





Comparative *In Silico* Analysis of Randomly Selected Heat Shock Proteins in *Caenorhabditis elegans* and *Photorhabdus temperata*

Urvashi Dhiman ¹, Ripu Daman Parihar ^{2,*}, Shivani Rana ¹, Sushil Kumar Upadhyay ³, Vinod Kumar ⁴

¹ Department of Zoology, DAV University, Jalandhar, India

² Department of Zoology, Central University of Jammu, Jammu & Kashmir, India

³ Department of Biotechnology, Maharishi Markandeshwar (Deemed to be University), Mullana-Ambala-133207 (Haryana), India

⁴ Department of Botany, Government Degree College Ramban, Jammu & Kashmir, India

* Correspondence: ripuparihar89@gmail.com;

Scopus Author ID 56692536400

Received: 20.11.2020; Revised: 6.12.2020; Accepted: 9.12.2020; Published: 11.12.2020

Abstract: Heat shock proteins (HSPs) such as HSP70A, HSP90 etc. (also known as Chaperons) play an important role in folding and unfolding of proteins, an assemblage of multiprotein complexes, transportation and sorting of proteins in subcellular compartments, cell cycle control, signaling pathways, protection against stress and programmed cell death. Studies have also linked heat shock proteins with a sudden rise in temperature, which can be related to anhydrobiosis in nematodes. Considering the significance of HSPs in nematodes and bacteria, the present study was designed for their *in silico* analysis in *Caenorhabditis elegans* and *Photorhabdus temperata*. The availability of a vast amount of sequence data generated through various bioinformatics tools, coupled with computational biology advancements, provides an ideal framework for *in silico* gene expression and its analysis. A detailed *in silico* insight into these proteins include physicochemical properties, secondary structure prediction, homology modeling, and different models. The amino acid composition data were subjected to multivariate techniques, Pearson correlation, and phylogenetic analysis. In the present study, the authors characterized different HSPs according to different stability parameters and valid structures. A detailed *in silico* analysis of these proteins and prediction of their activity in different conditions can be very useful in both *in vitro* and *in vivo* experiments.

Keywords: heat shock proteins; physicochemical properties; secondary structure prediction; homology modeling; homology validation.

Abbreviations: HSPs: Heat Shock Proteins, NCBI: National Centre for Biotechnology Information, ExPasy: Expert Protein Analysis System, SOPMA: Self-Optimized Prediction Method with Alignment, I-TASSER: Iterative Threading Assembly Refinement, RAMPAGE: RNA Annotation and Mapping of Promoters for the Analysis of Gene Expression, MEGA: Molecular Evolutionary Genetic Analysis, nsSNPs: Non-Synonymous Single Nucleotide Polymorphisms, MTHFR: Methylene tetrahydrofolate Reductase, BCL11A: B-cell Lymphoma/leukemia 11A, SMPX: Small Muscle Protein X-Linked, NMR: Nuclear Magnetic Resonance.

© 2020 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

All organisms undergo tremendous changes, either physiological or environmental, till life comes to an end. Every organism develops different strategies to survive and withstand

harsh conditions to maintain homeostasis [1]. Changes with which these organisms have to deal are environmental (atmospheric conditions, toxicity, and water scarcity) and physical (thermal sensitivity, thermal tolerance limits, and hypoxia conditions). Numerous invertebrate species developed various mechanisms to cope with severe stress conditions such as heat stress, water stress, food stress, etc. [2-5]. *Distolabrellus veechi* KPI, a pathogenic nematode, showed differential habitat preference in the river catchment area, wetland, cropland, playground, roadside, and railway track, indicating distinct survival strategies [6,7,8]. Their high reproductive rate, transparent body, and easy invitro culture make them a reliable tool to study fluorescent reactions with chemical compounds and require more insights [9,10,11]. Nematode and bacteria evolve over time and develop a strategy for their survival in response to varying conditions. The biological phenomenon is called anhydrobiosis [12-13]. In this process, an organism completely undergoes an ametabolic state (hypometabolic state). The organism survives indefinitely till the power to perform its basic metabolic activities is resumed [14-15]. In other words, it is a transition state between life and dormancy. Anhydrobiotic organisms possess the ability to survive extreme dehydration up to the level of 99% [16-17]. Heat Shock Protein (HSP) plays a diverse role in keeping the organism alive, even in 99% of the dehydrated state. Different groups of HSPs have been reported to be upregulated or downregulated at different stages of anhydrobiosis, probably reflecting their different functionality [2, 18-19]. HSPs are an assorted superfamily of conserved proteins that are found in all living beings in differing blends. They vary in atomic size and activity method, for the most part falling into two significant classifications: ATP-dependent large HSP and ATP-independent small HSP (sHSP). In turn, the large Hsp can be divided into four ubiquitous conserved families: HSP100s, HSP90s, HSP70s, and HSP60s. The number of large Hsp family names is around a demonstrative molecular mass, as the small HSPs are from 12 to 42 kDa in size [20]. Large HSPs can have both stress-inducible and constitutively expressed variants. At the same time, sHSP are frequently expressed uniquely because of the only stress response [21]. Hsp was first discovered as key defensive components of the heat shock response but were shown subsequently to participate in a wide range of cellular processes relating to proteome maintenance, a part of both stress responses and normal physiology [22]. Nowadays, they are often called molecular chaperones. Their main activities include enabling the correct folding of newly synthesized proteins, the refolding of denatured proteins, protein oligomers' assembly, protein trafficking, and assistance in protein degradation [22]. Recently heat shock proteins are also recognized for regulating immune responses. HSPs bind non-covalently to immunogenic peptides, induce dendritic cell maturation, releases pro-inflammatory cytokine, and activates Natural Killer cells (NK cells) [23-24]. Bat longevity and stress resilience have also been linked with the expression of HSPs [25-28].

With the growing advancement in science and technology, *in silico* studies address many questions and have emerged as a reliable tool for various studies [29-30]. *In silico* studies involving protein functions and structure have been used to uncover mutations, make predictions for classifying deleterious Non-Synonymous Single Nucleotide Polymorphisms (nsSNPs) in MTHFR (Methylenetetrahydrofolate Reductase), BCL11A (B-cell Lymphoma/leukemia 11A), SMPX (Small Muscle Protein X-Linked) gene related to protein-protein interactions [31-37]. *In silico* studies have also been reported to reveal B and T cell epitopes against fatal strains of the Aphthovirus serotypes and to explore strictosidine synthase homology modeling in alkaloid biosynthesis [38-39]. In the present study, the authors tried to characterize various physical and chemical properties of HSPs by using different freely

accessible bioinformatics tools. HSPs are found in both *Caenorhabditis elegans* and *Photobacterium aerophilum*. They might exhibit a different phenomenon in different conditions involving similar proteins. So insight knowledge of their physiochemical parameters, structure, and homology modeling can prove useful in studying biological pathways involved in this life-saving process, i.e., anhydrobiosis.

2. Materials and Methods

2.1. Collection of data / Retrieval of amino acid sequences.

The National Centre for Biotechnology Information (NCBI) is one of the world's leading networks for computational and biomedical research and was used to retrieve amino acid sequences or FASTA files for our work (<https://www.ncbi.nlm.nih.gov/>). A total of nine proteins and their FASTA format sequences were fetched from two different organisms, i.e., *C. elegans* and *P. temperata* (Table 1).

2.2. Primary structure prediction and physico-chemical characterization.

Prediction of primary structure and computation of various physical and chemical properties of protein sequences was made by using web tool ProtParam ExPasy (Expert Protein Analysis System). The computed parameters include the molecular weight, theoretical pI, amino acid composition, i.e., number of positively and negatively charged amino acids, instability index etc. A comparative study of physical and chemical parameters is important in determining the role of protein and its molecular evolution [40].

Table 1. Various proteins with their accession numbers.

PROTEIN NAME	GENE ID /ACCESSION NUMBER
HSP 110 (<i>C. elegans</i>)	CCD66159
HSP (<i>C.elegans</i>)	CCD64545
HSP16.2 (<i>C.elegans</i>)	AAA28071
HSP90 (<i>C.elegans</i>)	CAA99793
HSP 70A (<i>C.elegans</i>)	AAA28078
HSPQ (<i>P.temperata</i>)	ERT11781.1
HtpX (<i>P.temperata</i>)	ERT14044
Metal Binding HSP (<i>P.temperata</i>)	ERT13418
Ribosome Associated HSP 15 (<i>P.temperata</i>)	ERT14175

2.3. Secondary structure prediction.

Self-Optimized Prediction Method with Alignment (SOPMA) was used to predict secondary features of proteins. This tool evaluates the percentage of alpha-helix, beta-sheets, turns, random coils, extended strands etc. [41].

2.4. 3D structure prediction.

Iterative Threading Assembly Refinement (I-TASSER) is the widely used bioinformatics tool for predicting the three-dimensional structure model of protein molecules from amino acid sequences. I-TASSER is regarded as one of the best tools for structure prediction of proteins at a computerized level [42]. It is freely accessible and generates five full-length models [43].

2.5. Validation of generated models.

All the five models generated via I-TASSER were validated using different computational methods such as 3D-VERIFY, ERRAT, RAMPAGE, and PROCHECK.

2.5.1. 3D-Verify.

It determines the compatibility of an atomic 3D model (<http://servicesn.mbi.ucla.edu/Verify3D/>) with its amino acid sequence (1D) by assigning a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar etc.) and compares the results to ideal structures [44-45].

2.5.2. Errat.

It analyzes the statistics of non-bonded interactions between different atom types and plots the value of the error function versus position of a 9-residue sliding window, calculated by comparison with statistics from highly refined structures (<http://servicesn.mbi.ucla.edu/ERRAT/>) [46].

2.5.3. Rampage.

RNA Annotation and Mapping of Promoters for the Analysis of Gene Expression (RAMPAGE) is used for the assessment of Ramachandran plot (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) [47].

2.5.4. Procheck.

It checks the stereochemical quality of a protein structure, producing several PostScript plots analyzing its overall and residue-by-residue geometry (<http://www.ebi.ac.uk/thorntonsrv/databases/pdbsum/Generate.html>). It includes Procheck-NMR for checking the quality of structures solved by nuclear magnetic resonance (NMR). Results of procheck include plots and detailed information about a number of residues present in the favored region and the outlier region [48].

2.6. Phylogenetic analysis.

Molecular Evolutionary Genetic Analysis (MEGA) is used for analyzing genomic data and constructing evolutionary relationships [49].

2.7. Statistical analysis.

Multivariate techniques and Pearson correlation were used to analyze the correlation between different amino acids in the proteins.

3. Results and Discussion

Prediction of the physicochemical properties of heat shock proteins was made using the Protparam tool from ExPasy (Tables 2-5). These analyzed properties of protein play a crucial role in determining the sustainability and stability of a protein in a biological system [17]. Among heat shock proteins of *C. elegans* HSP 110 showed the highest molecular weight, i.e., 86896.55 KDa, and HSP16.2 showed the lowest molecular weight, i.e., 16242.9 KDa (Table

3). On the other hand, among Heat shock proteins in *Photorhabdus temperata*, HtPX showed the highest molecular weight, i.e., 32242.52 KDa, and HSPQ showed the lowest molecular weight, i.e., 11916.24 KDa (Table5).

Table 2. Amino Acid (%) of Heat Shock Protein in *Caenorhabditis elegans*.

Proteins	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
HSP110	8.4	5.4	4.5	6.6	0.9	4.3	8.6	5.2	1.4	5.9	6.3	7.2	2.1	4	6.2	5.9	4.6	0.8	2.8	8.9
HSP	3.3	12.5	6.9	9.2	0.3	3.3	5.3	4.8	2.8	4.1	4.6	3.3	2.5	4.1	5.9	10.7	5.9	2.8	3.3	4.6
HSP90	6.3	3.7	4.3	7.1	1.1	2.7	12.3	4.3	1.1	6.4	7.8	11	2.8	4.3	2.6	7	4.3	0.6	3.3	7.1
HSP70A	6.3	3.7	4.3	7.1	1.1	2.7	12.3	4.3	1.1	6.4	7.8	11	2.7	4.3	2.6	7	4.3	0.6	3.3	7.1
HSP 16.2	6.9	5.5	4.1	6.2	0.7	6.2	9	4.8	1.4	6.9	8.3	7.9	2.8	4.1	4.8	11	1.4	0	2.1	6.9

A=Alanine, R=Arginine, N=Asparagine, D=Aspartic Acid, C=Cysteine, Q= Glutamine, E=Glutamic Acid, G= Glycine, H=Histidine, I=Isoleucine, L=Leucine, K=Lysine, M=Methionine, F= Phenylalanine, P=Proline, S=Serine, T=Threonine, W=Tryptophan, Y=Tyrosine, V=Valine

Table 3. Physico-chemical properties of different Heat Shock Proteins in *Caenorhabditis elegans*.

PROTEIN	MW	pI	NO. OF AA	+ve AA.	-ve AA.	Instability index
HSP110	86896.55	5.3	776	98	118	46.21
HSP	46242.86	9.17	393	62	57	54.98
HSP 16.2	16242.9	5.25	145	18	22	53.57
HSP 90	80283.23	4.96	702	103	136	37.39
HSP 70A	80152.03	4.96	701	103	136	36.95

Table 4. Amino Acid (%) of Heat Shock Protein in *Photorhabdus temperata*.

PROTEIN	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
HSPQ	9.4	4.7	2.8	9.4	0	7.5	8.5	4.7	3.8	5.7	12.3	1.9	1.9	0.9	5.7	7.5	0.9	0.9	4.7	6.6
HtPX	9.8	4.7	5.1	1.7	0.3	4.1	6.8	8.1	1.7	8.1	10.5	2.7	5.8	5.1	2.4	8.5	4.1	1	1.7	7.8
MB HSP	5.2	2.6	2.6	7.1	1.3	5.8	15.5	3.9	3.9	5.2	12.3	3.9	2.6	2.6	5.2	5.8	3.9	1.3	2.6	7.1
RA HSP15	8.7	11.6	5.1	5.1	0	4.3	9.4	4.3	2.2	5.1	6.5	10.1	3.6	1.4	4.3	5.8	4.3	1.4	2.2	4.3

A=Alanine, R=Arginine, N=Asparagine, D=Aspartic Acid, C=Cysteine, Q= Glutamine, E=Glutamic Acid, G= Glycine, H=Histidine, I=Isoleucine, L=Leucine, K=Lysine, M=Methionine, F= Phenylalanine, P=Proline, S=Serine, T=Threonine, W=Tryptophan, Y=Tyrosine, V=Valine, X=Unspecified Amino Acid

Studies have linked high molecular weight with an amino acid composition of proteins, i.e., high molecular weight of a protein commends a high percentage of amino acid in their respective protein [50]. This supports our study in which we also observed that high molecular weighted HSP110 and HtPX have the highest number of amino acids, i.e., 776 and 295, respectively (Tables 2,4). Instability index is an intrinsic property of proteins that helps determine the *in vivo* stability [51]. Protein whose instability index value is less than 40 is regarded as stable, whereas a value more than 40 accounts for the instability of the proteins. In our study, only HSP90 and HSP70A in the case of *C.elegans* were found to be stable, whereas in *P. temperata* all the proteins were unstable proteins according to the value of their instability index. Isoelectric point (pI) is the pH at which a molecule carries no net electrical charge. At pI, proteins are generally present in the compact form [52]. If the pI value is less than 7, it indicates an acidic nature of the protein. Similarly, a pI value of more than 7 infers the basic nature of the protein. According to the computed values of pI from ProtParam tool HSP 90, HSP70A, HSP110, HSP16.2 (*C. elegans*), HSPQ, HtPX, Metal Binding HSP (*P. temperata*) were found to have acidic nature (pI<7) whereas HSP (*C.elegans*), Ribosome Associated HSP15 were basic proteins (pI>7) (Tables 3,5). This isoelectric point plays a significant role in the purification process. It marks the pH where the solubility of protein is typically minimal. Studies have linked the activity of various Heat Shock Proteins to their pH levels. Tiwari and his coworkers found the predominant activity of HSPs in acidic pH. The activity level goes down in an alkaline pH [53]. It has been reported that the expression of HSPs is maximum in the presence of a higher concentration of glutamine, and expression becomes minimum on a

low level of glutamine [54]. Precisely, glutamine or any of its metabolites positively affects the signaling pathway that controls the expression of heat shock proteins [54]. In our study, we found that HSP70A has the highest content of an acidic amino acid, i.e., Glutamic acid (Table 2).

Table 5. Physicochemical properties of different Heat Shock Proteins of *Phototribdus temperata*.

PROTEIN	MW	pI	Number of AA	+ve A.A.	-ve A.A.	Instability Index
HSPQ	11916.24	4.39	106	7	19	46.20
HtPX	32242.58	5.99	295	22	25	46.34
MB HSP	32242.58	4.25	155	10	35	57.67
RA HSP15	16073.34	10.03	138	30	20	52.61

Table 6. Calculated secondary structure percentages by SOPMA of HSPs of *Caenorhabditis elegans*.

Protein	Alpha helix	310 helix	Pi helix	Beta bridge	Extended strand	Beta turn	Bend region	Random coil	Ambiguous states	Other
HSP110	46.39%	0.00%	0.00%	0.00%	14.56%	3.48%	0.00%	35.57%	0.00%	0.00%
HSP	17.80%	0.00%	0.00%	0.00%	16.79%	7.38%	0.00%	58.02%	0.00%	0.00%
HSP90	53.56%	0.00%	0.00%	0.00%	13.68%	4.84%	0.00%	27.92%	0.00%	0.00%
HSP70A	42.34%	0.00%	0.00%	0.00%	18.59%	7.97%	0.00%	31.09%	0.00%	0.00%
HSP16.2	35.17%	0.00%	0.00%	0.00%	18.62%	5.52%	0.00%	40.69%	0.00%	0.00%

3.1. Secondary structure prediction.

The secondary structure of proteins was predicted using SOPMA (Self- optimized prediction method with alignment) tool [43]. The tool evaluates the percentage of Alpha helix, Beta extended strands, Random coils like parameters by using Homology methodology [55]. Among HSPs in *C. elegans*; HSP 90, HSP 110, HSP70A showed high percentage of Alpha helix i.e. 53.56%, 46.39% and 42.34% respectively (Table 6). Whereas among HSPs in *P. temperata*; HtPX, Metal Binding HSP, Ribosome Associated HSP 15 showed a high percentage of Alpha helix, i.e., 59.32%, 46.45%, and 42.75%, respectively (Table 7). To study the percentage of various amino acids in the protein, we used ProtParam, which disclosed that HSP 90, HSP70A, HtPX, Metal Binding HSP, and Ribosome Associated HSP15 possess a high concentration of Glutamic acid, Leucine, and Arginine in their polypeptide chains. The amino acid percentage has a significant role in forming the helical structure of the protein. Studies have revealed that amino acids Methionine, Alanine, Leucine, Glutamate, and Lysine have high helix forming tendency, whereas Proline and Glycine have poor [56]. Leucine and Arginine have also been acknowledged as strong helix forming components due to their frequent repetition in the polypeptide chain [57]. These studies also support our results where we observed dominance of the above-mentioned amino acids with respect to good alpha-helical structure in a protein. Over a period of time, this tool has been used invariably for secondary structure predictions of proteins [58].

Table 7. Calculated secondary structure percentages by SOPMA of HSPs of *Phototribdus temperata*.

Protein	Alpha helix	310 helix	Pi helix	Beta bridge	Extended strand	Beta turn	Bend region	Random coil	Ambiguous states	Other
HSPQ	33.02%	0.00%	0.00%	0.00%	19.81%	7.55%	0.00%	39.62%	0.00%	0.00%
HTPX	59.32%	0.00%	0.00%	0.00%	12.54%	3.73%	0.00%	24.41%	0.00%	0.00%
MB HSP	46.45%	0.00%	0.00%	0.00%	15.48%	3.87%	0.00%	34.19%	0.00%	0.00%
RS HSP15	42.75%	0.00%	0.00%	0.00%	12.32%	5.80%	0.00%	39.13%	0.00%	0.00%

3.2. Homology modeling and validation.

In the present study, homology modeling was performed by using I-TASSER. For all nine Heat Shock Proteins, five 3D models were generated. The quality of these models was validated by using different bioinformatic tools (Tables 8,9). The bioinformatics tool called

ERRAT or “Overall Quality Factor” validated models by statistical relation of non-bonded interactions among different atom types on their characteristic atomic interactions [44][38]. This tool assesses the overall quality of the model in the form of the Quality Factor, which is used to determine the reliability of the generated model [59]. The higher the value of the Quality Factor, high will be the quality of the generated model [60]. For a high-quality model, the generally accepted range is greater than 50. In case of *C.elegans*, errat scores of the best model were 89.48%, 84.37%, 88.97%, 92.55% and 93.02% for HSP110, HSP, HSP16.2, HSP90 and HSP70A respectively (Table 8). In case of *P. temperata*, errat scores were 96.93%, 91.63%, 91.72% and 96.15% for HSPQ, HtPX, MB HSP and RA HSP15 respectively (Table 9).

Table 8. Validation Scores of different 3D models of HSPs in *Caenorhabditis elegans* generated via I-TASSER.

PROTEIN	3DVERIFY (%)	ERRAT (%)	PROCHECK (%)	RAMPAGE (%)
HSP110	79.90	89.48	74.3	80.20
HSP	63.36	84.37	63.6	74.40
HSP16.2	88.97	88.97	76.7	88.80
HSP90	81.48	92.55	77.08	83.60
HSP70A	94.63	93.02	82.6	89.30

According to the concept, all the models of HSPs (*C.elegans* and *P. temperata*) showed their quality factor above 50; hence all the generated models are good and can be used for future proteomic studies. The 3D-Verify score is used to evaluate the protein structure by comparing its structural environments with amino acids' preferred environment [47]. In case of *C.elegans*, 3D verify scores of the best model were 79.9%, 63.36%, 88.97%, 81.48% and 94.63% for HSP110, HSP, HSP16.2, HSP90 and HSP70A respectively (Table 8). In case of *P. temperata*, 3D scores were 91.51%, 51.19%, 92.19% and 78.26% for HSPQ, HtPX, MB HSP and RA HSP15 respectively (Table 9).

Table 9. Validation Scores of different 3D models of *Photorhabdus temperata* generated via I-TASSER.

PROTEIN	3D VERIFY (%)	ERRAT (%)	PROCHECK (%)	RAMPAGE (%)
HSPQ	91.51	96.93	81.7	93.3
HtPX	51.19	91.63	83.6	89.10
MB HSP	92.19	91.72	82.7	90.20
RA HSP15	78.26	96.15	81.5	86.00

RAMPAGE another 3D validation model whose results show the number of residues in favored regions, allowed, and outlier regions. PROCHECK checked the stereochemical quality of protein structure by analyzing residue-by-residue geometry and overall structure geometry via Ramachandran plot. The plot has different regions – most favored regions, residues in allowed and disallowed regions. According to PROCHECK calculations, good Quality models should possess over 90% of amino acids in most favored regions. In the case of *C. elegans*, the highest procheck scores (number of residues in the favored region of the best model) was 74.3%, 63.6%, 76.7%, 77.08%, and 82.6% for HSP110, HSP, HSP16.2, HSP90, and HSP70A, respectively (Table 8). In the case of *P. temperata*, the highest procheck scores were 81.7%, 83.6%, 82.7%, and 81.5% for HSPQ, HtPX, MB HSP, and RA HSP15, respectively (Table 9). In the case of *C. elegans*, rampage scores (number of residues in the favored region of the best model) were 80.20%, 74.40%, 88.8%, 83.6%, and 89.3% for HSP110, HSP, HSP16.2, HSP90, and HSP70A, respectively (Table 8). In case of *P. temperata*, rampage scores were 93.3%, 89.10%, 90.20% and 86.00% for HSPQ, HtPX, MB HSP and RA HSP15 respectively (Table 9). A higher value of these models by using all validity tools indicates the high quality of all models.

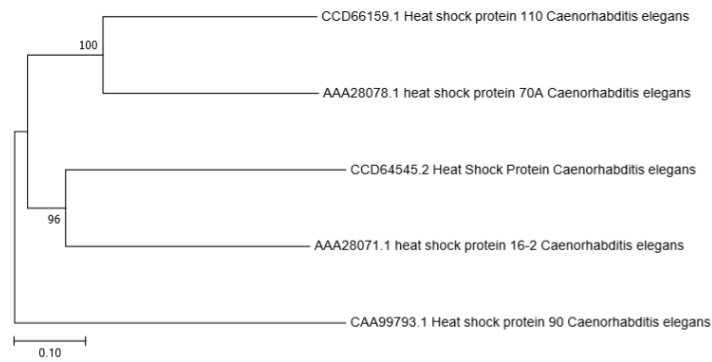


Figure 1. Phylogenetic analysis of Heat Shock Proteins of *Caenorhabditis elegans*.

3.2.1. Phylogenetic analysis.

To study the Heat Shock Proteins' evolutionary relationship, the phylogenetic tree construction was done using MEGA 7 software with other selected HSPs sequences retrieved from NCBI. The results reveal intra evolutionary relationships between different Heat Shock Proteins of *C. elegans* and *P. temperata*. For *C. elegans* Phylogenetic tree reveals that out of all five Heat shock proteins, HSP90 evolved earlier. The remaining four heat shock proteins evolved from HSP90. HSP110 and HSP 70A shares a common ancestor, which is HSP90. HSP and HSP16.2 also share a common ancestor, which earlier got diverged from HSP90. HSP 110 and HSP70A might have evolved together as their divergence branch is the same (Fig. 1). For *P. temperata* all four Heat Shock proteins evolved simultaneously. Along with this one, HSPs also get evolved, which further got diverged into heat shock protein HtPX and HSPQ (Fig. 2).

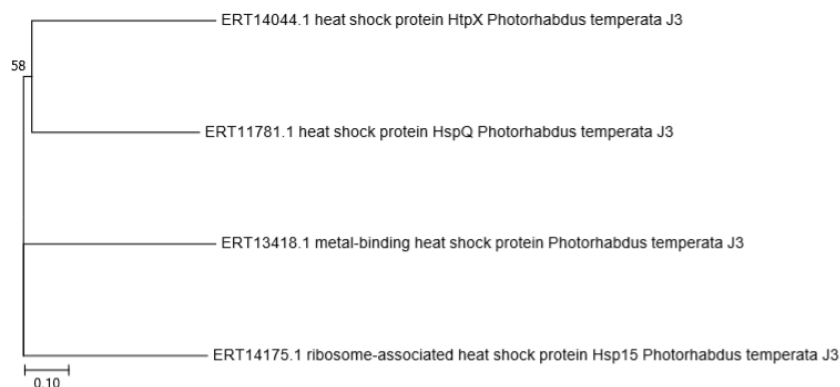


Figure 2. Phylogenetic analysis of Heat Shock Proteins of *Photobacterium aerophilum*.

3.2.2. Pearson correlation.

Pearson's correlation analysis was applied to different amino acids by using R-programming software. In the case of *C. elegans* statistically, Aspartic acid and Alanine, Cysteine, Glutamic Acid, Isoleucine, Leucine, Lysine, and Arginine, all are negatively correlated to each other, whereas Tryptophan is positively correlated to Arginine, Asparagine, and Aspartic acid (Fig. 3). In the case of *P. temperata* statistically, Isoleucine and Glycine, Proline and Aspartic acid, Tyrosine and Glutamine are positively correlated with each other, whereas Glutamic Acid and Alanine, Histidine and Asparagine, Leucine and Arginine, Methionine and Aspartic acid are negatively correlated to each other (Fig. 4).

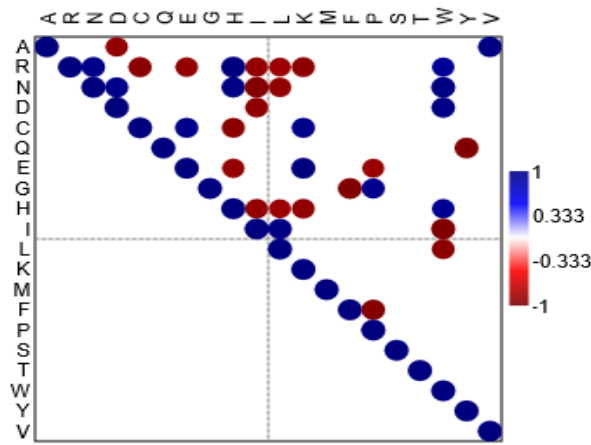


Figure 3. Pearson correlation of different HSPs in *Caenorhabditis elegans*.

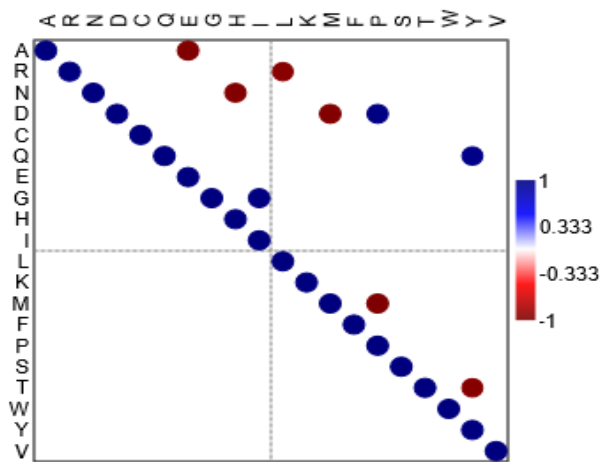


Figure 4. Pearson correlation of different HSPs in *Photothabdus temperata*.

3.2.3. Multivariate technique.

The Heat Map was conducted to different amino acids and Heat Shock Proteins. Results showed that in *C. elegans* HSP 90, HSP70A belongs to one group, and HSP 16.2, HSP 110 belongs to another group. The composition of cysteine is the same in HSP, HSP110, and HSP16.2. The composition of tryptophan is the same in the case of HSP110, HSP16.2, HSP70A, and HSP90. The composition of leucine is the same in the case of HSP110, HSP 16.2, HSP70A, and HSP90.

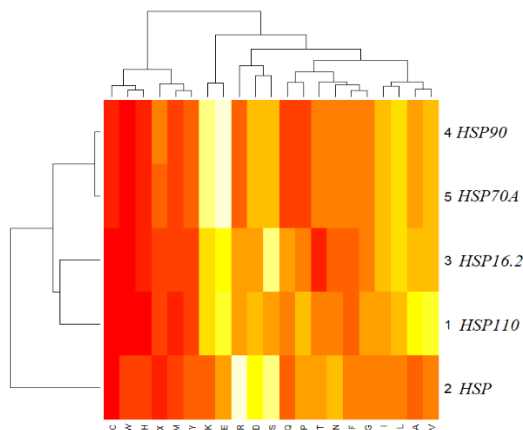


Figure 5. Heat maps showing the amino acid composition of different HSP in *Caenorhabditis elegans*.

The composition of histidine, glycine, isoleucine, leucine, and valine is the same in HSP 90, HSP70A, and HSP16.2. However, HSP composition of 20 amino acids is different in all Heat Shock Proteins of *C. elegans* (Fig. 5). In *P. temperata* Ribosome Associated HSP15, HtPX belongs to one group, whereas Metal Binding HSP, HSPQ belongs to the other group. The composition of cysteine, tryptophan, tyrosine, and histidine is the same in Heat Shock Proteins HtPX and Ribosome Associated HSP15. The valine composition is the same in HSPQ, Metal Binding HSP, and HtPX (Fig. 6).

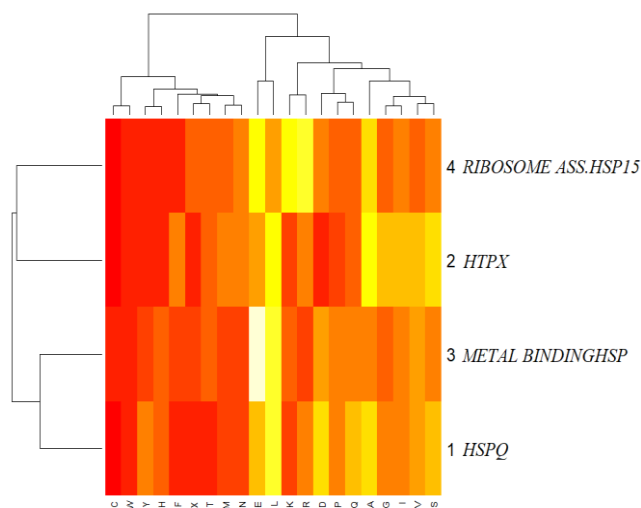


Figure 6. Heat maps showing the amino acid composition of different HSP in *Photorhabdus temperata*.

4. Conclusions

Our study has presented a compressive *in silico* assessment and structure prediction of Heat Shock Proteins (HSPs). Out of the nine HSPs, only two, i.e., HSP90 and HSP70 were found to be stable following the instability index. Classification of acidic and basic proteins based on the isoelectric point was done. Secondary structure prediction revealed the dominance of alpha-helical structure in these proteins. For all nine HSPs, five 3D models were generated, and the most reliable predicted structure was validated through validity scores, i.e., 3D VERIFY (%), ERRAT (%), PROCHECK (%), and RAMPAGE (%). This data can be very useful for the study of the structure of these proteins by NMR and X-ray crystallography as well as their industrial utility. Our study also provides insights into the functional analysis of these proteins, which will enable researchers to design *in vivo* assays.

Funding

This research received no external funding.

Acknowledgments

The authors are thankful to the Head, Department of Zoology, DAV University Jalandhar, and Department of Biotechnology, Maharishi Markandeshwar (Deemed to be University) for permission of interuniversity collaborative research work and incessant moral support.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Storey, K.B.; Storey, J.M. Heat shock proteins and hypometabolism: Adaptive strategy for proteome preservation. *Res. Rep. Biol.* **2011**, *2*, 57-68, <https://doi.org/10.2147/RRB.S13351>.
2. Malhotra, A.; Jaiswal, N.; Upadhyay, S.K.; Kumar, S.; Malhotra, S.K. Parasite stress response in zoonoses of *Cephalogonimus* sp. *Indian J. Helminthol.* **2009**, *27*, 73-80.
3. Jaiswal, N.; Upadhyay, S.K.; Malhotra, A.; Malhotra, S.K. Multifactorial etiology of infections by larvae of *Eustrongylides tubifex* (Nematoda: Dioctophymidae) in silver whiting of the central west coast of India at Goa. *Asian J. Biol. Sci.* **2013**, *6*, 21-39, <https://doi.org/10.3923/ajbs.2013.21.39>.
4. Upadhyay, S.K.; Jaiswal, N.; Malhotra, A.; Malhotra, S.K. Ecological morphotaxometry of trematodes of garfish (Teleostomi: Belonidae) from Gangetic riverine ecosystem in India. II. Correlation of seasonality and host biology with distribution pattern of *Cephalogonimus yamunii* n.sp. *J. Parasit. Dis.* **2013**, *37*, 211-217, <https://doi.org/10.1007/s12639-012-0168-2>.
5. Jaiswal, N.; Upadhyay, S.K.; Malhotra, A.; Malhotra, S.K. Ecological morphotaxometry of trematodes of garfish (Teleostomi: Belonidae) from Gangetic riverine ecosystem in India. III. Principal component analysis for hydrobiological correlates to dynamics of infections by *Cephalogonimus yamunii* (Upadhyay, Jaiswal, Malhotra and Malhotra, 2012). *J. Parasit. Dis.* **2014**, *38*, 153-162, <https://doi.org/10.1007/s12639-012-0200-6>.
6. Parihar, R.D.; Singh, I.; Sohal, S.K.; Kesavan, A.K.; Ohri, P. Molecular characterization and study of abiotic soil parameters to understand the natural habitat preference of entomopathogenic nematode *Distolabrellus veechi* isolates. *Indian J. Nematol.* **2019**, *49*, 146-155.
7. Parihar, R.D.; Singh, I.; Verma, V.; Sohal, S.K.; Kesavan, A.K.; Ohri, P. Biocontrol potential of *Distolabrellus veechi* isolate KPI (Nemata: Rhabditida) against cotton cutworm, *Spodoptera litura* (Lepidoptera: Noctuidae). *Indian J. Nematol.* **2019**, *49*, 71-82.
8. Parihar, R.D.; Kaur, S.S.; Ohri, P. Morphometric variations in *Distolabrellus veechi* (Nematoda: Rhabditidae) collected from different localities in Punjab, India. *Indian J. Nematol.* **2016**, *46*, 193-197.
9. Reja, S.I.; Sharma, N.; Gupta, M.; Bajaj, P.; Bhalla, V.; Parihar, R.D.; Kumar, M. A highly selective fluorescent probe for detection of hydrogen sulfide in living systems: *In vitro* and *in vivo* applications. *Chem.–A European J.* **2017**, *23*, 9872-9878, <https://doi.org/10.1002/chem.201701124>.
10. Gupta, N.; Kaur, T.; Bhalla, V.; Parihar, R.D.; Ohri, P.; Kaur, G.; Kumar, M. A naphthalimide-based solid state luminescent probe for ratiometric detection of aluminum ions: *In vitro* and *in vivo* applications. *Chem. Comm.* **2017**, *53*, 12646-12649, <https://doi.org/10.1039/C7CC07996F>.
11. Sharma, P.; Gupta, N.; Kaur, S.; Kaur, S.; Ohri, P.; Parihar, R.D.; Kumar, M. Imaging of lysosomal activity using naphthalimide-benzimidazole based fluorescent probe in living cells. *Sensors and Actuators B: Chemical* **2019**, *286*, 451-459, <https://doi.org/10.1016/j.snb.2019.01.134>.
12. Upadhyay, S.K. Allelopathic activities of specific microbial metabolites in the inland prawn fisheries off eastern Uttar Pradesh, India. *Int. J. Scient. Res.* **2016**, *5*, 415-416, <https://doi.org/10.15373/22778179>.
13. Upadhyay, S.K. Activity patterns of cell free supernatant of antagonistic microbial strains in rodents host-parasite systems. *Int. J. Sci. Res.* **2016**, *5*, 332-336.
14. Crowe, J.H.; Hoekstra, F.A.; Crowe, L.M. Anhydrobiosis. *Ann. Rev. Physiol.* **1992**, *54*, 579-599, <https://doi.org/10.1146/annurev.ph.54.030192.003051>.
15. Danks, H.V. Dehydration in dormant insects. *J. Insect Physiol.* **2000**, *46*, 837-852, [https://doi.org/10.1016/S0022-1910\(99\)00204-8](https://doi.org/10.1016/S0022-1910(99)00204-8).
16. Upadhyay, S.K. Anthelmintic and food supplementary conscientiousness of apitoxin in poultry model. *Res. J. Rec. Sci.* **2016**, *5*, 09-14.
17. Kumar S.; Upadhyay, S.K. Pathogenesis of *Flavobacterium colunare* in fish of fresh water riverine ecosystem from eastern region of Uttar Pradesh, India. *Int. J. Rec. Scient. Res.* **2016**, *7*, 13676-13679.
18. Hayward, S.A.L.; Rinehart, J.P.; Denlinger, D.L. Desiccation, and rehydration elicit distinct heat shock protein transcript responses in flesh fly pupae. *J. Exp. Biol.* **2004**, *207*, 963-971, <https://doi.org/10.1242/jeb.00842>.
19. Wang, C.; Grohme, M.A.; Mali, B.; Schill, R.O.; Frohme, M. Towards decrypting cryptobiosis – analyzing anhydrobiosis in the tardigrade *Milnesium tardigradum* using transcriptome sequencing. *PLoS One* **2014**, *9*, 1-15, <https://doi.org/10.1371/journal.pone.0092663>.
20. Haslbeck, M.; Vierling, E. A first line of stress defense: small heat shock proteins and their function in protein homeostasis. *J. Mol. Biol.* **2015**, *427*, 1537-1548, <https://doi.org/10.1016/j.jmb.2015.02.002>.
21. Richter, K.; Haslbeck, M.; Buchner, J. The heat shock response: life on the verge of death. *Mol. Cell.* **2010**, *40*, 253-266, <https://doi.org/10.1016/j.molcel.2010.10.006>.
22. Hartl, F.U.; Bracher, A.; Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature.* **2011**, *475*, 324-332, <https://doi.org/10.1038/nature10317>.
23. Ostberg, J.R.; Kablingu, E.; Repasky, E.A. Thermal regulation of dendritic cell activation and migration from skin explants. *Int. J. Hypertherm.* **2003**, *19*, 520-533, <https://doi.org/10.1080/02656730310001607986>.

24. Menoret, A.; Chaillot, D.; Callahan, M.; Jacquin, C. Hsp70, an immunological actor playing with the intracellular self under oxidative stress. *Int. J. Hypertherm.* **2002**, *18*, 490-505, <https://doi.org/10.1080/02656730210146926>.
25. Ball, H.C.; Levari-Shariati, S.; Cooper, L.N.; Aliani, M. Comparative metabolomics of aging in a long-lived bat: Insights into the physiology of extreme longevity. *PLoS One.* **2018**, *13*, <https://doi.org/10.1371/journal.pone.0196154>.
26. Hughes, G.M.; Leech, J.; Puechmaile, S.J.; Lopez, J.V.; Teeling, E.C. Is there a link between aging and microbiome diversity in exceptional mammalian longevity? *Peer. J.* **2018**, *6*, <https://doi.org/10.7717/peerj.4174>.
27. Wilkinson, G.S.; Adams, D.M. Recurrent evolution of extreme longevity in bats. *Biol. Lett.* **2019**, *15*, <https://doi.org/10.1098/rsbl.2018.0860>.
28. Chionh, Y.T.; Cui, J.; Koh, J.; Mendenhall, I.H.; Ng, J.H.J.; Low, D.; Itahana, K.; Irving, A.T. Wang, L.F. High basal heat-shock protein expression in bats confers resistance to cellular heat/oxidative stress. *Cell Stress Chap.* **2019**, *24*, 835–849, <https://doi.org/10.1007/s12192-019-01013-y>.
29. Upadhyay, S.K. Transmission dynamics and environmental influence on food borne parasitic helminthes of the Gangetic plains and central west coast of India. *Unpubl. D.Phil. Thesis. Univ. Allahabad* **2012**, 1-400.
30. Upadhyay, S.K. Morphotaxometry and molecular heterogeneity of *Sturdynema multiembryonata* gen. et sp.n. (Spiruroidea: Gnathostomatinae) of fresh water garfish, *Xenentodon cancilla* from the Gangetic riverine ecosystem in northern India with a revised key to genera of Gnathostomatinae. *Species* **2017**, *18*, 1-13.
31. Agrahari, A.K.; George, P.D.C.; Siva, R.; Magesh, R.; Zayed, H. Molecular insights of the G2019S substitution in LRRK2 kinase domain associated with Parkinson’s disease: A molecular dynamics simulation approach. *J. Theoret. Biol.* **2019**, *469*, 163–171, <https://doi.org/10.1016/j.jtbi.2019.03.003>.
32. Agrahari, A.K., Priya, M.K.; Kumar, M.P.; Tayubi, I.A.; Siva, R.; Christopher, B.P.; Doss, C.G.P.; Zayed, H. Understanding the structure–function relationship of HPRT1 missense mutations in association with Lesch–Nyhan disease and HPRT1-related gout by in silico mutational analysis. *Comput. Biol. Med.* **2019**, *107*, 161–171, <https://doi.org/10.1016/j.compbiomed.2019.02.014>.
33. Dash, R.; Junaid, M.; Mitra, S.; Arifuzzaman, M.; Hosen, S.Z. Structure-based identification of potent VEGFR-2 inhibitors from in vivo metabolites of a herbal ingredient. *J. Mol. Model.* **2019**, *25*, <https://doi.org/10.1007/s00894-019-3979-6>.
34. Karimian, M.; Hosseinzadeh C.A. Human MTHFRG1793A transition may be a protective mutation against male infertility: A genetic association study and in silico analysis. *Human Fertility Camb.* **2018**, *21*, 128–136, <https://doi.org/10.1080/14647273.2017.1298161>.
35. Abdulazeez, S.; Sultana, S.; Almandil, N.B.; Almohazey, D.; Bency, B.J.; Borgio, J.F. The rs61742690 (S783N) single nucleotide polymorphism is a suitable target for disrupting BCL11A-mediated foetal-to-adult globin switching. *PloS One.* **2019**, *14*, <https://doi.org/10.1371/journal.pone.0212492>.
36. Arifuzzaman, M.; Mitra, S.; Jahan, S.I.; Jakaria, M.; Abeda, T.; Absar, N.; Dash, R. A computational workflow for the identification of the potent inhibitor of type II secretion system traffic ATPase of *Pseudomonas aeruginosa*. *Comput. Biol. Chem.* **2018**, *76*, 191–201, <https://doi.org/10.1016/j.compbiolchem.2018.07.012>.
37. Arshad, M.; Bhatti, A.; John, P. Identification and in silico analysis of functional SNPs of human TAGAP protein: A comprehensive study. *PloS One.* **2018**, *13*, <https://doi.org/10.1371/journal.pone.0188143>
38. Sohail, R.; Kalsoom S.; Masood, R.; Tahir, Y.; Aftab, A.A.; Muhammad, I.; Muhammad, A.; Farrukh, J.; Muhammad, A.R. In silico analysis of four structural proteins of aphthovirus serotypes revealed significant B and T cell epitopes. *Microb. Pathogen.* **2019**, *128*, 254–262, <https://doi.org/10.1016/j.micpath.2019.01.007>.
39. Sahay, A.; Piprodhe, A.; Pise, M. In silico analysis and homology modeling of strictosidine synthase involved in alkaloid biosynthesis in *Catharanthus roseus*. *J. Genet. Eng. Biotechnol.* **2020**, *18*, <https://doi.org/10.1186/s43141-020-00049-3>.
40. Garg, V.K.; Avashthi, H.; Tiwari, A.; Jain, P.A.; Ramkete, P.W.; Kayastha, A.M.; Singh, V.K. MFPPi–Multi FASTA ProtParam interface. *Bioinformatics.* **2016**, *12*, <https://doi.org/10.6026/97320630012074>.
41. Geourjon, C.; Deleage, G. SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics.* **1995**, *11*, 681-684, <https://doi.org/10.1093/bioinformatics/11.6.681>.
42. Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocol.* **2010**, *5*, <https://doi.org/10.1038/nprot.2010.5>.
43. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinfo.* **2008**, *9*, <https://doi.org/10.1186/1471-2105-9-40>.
44. Bowie, J.U.; Luthy, R.; Eisenberg, D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **1991**, *253*, 164-170, <https://doi.org/10.1126/science.1853201>.
45. Lüthy, R.; Bowie, J.U.; Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* **1992**, *356*, 83-85, <https://doi.org/10.1038/356083a0>.
46. Colovos, C.; Yeates, T.O. Verification of protein structures: Patterns of non-bonded atomic interactions. *Protein Sci.* **1993**, *2*, 1511-1519.

47. Lovell, S.C.; Davis, I.W.; Arendall, W.B.; de Bakker, P.I.W.; Word, J.M.; Prisant, M.G.; Richardson, J.S.; Richardson, D.C. Structure validation by Calpha geometry: Phi, psi and beta deviation. *Proteins Struct. Funct. Genet.* **2002**, *50*, 437-450.
48. Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystal.* **1993**, *26*, 283-29, <https://doi.org/10.1107/S0021889892009944>.
49. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870-1874, <https://doi.org/10.1093/molbev/msw054>.
50. Guan, Y.; Zhu, Q.; Huang, D.; Zhao, S.; Lo, L.J.; Peng, J. An equation to estimate the difference between theoretically predicted and SDS PAGE-displayed molecular weights for an acidic peptide. *Scient. Rep.* **2015**, *5*, <https://doi.org/10.1038/srep13370>.
51. Guruprasad, K.; Reddy, B.B.; Pandit, M.W. Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng. Design Select.* **1990**, *4*, 155-161, <https://doi.org/10.1093/protein/4.2.155>.
52. Sivakumar, K.; Balaji, S. *In silico* characterization of antifreeze proteins using computational tools and servers. *J. Chem. Sci.* **2007**, *119*, 571-579, <https://doi.org/10.1007/s12039-007-0072-y>.
53. Tiwari, S.; Thakur, R.; Shankar, J. Role of heat-shock proteins in cellular function and in the biology of fungi. *Biotechnol. Res. Int.* **2015**.
54. Sanders, M.M.; Kon, C. Glutamine is a powerful effector of heat shock protein expression in *Drosophila* Kc cells. *J. Cell. Physiol.* **1991**, *146*, 180-190.
55. Singh, N.; Upadhyay, S.; Jaiswar, A. Mishra, N. *In silico* analysis of protein. *J Bioinform Genom. Proteom.* **2016**, *1*.
56. Pace, C.N.; Scholtz, J.M. A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.* **1998**, *75*, 422-427, [https://doi.org/10.1016/S0006-3495\(98\)77529-0](https://doi.org/10.1016/S0006-3495(98)77529-0).
57. Chou, P.Y.; Fasman, G.D. Empirical predictions of protein conformation. *Ann. Rev. Biochem.* **1978**, *47*, 251-276, <https://doi.org/10.1146/annurev.bi.47.070178.001343>.
58. Pradeep, N.V.; Anupama, A.; Vidyashree, K.G.; Lakshmi, P. *In silico* characterization of industrial important cellulases using computational tools. *Adv. Life Sci. Technol.* **2012**, *4*, 2224-7181.
59. Pandey, V.P.; Singh, S.; Jaiswal, N.; Awasthi, M.; Pandey, B.; Dwivedi, U.N. Papaya fruit ripening: ROS metabolism, gene cloning, characterization and molecular docking of peroxidase. *J. Mol. Catalys. Enzym.* **2013**, *98*, 98-105, <https://doi.org/10.1016/j.molcatb.2013.10.005>.
60. Messaoudi, A.; Belguith, H.; Hamida, J. B. . Homology modeling and virtual screening approaches to identify potent inhibitors of VEB-1 β -lactamase. *Theoret. Biol. Med. Model.* **2013**, *10*, <https://doi.org/10.1186/1742-4682-10-22>.