

Research Article

Computational Approach to Search for Plant Homologues of Human Heat Shock Protein

Animesh Sarker^{1*}, Marufa Nasreen¹, Md. Rafiad Islam¹, G M Mahmud Arif Pavel² and Al Amin¹

¹Department of Biotechnology and Genetic Engineering, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902, Bangladesh ²Department of Biotechnology and Genetic Engineering, St. John's University, Queens, New York, USA

Abstract

A number of homologous human heat shock proteins (HSPs) are available in plants. Human heat shock proteins (HSPs), which are expressed at higher temperature or other stress, have chaperone activity belong to five conserved families: HSP33, HSP60, HSP70, HSP90 and HSP100. A well known bioinformatics program BLASTp reveals that each of the human HSP families possesses a number of plant homologues except HSP33. Out of the rest four families, HSP70 carries best plant homologue. 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 between human HSP_7C and the designated plant protein. Secondary and three-dimensional (3D) structure analysis of the predicted plant protein strongly supports its functional relationship to the class of human HSP70.

Keywords: HSPs; Plant homologue; Comparative studies; Functional relationship

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 humans. 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 [1]. In plants, Hsps are greatly varied in level of expression as well as their type [5]. 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 small sHsps, but one single species could contain 40 types of these sHsps [6].

In plants, most common class protein, Hsp70 has been characterised in diverse species [6,7]. 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 [3,8,9]. At least 12 Hsp70 members have been found in the spinach genome [10]. 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 [10-13].

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 corelation 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 [14]. It seems that the synthesis of these proteins require enormous energy that impact on the yield of the organism [15].

Materials and Methods

Sequence analysis

The full-length protein sequences of all the human heat shock proteins (HSPs) families except HSP33 were retrieved from NCBI protein Database (http://www.ncbi.nlm.nih.gov/), with GI numbers, HSP60 (GI: 41399285), HSP70 (GI:123648), HSP90 (GI:92090606), HSP100 (GI:97536358). Detail protein information was retrieved from UniProt protein database (http://www.uniprot.org/) with accession number P10809(HSP60_Human), P11142(HSP7C_

*Corresponding author: Animesh Sarker, Department of Biotechnology and Genetic Engineering, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902, Bangladesh, Tel: +088-092162407; E-mail: animesh_du_geb@yahoo.com

Received April 04, 2013; Accepted May 16, 2013; Published May 20, 2013

Citation: Sarker A, Nasreen M, Islam MR, Mahmud Arif Pavel GM, Amin A (2013) Computational Approach to Search for Plant Homologues of Human Heat Shock Protein. J Comput Sci Syst Biol 6: 099-105. doi:10.4172/jcsb.1000106

Copyright: © 2013 Sarker A, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

HUMAN), P07900(HS90A_HUMAN), P07900 (HS90A_HUMAN), Q92598(HSP105_Human). 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.

Structure modeling

Secondary structures 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-MODEL is a structural bioinformatics web-server dedicated to homology modeling of protein 3D structures. Homology modeling is currently the most accurate method to generate reliable threedimensional protein structure models and is routinely used in many practical applications. Homology (or comparative) modelling methods make use of experimental protein structures ("templates") to build models for evolutionary related proteins ("targets").

Function prediction

To comprehend the functional relationship between these two proteins, SVMProt was used to predict their functions. One approach for function prediction is to classify a protein into functional family. Support vector machine (SVM) is a useful method for such classification, which may involve proteins with diverse sequence distribution. SVMProt can be accessed at http://jing.cz3.nus.edu.sg/ cgi-bin/svmprot.cgi.

Results and Discussion

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 (Table 2). Here, E-value means expect value that represents significance of sequence homology. It decreases exponentially with the Score (S) that is assigned to a match between two sequences. So, lower E-value is result for better match.

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 (Hsp90a and Hsp90 β , for instance) or conflicts of nomenclature (Hsp72 is sometimes called Hsp70) [16].

Since UniProt and NCBI database carries enormous sequence of human heat 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).

Position specific iterated BLAST scanning of each of the HSP candidate 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).

Phylogenetic study

According to BLAST result it is clear that two proteins [Populus trichocarpa gb|EEE71404.1|] and [Populus trichocarpa gb|EEE70430.1|] have 88% sequence homology and 79% and 78% sequence identity with HSP7C_HUMAN (Table 2). 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 multiple sequence alignment (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|EEE71404.1| (gi|224115828|Hsp_Populus1) (Figure 1).

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)

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

HSPs	Predicted Protein Description	E value	Similarity	Identity
HSP 105_Human	gb EFH64112.1 hypothetical protein ARALYDRAFT_896005 [<i>Arabidopsis</i> lyrata subsp. <i>lyrata</i>]	1e-165	59%	40%
	gb EES19148.1 hypothetical protein SORBIDRAFT_09g005570 [Sorghum bicolor]	4e-162	58%	39%
HS90A_HUMAN	gb EEE74469.1 predicted protein [<i>Populus trichocarpa</i>]	0	83%	71%
	gb EEE85774.1 predicted protein [<i>Populus trichocarpa</i>]	0	84%	71%
HSP7C_HUMAN	gb EEE71404.1 predicted protein [<i>Populus trichocarpa</i>]	0	88%	79%
	gb EEE70430.1 predicted protein [<i>Populus trichocarpa</i>]	0	88%	78%
HSP60_Human	gb EEE82364.1 predicted protein [<i>Populus trichocarpa</i>]	0	79%	60%
	gb EER94121.1 hy- potheticalprotein SORBIDRAFT_01g020010 [Sorghum bicolor]	0	79%	60%

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

Sequence Conservation study

Global alignment of the coding sequence of Human HSP7C and predicted populas_Hsp (gi|224115828|) protein gene sequence was performed with the program AVID using a window size of 100 bp and a conservation level of 70%. Basically AVID aligns DNA sequences of arbitrary length for the purpose of annotation and biological discovery using syntenic genomic sequences from two organisms. Results were viewed with the program VISTA (Figure 2). The VISTA plot is based on moving a user-specified window over the entire alignment and calculating the percent identity over the window at each base pair. From here it can be concluded that human HSP7C (gi|123648|) and populas_Hsp (gi|224115828|) are highly conserved with each other.

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

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	EEE71404.1	Populus trichocarpa	655	79	88	0
224113495	EEE70430.1	Populus trichocarpa	648	78	88	0

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





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).

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 [17]. 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 [18]. 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).

3D structure modelling

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.

Hypothetical protein function

Prediction of protein function is of significance in studying biological processes. 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). Here, P-Value is expected classification accuracy (probability of correct classification) in terms of percentage.

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 [19].

Even though these hypothetical functions of this predicted protein



Citation: Sarker A, Nasreen M, Islam MR, Mahmud Arif Pavel GM, Amin A (2013) Computational Approach to Search for Plant Homologues of Human Heat Shock Protein. J Comput Sci Syst Biol 6: 099-105. doi:10.4172/jcsb.1000106



Features	HSP7C_Human (no. of a.a and percentage)	Predicted Protein (no. of a.a and percentage)
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 (li) :	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%

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

Citation: Sarker A, Nasreen M, Islam MR, Mahmud Arif Pavel GM, Amin A (2013) Computational Approach to Search for Plant Homologues of Human Heat Shock Protein. J Comput Sci Syst Biol 6: 099-105. doi:10.4172/jcsb.1000106



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.

HSP7C_Human		Hsp_POPULUS1(predicted protein)		
Function	P Value (%)	Function	P value (%)	
All lipid-binding proteins All DNA-binding Metal-binding	97.0 91.3 80.4	All lipid-binding proteins All DNA-binding	93.6 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	
hydrolysis-driven transporters Actin binding	65.4	hydrolysis-driven transporters Actin binding		
Copper-binding		Copper-binding	58.6	
	58.6		62.2	
	58.6		58.6	

(*P-Value is the expected classification accuracy in terms of percentage) **Table 5:** Hypothetical function of HSP7C_Human and Hsp_Populus1.

(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 (1999) Heat shock proteins: facts, thoughts, and dreams. Shock 11: 1-12.
- Crevel G, Bates H, Huikeshoven H, Cotterill S (2001) The Drosophila Dpit47 protein is a nuclear Hsp90 co-chaperone that interacts with DNA polymerase alpha. J Cell Sci 114: 2015-2025.
- 3. Schlesinger MJ (1990) Heat shock proteins. J Biol Chem 265: 12111-12114.
- Kregel KC (2002) Heat shock proteins: modifying factors in physiological stress responses and acquired thermotolerance. J Appl Physiol 92: 2177-2186.
- Hamilton EW, Heckathorn SA, Downs CA, Schwarz TE, Coleman JS, ET AL. (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: 180
- 6. Vierling E (1991) The roles of heat shock proteins in plants. Ann Rev Plant Physiol Plant Mol Biol 42: 579-620.
- Boston RS, Viitanen PV, Vierling E (1996) Molecular chaperones and protein folding in plants. Plant Mol Biol 32: 191-222.

- Lin BL, Wang JS, Liu HC, Chen RW, Meyer Y, et al. (2001) Genomic analysis of the Hsp70 superfamily in Arabidopsis thaliana. Cell Stress Chaperones 6: 201-208.
- Sung DY, Vierling E, Guy CL (2001) Comprehensive expression profile analysis of the Arabidopsis Hsp70 gene family. Plant Physiol 126: 789-800.
- Guy CL, Li QB (1998) The organization and evolution of the spinach stress 70 molecular chaperone gene family. Plant Cell 10: 539-556.
- Swindell WR, Huebner M, Weber AP (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.
- Wang W, Vinocur B, Shoseyov O, Altman A (2004) Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends Plant Sci 9: 244-252.
- 13. Al-Whaibi MH (2010) Plant heat-shock proteins: A mini review. J Jksus.
- Vígh L, Török Z, Balogh G, Glatz A, Piotto S, et al. (2007) Membrane-regulated stress response: a theoretical and practical approach. Adv Exp Med Biol 594: 114-131.
- Flaherty KM, DeLuca-Flaherty C, McKay DB (1990) Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. Nature 346: 623-628.
- 16. Tavaria M, Gabriele T, Kola I, Anderson RL (1996) A hitchhiker's guide to the human Hsp70 family. Cell Stress Chaperones 1: 23-28.

Citation: Sarker A, Nasreen M, Islam MR, Mahmud Arif Pavel GM, Amin A (2013) Computational Approach to Search for Plant Homologues of Human Heat Shock Protein. J Comput Sci Syst Biol 6: 099-105. doi:10.4172/jcsb.1000106

 Takayama S, Bimston DN, Matsuzawa S, Freeman BC, Aime-Sempe C, et al. (1997) BAG-1 modulates the chaperone activity of Hsp70/Hsc70. EMBO J 16: 4887-4896. the genome era. Cell Mol Life Sci 61: 930-944.

- 19. Lindquist S (1986) The heat-shock response. Annu Rev Biochem 55: 1151-1191.
- 18. Gabaldón T, Huynen MA (2004) Prediction of protein function and pathways in