

# Homology Modeling and Functional Annotation of Heat Shock Protein (HSP) Gene from *Agave Americana*

Anicet Batcho\*, Bilal Sarwa and Bushra Rachid

Plant Genomic Lab, Centre of Excellence in Molecular Biology, University of the Punjab, Lahore  
Email: [batcho.anicet@cemb.edu.pk](mailto:batcho.anicet@cemb.edu.pk)

**Abstract.** Identifying active domain in a protein sequence is one of the most important functions in molecular biology. This is usually studied with three-dimensional (3D) structure of the active sites protein using Bioinformatics tools. In the absence of an experimentally determined structure, homology modeling can sometimes supply a useful 3D model for protein that is linked to at least one known protein structure. In this work, we have evaluated the sequence information of Heat Shock Protein (HSP) gene retrieved from *Agave americana*. Homology modeling, functional annotation of *A. americana* heat shock protein (HSP) sequence. Based on homology modeling, 3D structure of the gene was constructed. Validation tests were used to calculate and determine the reliability of the structure. In this study, details structure of HSP led to develop a better understanding of the HSP's role in different abiotic stress response pathways.

**Keywords:** Functional annotation, bioinformatics tools, HSP gene, *Agave americana*.

## 1 Introduction

*Agave americana* is one of Agavaceae species found in deserts. It is well adapted to arid and semiarid regions [1]. As most succulent plants, agaves have undergone morphological and physiological adaptations to survive under adverse environmental conditions. Survival of agaves in such regions is often related to their fleshy massiveness, or succulence. The such water capability occurs on a cellular level and is also related to the photosynthetic pathway used by most of other related plant species, known as Crassulacean acid metabolism (CAM) [2]. Plants exhibiting CAM open their stomata and have a net uptake of CO<sub>2</sub> mainly at night, when the lower tissue temperature and higher ambient relative humidity lead to less water than daytime stomata opening. CAM can be crucial for growth and survival in arid habitats and has evolved many times in diverse groups of related plant. Agavaceae have gained attention around the world both as its potential lignocellulosic bioenergy feed stocks and models for exploring gene response to abiotic stress.

Plants growth and developmental processes are substantially affected by high temperature (HT) [3, 4], which causes multifarious, and often adverse, alterations in plant growth, development, physiological processes, and yield [5, 6]. A combination of approaches based on the recent advances in genomic research has been formulated to address these challenges by identifying important heat stress responsive genes [7, 8]. Recent studies at molecular level have provide ways to develop stress tolerant varieties and to grow agriculturally important crop plants under heat stress.

Some specific genes were identified in response to stress. Their stress adaptation provides a solid foundation for generating effective strategies to better stress tolerance in plants [9]. Heat shock proteins (HSPs) are found in most prokaryotes and eukaryotes, and even some viruses [10]. They are said to be isolated in many plant species, such as *Arabidopsis*, soybean, tobacco, rice, maize and wheat [11]. These proteins enable the cell to survive and recover from stressful conditions [12, 13]. Functional annotation of three-dimensional (3D) structure has big potential to examine the hereditary pattern of genes at the molecular level [14]. The combination of genetic mapping techniques with Bioinformatics tools will facilitate the detailed molecular dissection of HSP from *A. americana* and the results will therefore be valuable to study its relative species.

## 2 Materials and Methods of Construction

### 2.1 Nucleotide Sequencing and Analysis

The sequence of HSP gene of *Agave americana*, was retrieved from the NCBI database using the primary accession name JQ671429. Approximately 238 amino acid sequences were used for sequence analysis.

### 2.2 Homology Modeling of HSP Gene

The protein sequence of *A. americana* HSP gene (NCBI GenBank accession number JQ671429.1) was submitted to SWISS-Model server (<https://swissmodel.expasy.org/interactive/nYMCSb/templates/>) for homology modeling of the HSP proteins. A total of 319 templates were found to match the target sequence. This list was filtered by a heuristic down to 50. Five models (5DS2.1.A, 2BYU.1.A, 2BYU.1.B, 1GME.2.A and 1GME.2.B) have been taken as the guide based on their similarity and identity to the target sequence. Validation of 3D structure was done by plotting Ramachandran plot [15] (<http://services.mbi.ucla.edu/SAVES/Ramachandran>) and calculation assessed by a versatile protein structure analysis PROCHECK [16].

### 2.3 Functional Annotation

The ProSA-web server [17] (<https://prosa.services.came.sbg.ac.at/prosa.php>) was utilized to find the energetic architecture of the protein fold. The multiplication of energy graphs and calculation of z-score were done by uploading the PDB format target and template file. The PROCHECK was used to assess the overall stereo-chemical character of the model construction. A Ramachandran plot was generated for the model using PROCHECK and was viewed using a ghost script viewer for quality assessment of 3D models. The protein model structure was compared to differentiate the patterns of distribution of Alpha helices, Beta plated sheets and interconnecting loops.

### 2.4 Identification of Active Site

It is implicitly assumed that having the protein's structure will provide insights into its function. However, knowing a protein's three-dimensional structure is insufficient to determine its function [18]. To analyze and predict the multifunctional aspects of proteins to recognize active sites and binding regions in these protein structures, HSP gene was examined for the presence or absence of conserved domains using SMART (Simple Modular Architecture Research Tool) (<http://smart.embl-heidelberg.de>). SMART allows the identification and annotation of genetically mobile domains and the analysis of domain architectures.

## 3 Results

The current work focused to describe homology modeling, functional annotation of HSP gene in *Agave americana* using different Bioinformatics approaches.

### 3.1 Homology Modeling

Homology detection by iterative (HHblits) [19] for evolutionary and a three-dimensional structure of HSP was predicted (<https://swissmodel.expasy.org/interactive>) (Figure 2) to study the localization and interaction of proteins in stable confirmation. Among the 50 templates, 5SD2.1.A showed the highest similarity with 39% and 41.94% identities (Table 1). Moreover, the results were further analyzed by verifying 3D. The model quality shows 90% (Figure 3) which further confirms the good quality of the generated model. This predicted information will help in better understanding of mechanisms underlying to heat tolerance, and the use this information in protein engineering will help to improve plants to high temperature conditions.

### 3.2 Functional Annotation

The reliability of 3D was confirmed using Ramachandran plot. The point at which charge on protein become zero is called an isoelectric point (PI). The ratio of PI/Mw was estimated by PROCHECK, to assess the overall stereo-chemical quality of the 3D model. The result showed in terms of Z-score -5.19 indicated the overall quality of protein structure (Figure 1B). In this plot, groups of structures of different sources (X-ray, NMR) were marked by different colors. The value -5.19 lies in the range of native conformation. Generally the positive values corresponded to those parts of the inputs which were problematic or erroneous. The residue energy was negative without points, so this implies that no membrane spanning regions were present. Ramachandran plot was drawn between the Phi and Psi torsion angles and generated four coordinated colours. The red coloured region, represented the most favoured combination of torsion value and the generously regions were in yellow (Figure 3).

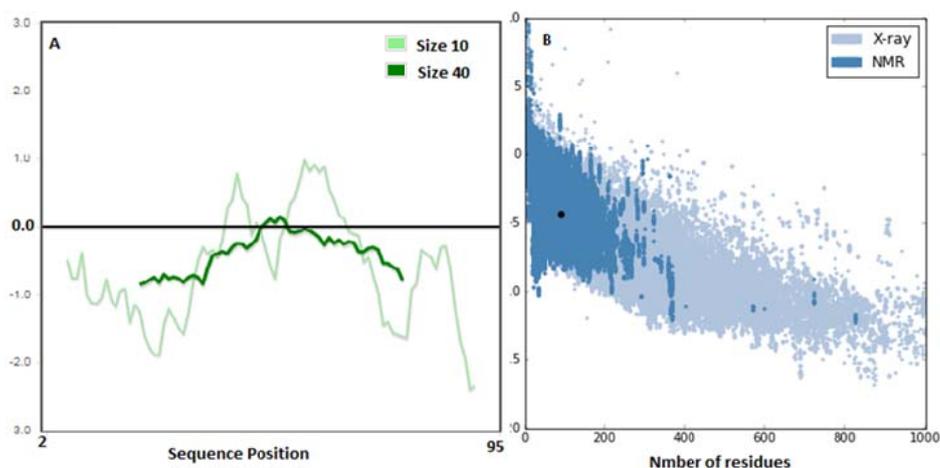
### 3.3 Active-site Identification

The active sites in proteins are better conserved than the overall fold. This suggests that one could not only identify distant ancestors with the same global fold and the same activity but also proteins with similar functions but distantly related, or possibly unrelated, global folds [18]. Figure 4 represents a summary of active site, within four domains. Orbits with a score less significant than established cutoffs are not pictured in the figure. The conserved domain of the annotated HSP protein sequence starts at 133 and ends at 238 with E-value =  $1.3e30$  while, its homologue D1GM sequence in Structural Classification of Protein Database (SCOP\_D1GM), starts at 90 and ends at 235 with E-value =  $5e37$ . The LALIGN pair wise (<http://www.ebi.ac.uk/Tools/>) of both sequence showed a total of 145 amino acid overlapped between two amino acid sequences. The alignment analysis further showed 100% identity and 100% similarity between Protein family database (Pfam\_HSP20 and SCOP\_D1GM).

**Table 1.** Templates sequences used for the 3D model building

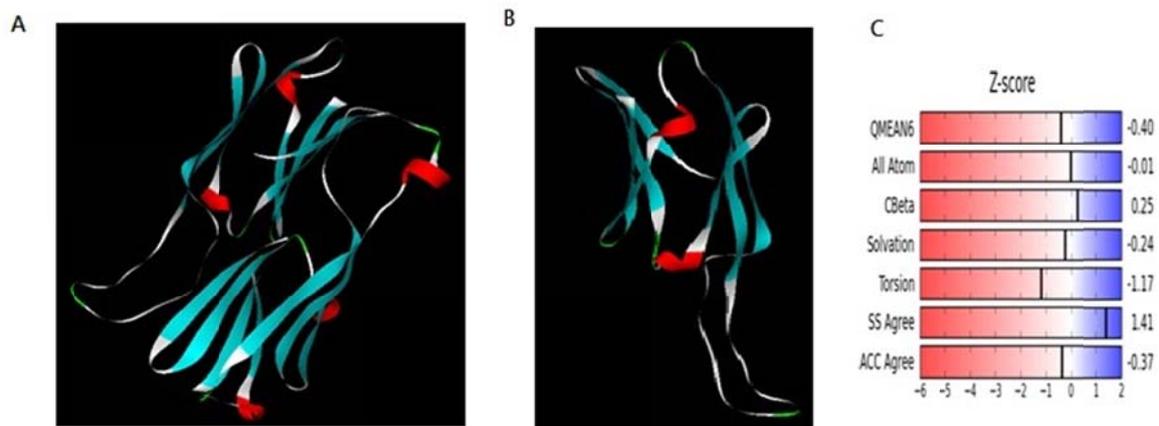
Template	Seq Identity	Oligo-state	Found by	Method	Resolution	Seq similarity	Coverage	Description
1GME.2.A	25.71	homo-12-mer	HHblits	X-ray	2.70Å	0.32	0.15	HSP 16.9B
1GME.2.B	25.71	homo-12-mer	HHblits	X-ray	2.70Å	0.32	0.15	HSP 16.9B
2BYU.1.A	37.50	homo-12-mer	HHblits	EM	16.50Å	0.37	0.44	HSP 16.9B
2BYU.1.B	37.50	homo-12-mer	HHblits	EM	16.50Å	0.37	0.44	HSP 16.9B
5SD2.1.A	41.94	homo-dimer	HHblits	X-ray	1.85Å	0.39	0.39	18.1 kDa class I heat shock protein

\*Seq: sequence



**Figure 1.** Predicted local similarity of HSP sequence, (A) Local quality estimate (B) Z-Score plot. The z-score indicates overall model quality. Its value is displayed in a plot that contains the z-scores of all experimentally

determined protein chains. In this plot, groups of structures from different sources (X-ray, NMR) are marked by different colors.



**Figure 2.** Homology model of *A. americana* HSP gene, residues are colored from blue to red in the order of increasing residue energy (A) Predicted template-target 3D structure (B) Predicted target 3D structure (C) Statistical angles between different domain

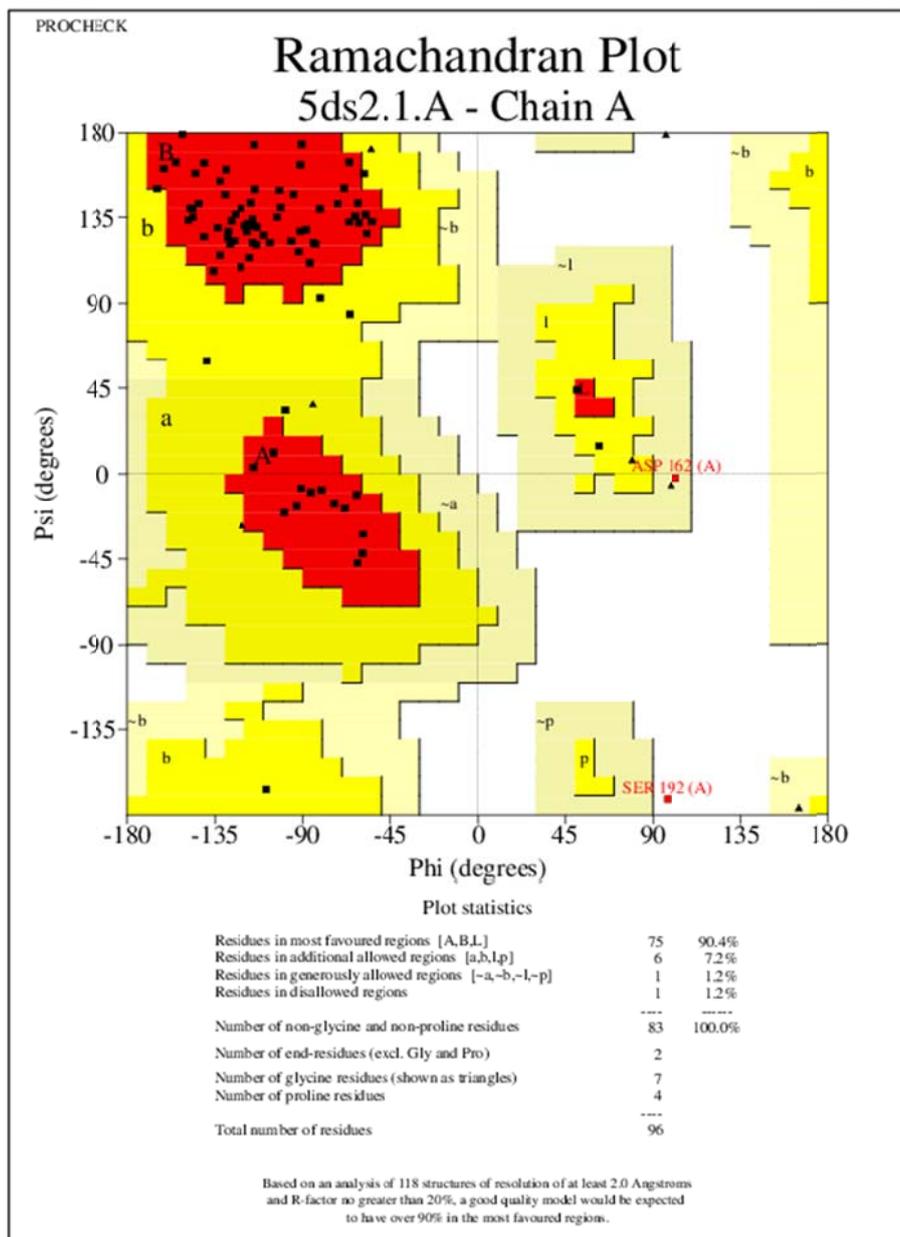


Figure 3. Ramadachan plot of HSP gene for reliability of 3D structure

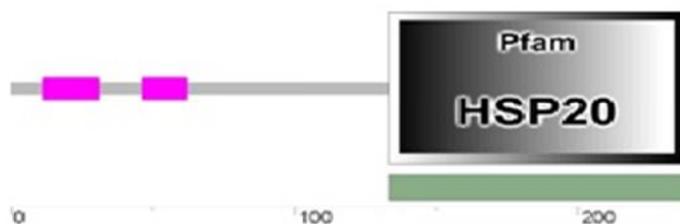


Figure 4. Domains within *A. americana* protein J7H8J9\_AGAAM (J7H8J9) Chloroplast small heat shock protein

## 4 Discussion

This research examined the homology between known HSP genes and HSP gene from *A. americana*, the result shows a Core domain of the class I small heat-shock protein HSP 18.1 from *Pisum sativum*, which is considerably more divergent in structure and function than the HSP70s. The sHSPs are found in bacteria, archaea and eukaryotes. They range in monomer size between 12 and 43kDa, and are characterized by a conserved  $\alpha$ -crystallin domain of about 80 residue [20]. The characteristic functions of this class have the ability to suppress the in vitro aggregation of denaturing protein, while in vivo their expression protects cells during stress [21]. The induction of transcription of this protein is a common phenomenon in all living organisms. It is believed that in plant this protein reflects an adaptation to tolerate stress [22]. Heat shock proteins have some kind of related roles in regulating a range of effect or components, all of which contribute to survival under abiotic stress by solving the problem of misfolding an aggregation, as well as its role as chaperones [23]. Finally, it is important to study stress combinations to end up with tolerant plant.

## 5 Conclusion

Protein structure determination of challenging molecular assemblies requires integration of different data type obtained by multiple methods. This research is based on functional annotation of heat shock protein from *Agave americana* revealed four active site. This result will certainly help to identify small heat shock protein from the relative species and enhance the heat stress tolerance in economical important plants.

**Acknowledgement.** Authors wish to thank anonymous reviewers for critically reading this transcript, valuable suggestions, and thoughtful comment.

## References

1. R. Lujan, F. Lledias, L. Martínez, R. Barreto, G.I. Cassab, & J. Nietosotelo, Small heat shock proteins and leaf cooling capacity account for the unusual heat tolerance of the central spike leaves in *Agave tequilana* var. Weber. *Plant, cell & environment*, 2009. 32(12): p. 1791-1803.
2. S.C. Davis, F.G. Dohleman, and S.P. Long, The global potential for *Agave* as a biofuel feedstock. *Gcb Bioenergy*, 2011. 3(1): p. 68-78.
3. D.B. Lobell, and G.P. Asner, Climate and management contributions to recent trends in US agricultural yields. *Science*, 2003. 299(5609): p. 1032-1032.
4. D.B. Lobell, and C.B. Field, Global scale climate-crop yield relationships and the impacts of recent warming. *Environmental research letters*, 2007. 2(1): p. 014002.
5. M. Hasanuzzaman, et al., Plant response and tolerance to abiotic oxidative stress: antioxidant defense is a key factor, in *Crop stress and its management: Perspectives and strategies*, 2012, Springer. p. 261-315.
6. M. Hasanuzzaman, K. Nahar, and M. Fujita, *Extreme temperature responses, oxidative stress and antioxidant defense in plants*, 2013: INTECH Open Access Publisher.
7. J. Krasensky, and C. Jonak, Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. *Journal of experimental botany*, 2012. 63(4): p. 1593-1608.
8. M.A. Semenov, and N.G. Halford, Identifying target traits and molecular mechanisms for wheat breeding under a changing climate. *Journal of experimental botany*, 2009. 60(10): p. 2791-2804.
9. Y.Y. Sin, et al., Small heat shock protein 20 (Hsp20) facilitates nuclear import of protein kinase D 1 (PKD1) during cardiac hypertrophy. *Cell Communication and Signaling*, 2015. 13(1): p. 16.
10. S.Y. Seong, and P. Matzinger, Hydrophobicity: an ancient damage-associated molecular pattern that initiates innate immune responses. *Nature Reviews Immunology*, 2004. 4(6): p. 469-478.
11. V.S. Murthy, and K.V. Ravishankar, *Molecular Mechanisms of Heat Shock Proteins and Thermotolerance in Plants*, in *Abiotic Stress Physiology of Horticultural Crops* 2016, Springer. p. 71-83.

- 12.K.C. Kregel, , Invited review: heat shock proteins: modifying factors in physiological stress responses and acquired thermotolerance. *Journal of applied physiology*, 2002. 92(5): p. 2177-2186.
- 13.D.M. Walther, et al., Widespread proteome remodeling and aggregation in aging *C. elegans*. *Cell*, 2015. 161(4): p. 919-932.
- 14.P.C. Phillips, Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, 2008. 9(11): p. 855-867.
- 15.G.t. Ramachandran, and V. Sasisekharan, Conformation of polypeptides and proteins. *Advances in protein chemistry*, 1968. 23: p. 283-437.
- 16.R.A. Laskowski, et al., PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography*, 1993. 26(2): p. 283-291.
- 17.M. Wiederstein, and M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*, 2007. 35(suppl 2): p. W407-W410.
- 18.J. Skolnick, and J.S. Fetrow, From genes to protein structure and function: novel applications of computational approaches in the genomic era. *Trends in biotechnology*, 2000. 18(1): p. 34-39.
- 19.M. Remmert, et al., HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature methods*, 2012. 9(2): p. 173-175.
- 20.A.P. Arrigo, and W.E. Müller, *Small stress proteins*. Vol. 28. 2012: Springer Science & Business Media.
- 21.W. Sun, M. Van Montagu, and N. Verbruggen, Small heat shock proteins and stress tolerance in plants. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, 2002. 1577(1): p. 1-9.
- 22.M.H. Al-Whaibi, Plant heat-shock proteins: a mini review. *Journal of King Saud University-Science*, 2011. 23(2): p. 139-150.
- 23.B. Asthir, Mechanisms of heat tolerance in crop plants. *Biologia plantarum*, 2015. 59(4): p. 620-628.