Table 3**), which were selected from BLAST result of HSP7C . The members of higher percentage of sequence similarity and identity were allowed for MSA and finally a phylogenetic tree was constructed. According to this cladistic data, a common clade with human heat shock protein(HSP7C) is formed by protein gb|EEE70430.1| (gi|224115828|HSP_ Populus1)(**Figure 1**).

Table 2. Plant homologues of each the human HSP class.

HSPs	Predicted Protein Description	E value	Similarity	Identity
HSP 105_Human	gb|EFH64112.1| hypothetical protein ARALYDRAFT_896005 [Arabidopsis lyrata subsp. lyrata]	1e-165	59%	40%
	gb|EES19148.1| hypothetical protein SORBIDRAFT_09g005570 [Sorghum bicolor]	4e-162	58%	39%

HS90A_HUMAN	gb|EEE74469.1| predicted protein [Populus trichocarpa]	0	83%	71%
	gb|EEE85774.1| predicted protein [Populus trichocarpa]	0	84%	71%
HSP7C_HUMAN	**gb|EEE71404.1| predicted protein [Populus trichocarpa]**	**0**	**88%**	**79%**
	gb|EEE70430.1| predicted protein [Populus trichocarpa]	**0**	**88%**	**78%**
HSP60_Human	gb|EEE82364.1| predicted protein [Populus trichocarpa]	0	79%	60%
	gb|EER94121.1| hypotheticalprotein SORBIDRAFT_01g020010[Sorghum bicolor]	0	79%	60

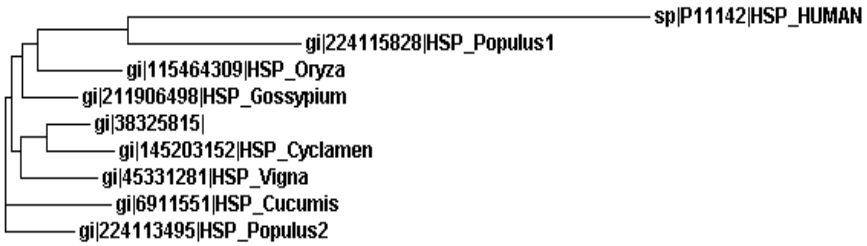

Figure 1. Phylogenetic tree of human HSP7C with its plant homologues.

Table 3. Sequence similarities and identities between HSP7C_Human and selected plant homologues.

GI No.	Accession No.	Organism	Length (aa)	Identity (%)	Similarity (%)	E-value
38325815	AAR17080.1	Nicotiana tabacum	648	78	89	0
115464309	ABF95267.1	Oryza sativa Japonica Group	653	79	88	0
211906498	ACJ11745.1	Gossypium hirsutum	648	78	88	0
45331281	AAS57912.1	Vigna radiata	649	78	88	0
145203152	ABP35942.1	Cyclamen persicum	650	77	87	0
6911551	CAB72129.1	Cucumis sativus	653	77	87	0
224115828	XP_002332067.1	Populus trichocarpa	655	78	88	0
224113495	XP_002332589.1	Populus trichocarpa	648	78	88	0

3.3 Sequence conservation study

Global alignment of the coding sequence of Human HSP7c and predicted protein gene sequence was performed with the program AVID using a window size of 100 bp and a conservation level

of 70%. Results were viewed with the program VISTA (**Figure** 2). From here it can be observed that these two sequences are highly conserved with each other.

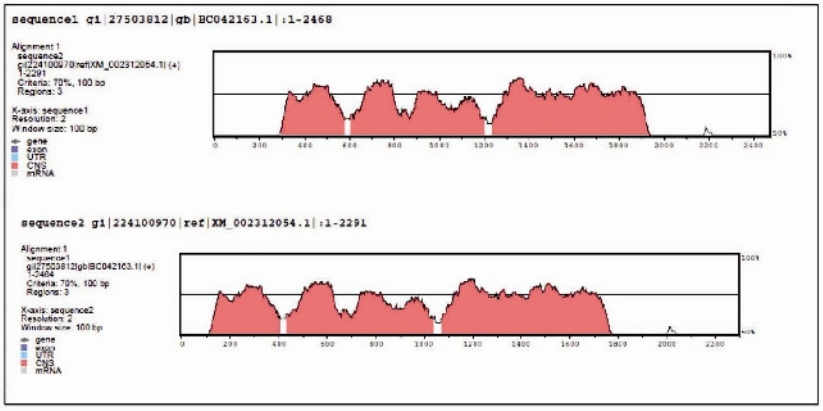

Figure 2. Conservation level of human HSP7C and predicted populus1 protein.

VISTA plot of the AVID alignment **(Figure 2)** indicated that major part in predicted gene is highly conserved (more than 70%) in human HSP7C. Pair wise sequence alignment has also showed that the predicted protein is 77.3% identical and 86.7% similar with human HSP7C. The score obtained from such alignment is 2581 with 1.4% gaps. Global sequence alignment of human HSP7C and other related flowering plant proteins including Populus HSP70 was also performed and it showed strong full length sequence conservation. But sequence conservation result only for peptide binding domain was displayed here (**Figure 3**).

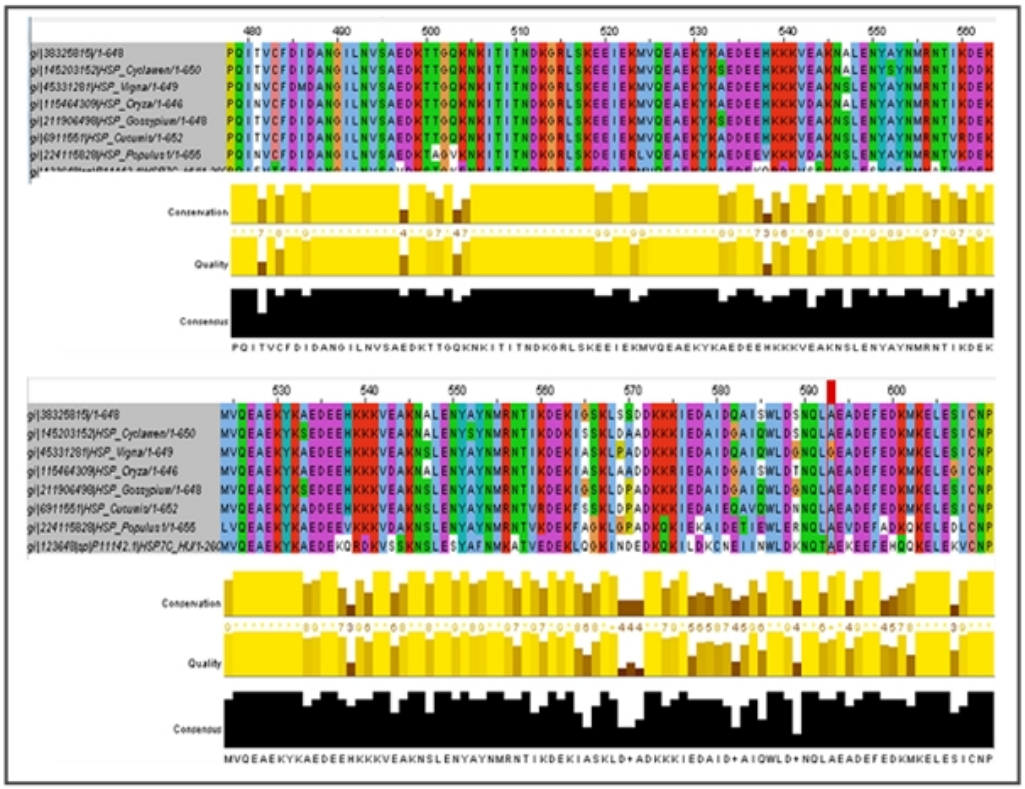

Figure 3. Global alignment of peptide binding domain of Human HSP7c protein with plant homologous.

3.4 Secondary structure study:

The Human HSP_7C includes two core functional domains: ATPase domain which located at N-terminal region and the peptide binding domain which consists of C-terminal region [18]. Helix and coil structure play strong role to form the architecture of these functional domains. It has already been established that structure determines corresponding function [19]. Here, secondary structure of this predicted protein was obtained by bioinformatics secondary structure prediction tool (HNN model) and successive comparative studies with Human_7C were carried on **(Figure 4)**. HNN model demonstrates that the hypothetical protein almost carries equal percentage of alpha helix, extended strands and random coils **(Table 4).**

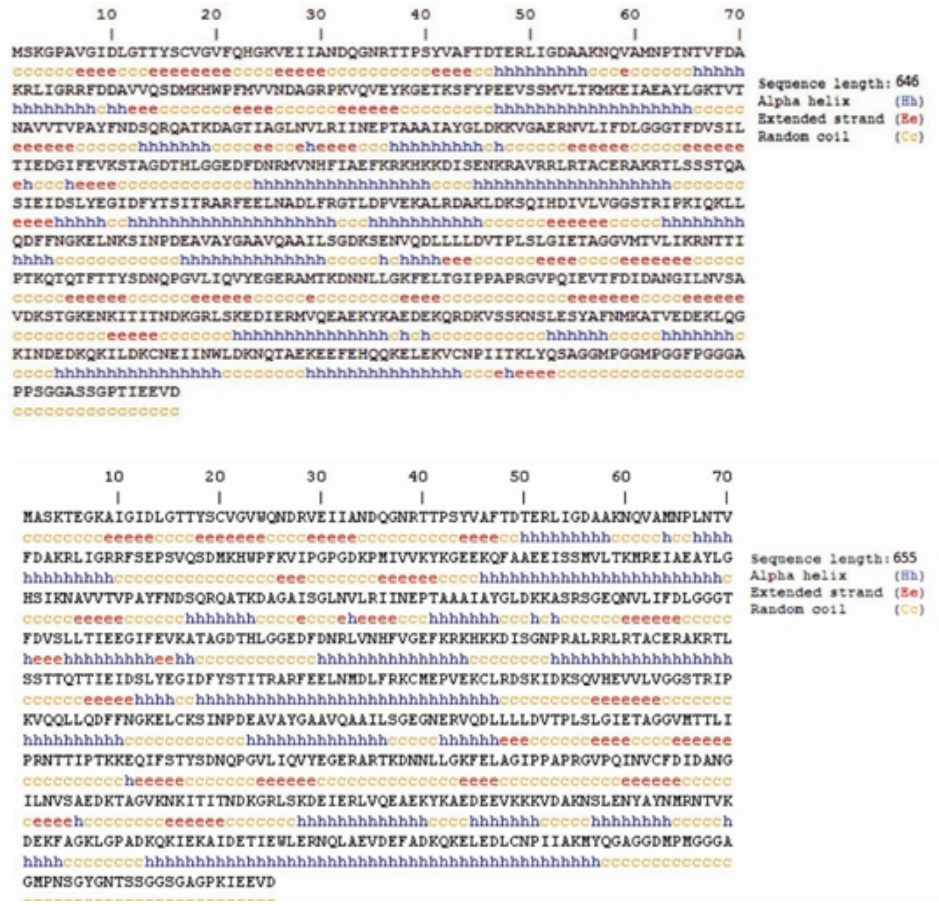

Figure 4. Secondary structure of human HSP7C and predicted populus1 protein.

Table 4: Secondary structure of human HSP7C and predicted populus1 protein in details.

Features	**HSP7C_Human**	**Predicted Protein**
Sequence length :	646	655
Alpha helix (Hh) :	228 is 35.29%	253 is 38.63%
3_{10} helix (Gg) :	0 is 0.00%	0 is 0.00%
Pi helix (Ii) :	0 is 0.00%	0 is 0.00%
Beta bridge (Bb) :	0 is 0.00%	0 is 0.00%

Extended strand (Ee) :	129 is 19.97%	108 is 16.49%
Beta turn (Tt) :	0 is 0.00%	0 is 0.00%
Bend region (Ss) :	0 is 0.00%	0 is 0.00%
Random coil (Cc) :	289 is 44.74%	294 is 44.89%
Ambigous states (?) :	0 is 0.00%	0 is 0.00%
Other states :	0 is 0.00%	0 is 0.00%

3.5 3D structure study

Three dimensional structure of predicted populus hsp70 and human HSP7c were constructed by using Swiss Pdb Viewer (SPDBV) **(Figure 5)**. 3D structures of these two proteins are quite similar and they have formed same functional clefts. Eletrostatic potential of their active sites are also alike that firmly indicates their equivalent function.

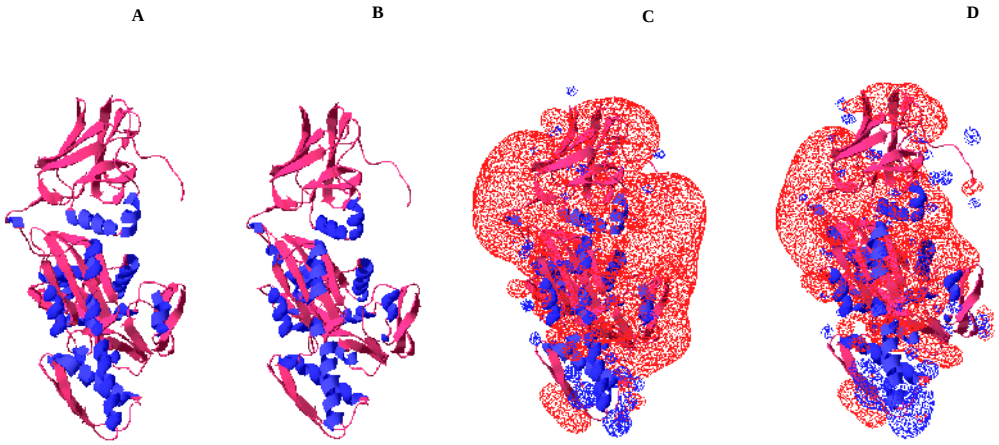

Figure 5: 3D structure and electrostatic potential of HSP7C_Human and Populus1 protein. A and B represent for 3D structure of human HSP7C and predicted populus1 protein respectively. C and D represent for electrostatic potential human HSP7C and predicted populus1 protein respectively.

Table 5. Hypothetical function of HSP7C_Human and HSP_Populus1.

HSP7C_Human		**HSP_POPULUS1(predicted protein)**	
Function	P Value (%)	Function	P Value (%)
All lipid-binding proteins	97.0	All lipid-binding proteins	93.6
All DNA-binding	91.3	All DNA-binding	94.7
Zinc-binding	76.2	Zinc-binding	82.2
TC 1.A. Channels/Pores - Alpha-Type channels	73.8	TC 1.A. Channels/Pores - Alpha-Type channels	97.5
TC 3.A. Primary Active Transporters - P-P-bond-hydrolysis-driven transporters	65.4	TC 3.A. Primary Active Transporters - P-P-bond-hydrolysis-driven transporters	58.6
Actin binding	58.6	Actin binding	62.2
Copper-binding	58.6	Copper-binding	58.6

3.6 Hypothetical protein function

To comprehend the functional relationship between these proteins, function of predicted protein was retrieved by SVMProt. The results have revealed that all of the functional domain of Human HSP_7C exist in the predicted populus1 protein and showed similar modular function with almost equal p value **(Table 5).**

4. CONCLUSION

Based on all the findings it can be concluded that the predicted protein may comprise human HSP_7C like function. Since plant HSP70 show higher sequence and structure homology with Human HSP70 family, it can further be concluded that the predicted protein may take part in protein folding, assembly, translocation and degradation in many normal cellular processes, stabilize proteins and membranes, and assist in protein refolding under stress conditions like plant hsp70[20].

Even though these hypothetical functions of this predicted protein (HSP_Populus1) should be further characterized by laboratory means before their existence can be conclusively affirmed, the results presented in this study suggested and identified structurally similar protein of heat shock protein family. The findings can provide new insight in plant protein characterization.

REFERENCES

[1] De Maio A (January 1999). "Heat shock proteins: facts, thoughts and dreams. Shock (Augusta, Ga) 11 (1): 1-12

[2] Crevel G, Bates H, Huikeshoven H, Cotterill S (1 June 2001). "The Drosophila Dpit47 protein is a nuclear Hsp90 co-chaperone that interacts with DNA polymerase alpha". *J. Cell. Sci.* **114** (Pt 11): 2015–25. PMID 11493638.

[3] Schlesinger, MJ (1990-07-25). "Heat shock proteins". *The Journal of Biological Chemistry* **265** (21): 12111–12114. PMID 2197269.

[4] Kregel, K.C., 2002. Heat shock proteins: modifying factors in physiological stress responses and acquired thermotolerance. J.Appl. Physiol. 92, 2177–2186.

[5] Schlesinger, M .J., 1990. Heat shock proteins. J. Biol. Chem. 265, 12111–12114.

[6] Hamilton, E.W., Heckathorn, S.A., Downs, C.A., Schwarz, T.E.,Coleman, J.S., Hallberg, R.L., 1996. Heat shock proteins are produced by field-grown naturally occurring plants in the summer in the temperate northeast US Bulletin of the Ecologic. Soc. Am.77, Suppl. Part 2: 180 (Abstr.)

[7] Vierling, E. (1991) The roles of heat shock proteins in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 42, 579–620

[8] Boston, R.S. et al. (1996) Molecular chaperones and protein folding inplants. Plant Mol. Biol. 32, 191–222

[9] Lin, B.L. et al. (2001) Genomic analysis of the Hsp70 superfamily in *Arabidopsis thaliana.* Cell Stress Chaperones 6, 201–208

[10] Sung, D.Y. et al. (2001) Comprehensive expression profile analysis ofthe Arabidopsis Hsp70 gene family. Plant Physiol. 126, 789–800

[11] Guy, C.L. and Li, Q.B. (1998) The organization and evolution of the spinach stress 70 molecular chaperone gene family. Plant Cell 10, 539–556

[12] Swindell, W.R., Huebner, M., Weber, A.P., 2007. Transcriptional profiling of Arabidopsis heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. BMC Genomics 8, 125 <http://www.biomed-central.com/1471-2164/8/125>

[13] Wangxia Wang1,3, Basia Vinocur1, Oded Shoseyov1,2 and Arie Altman1,2 Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. TRENDS in

Plant Science Vol.9 No.5 May 2004

[14] Al-Whaibi M.H (June 2010). Plant heat-shock proteins: A mini review. doi: 10.1016/j.jksus

[15] Vigh, L. et al. (2007) , Membrane Associated Stress Response:a Theoretical and Practical Approach

[16] Flaherty KM, DeLuca-Flaherty C, McKay DB (August 1990). "Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein". *Nature* **346** (6285): 623–8.

[17] Tavaria M, Gabriele T, Kola I, Anderson RL (April 1996). "A hitchhiker's guide to the human Hsp70 family". *Cell Stress Chaperones* 1 (1): 23–8.

[18] Takayama S., Bimston D.N., Matsuzawa S.-I."BAG-1 modulates the chaperone activity of Hsp70/Hsc70."**,** [PubMed: 9305631]

[19] Gabaldon, T; M.A. Huynen (2004). "Prediction of protein function and pathways in the genome era". Cellular and Molecular Life Sciences **61**: 930–944.

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

AUTHOR:

Animesh Sarker, Lecturer Department of Biotechnology and Genetic Engineering, Mawlana Bhashani Science and Technology University, Santosh,Tangail-1902.Bangladesh. Telephone:+088-092162407

Email: animesh_du_geb@yahoo.com