Biological Forum — An International Journal, **1**(2): 1-4(2009)

Comparative modeling and analysis of 3-D structure of Hsp 70, in *Cancer irroratus*

Barneet Singh, Ashish Gupta and Sharada M. Potukuchi

School of Biotechnology, Shri Mata Vaishno Devi University, Katra (J&K), INDIA

ABSTRACT : The present study was undertaken to predict the 3-D structure of heat shock protein/s (Hsp/s) from Cancer irroratus, a crab belonging to the family Cancridae. Hsps are a class of functionally related proteins whose expression is increased when cells are exposed to elevated temperatures or other stresses like starvation, hypoxia, etc. The structural template for Hsp was identified from structural database using homology modelling or comparative modelling approach. Based on the knowledge of the template, a 3-D model was predicted and processed through energy minimization, Ramachandran plot analysis and quality assessment. The sequence predicted has been deposited into PMDB having ID PM0075773.

Keywords : Modeller, Protein Model Database, RAM-Page, Errat, Swissprot, dope score, alignment

Abbreviations: Hsp-Heat shock protein, BLAST-Basic local alignment search tool, PMDB-Protein model database, PIR-Protein information resource, CW-Cephalothorax width, ATP- Adenosine triphosphate, ADP- Adenosine diphosphate, AMPK- AMP activated protein kinase, HSF-1 – Heat shock factor-1.

INTRODUCTION

Cancer irroratus (Atlantic rock crab), belongs to the kingdom Animalia and family Cancridae. It usually occurs on the eastern coast of North America, from Labrador to Florida and lives in the benthic zone of the ocean (Ristvey and Rebach, 1999). *Cancer irroratus* does not remain in one spot for too long. As a scavenger it is extremely mobile. Their main method of defense is to pinch if provoked (Gosner, 1978). This species of crabs is known to possess a protein Hsp 70 which show an upregulation on exposure to different kinds of environmental stress conditions, such as infection, inflammation, exercise, exposure of the cell to toxins (ethanol, arsenic, trace metals and ultraviolet light, among many others), starvation, hypoxia (oxygen deprivation), nitrogen deficiency (in plants), and amino acid analogues (De Maio A, 1999).

The protein Hsp 70 is named according to its molecular weight, for example, Hsp60, Hsp70 and Hsp90 refer to families of heat shock proteins on the order of 60, 70 and 90 kilodaltons in size, respectively (Li Z *et al.*, 2004). The synthesis of Hsp is induced when environmental variation perturbs the organism's physiological system to the extent that its proteins denature. Under such environmental conditions, Hsps and other molecular chaperones stabilize denaturing proteins, refold reversibly denatured proteins, and facilitate the degradation of irreversibly denatured proteins (Lindquist, 1986; Lindquist and Craig, 1988; Parsell and Lindquist, 1994). The main reason for choosing this protein was its role as a temperature stress indicator in virtually all species, from bacteria to humans. HSF-1 is the major regulator of Hsp 70 transcription in eukaryotes. In the absence of cellular stress, HSF-1 is inhibited by association with Hsps and is therefore not active. Cellular stresses, such as increased temperature, can cause proteins in the cell to misfold. Hsps bind to the misfolded proteins and dissociate from HSF-1. This allows HSF-1 to form trimers and translocate to the cell nucleus and activate transcription (Prahlad *et al.*, 2008).

Experiments performed on Antarctic Crustaceans(Rock Crabs) compared the AMPK activity and Hsp 70 protein levels at 2°C increments and it was found that Hsp 70 started to increase at 28°C; the crabs' critical temperature (the temperature at which the animals switch to anaerobic metabolism due to heat stress). AMPK activity was found to steadily increase above 18°C revealing that the crabs were already heat stressed at this temperature revealing their behavior threshold. (Frederich *et al.*, 2009).

Homology modeling, also known as **comparative modeling** of protein refers to constructing an atomic resolution model of the "*target*" protein from its amino acid sequence and an experimental 3-D structure of a related homologous protein "*template*". Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. The sequence alignment and template structure are then used to produce a structural model of the target(Marti-Renom *et al.*, 2000).The understanding of the 3-D structure of a protein would be a precious aid to understand the details of a particular protein. The main objective of this study is an attempt to predict the structural information of Hsp 70 from *Cancer irroratus*, which is produced in increased amounts during stress conditions.

MATERIAL AND METHOD

Retrieval of *Cancer irroratus* –Hsp 70 Protein sequence. The protein sequence of Hsp 70 in *Cancer irroratus* was retrieved from the Swissprot database (http://www.expasy.org) and taken as target sequence. An extensive searched revealed that the 3-D structure of this protein was not available in any structural databases. Hence, the current study of developing the 3-D structure of Hsp 70 from *Cancer irroratus* was undertaken.

Selection of Structural Template. An effort was made to find a suitable structural homolog or template for the modeling of Hsp 70. A structural template was obtained from protein BLAST (Altschul *et al*., 1990) and it used Protein Data Bank (Berman *et al*., 2003) as reference for identifying the closely related sequences.

Target-Template Alignment. The protein sequence of Hsp 70 was aligned with its corresponding template by using align-2D module in MODELLER 9V5 (Eswar *et al.*, 2008), which required two files, one containing target sequence and another the structural coordinates of template. This step is essential to identify the common conserved residues or active residues present in both the sequences.

Model Building. MODELLER 9V5 was used to predict the 3-D structure of Hsp 70 using model-single.py based on satisfaction of spatial restraints. It is a python script, used to predict the 3-D model from single template. Theoretical model was subjected into Swiss-PDB Viewer (Kaplan and Littlejohn, 2001) for energy minimization using the steepest descent and conjugate gradient technique to correct the stereochemistry of the model. Computational analysis was carried out in vacuum with the GROMOS96 43b1 parameters set, without reaction field in Swiss-PDB Viewer.

Model Evaluation. The refined model obtained was subjected to a series of tests for testing its internal stability and reliability. Backbone conformation of the refined model was assessed by the examination of the Psi/Phi Ramachandran plot obtained from RAMPAGE web server (Lovell *et al*., 2003). Errat web server (Colovos and Yeates, 1993) was used to explore the statistics of non-bonded interactions between different atom types to plot the graph.

RESULTS AND DISCUSSION

The amino acid sequence of Hsp 70 was retrieved from commonly used primary protein sequence database i.e. Swissprot (http://www.expasy.org) and the accession number was ACL13566 and the source was of *Cancer irroratus.* The results of Protein Blast (http://www.ncbi.nlm.nih.gov) search for suitable template structure related that the target sequence, Hsp 70 showed Chain A , structural basis of the 70-kilodalton heat shock cognate protein ATP hydrolytic activity, Ii. structure of the active site with ADP or ATP bound to wild type and mutant ATPase fragment with highest sequence similarity (78%), as the most suitable template for modelling (PDB ID: 1NGD). The alignment of Hsp 70 and its corresponding template was carefully examined and conserved regions were identified and it was concluded that this alignment can be assisted to generate a 3-D model. Jalview (Clamp *et al*., 2004) was used to display the conserved regions in graphical representation. (Fig. 1). Once target-template alignment was completed, a 3-D structure of Hsp 70, was predicted using the program MODELLER 9V5 produced twenty different conformations and the dope score in increasing order of Model 3 , model 8, model 18, model 4 were –11700.68359, –11580.00879, –11489.75684, –11406.33887 respectively,

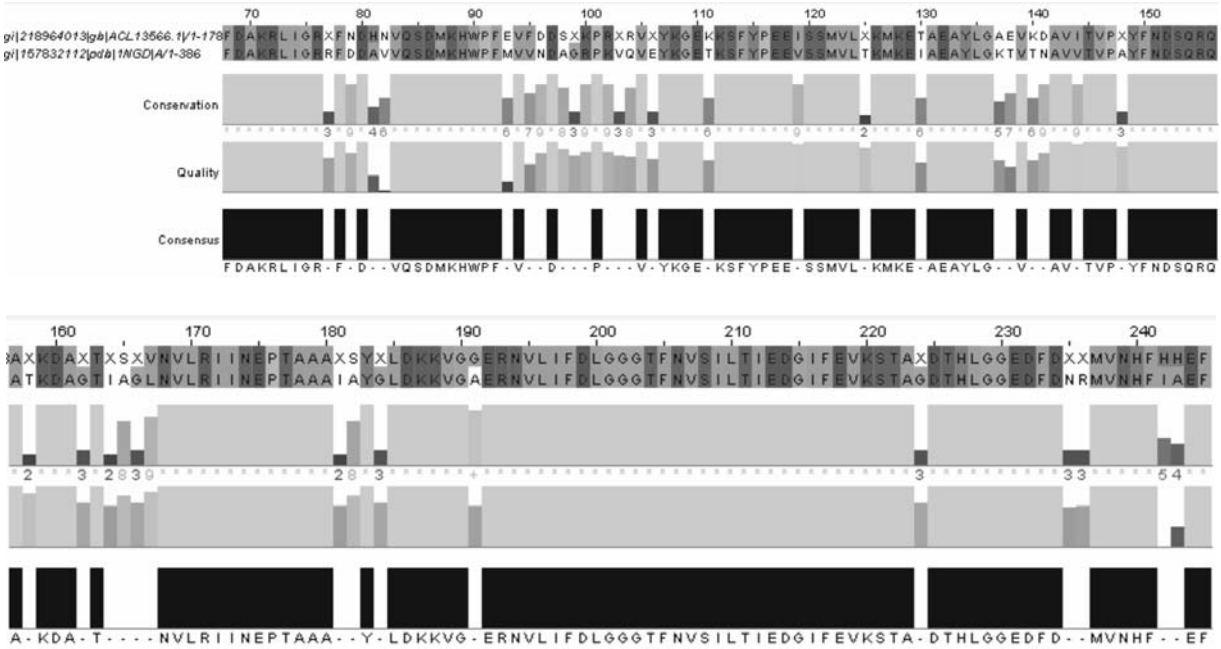

Fig. 1. Graphical representation of sequence alignment of both target [Hsp 70] and template [1NGDA]. Various conserved regions with highlighted 100% conserved residues.

Model 3 had lowest value compared to others and processed into Swiss- PDB Viewer for energy minimization. With the help of align2d.py file, length of alignment was found to be 378. The expect value shown in the BLAST result of the protein was found to be $4e^{-79}$ and the score was shown to be 289 bits (739).

An assessment of the refined model involved two independent tests. The first test was to compare the residue backbone conformations in our refined model with the referred values obtained from Protein Data Bank of known structures. The results of SAVES server Procheck (http://nihserver.mbi.ucla.edu/SAVES/) indicated that 91.9% residues were found to be in the most favoured region of the Ramachandran Plot of refined model of Hsp 70 which is more than cut-off of 90% in most reliable models (Lovell *et al*., 2003) (Fig. 2). The stereochemical quality of the predicted model was found to be satisfactory and low percentage of residues having phi/psi angles in the outlier region. A 3-D structure of Hsp 70 is shown (Fig. 3).

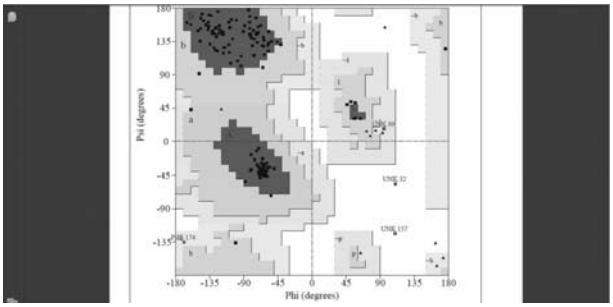

Fig. 2. Plot generated by Procheck (3.0).Ramachandran plot of model 3 of Hsp 70.

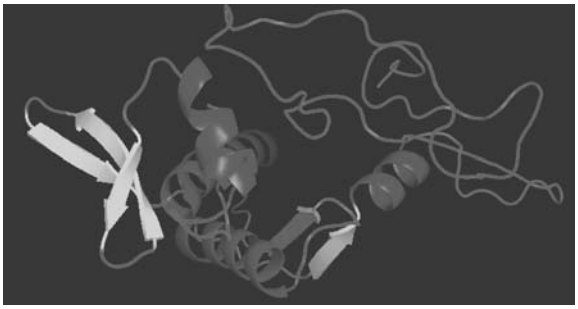

Fig. 3. The final three-dimensional structural representation of Hsp 70 [PyMOL] generated by MODELLER 9v5 Program.

Residues in most favoured regions [A,B,L]	147	91.9%
Residues in additional allowed regions [a,b,l,p]	9	5.7%
Residues in generously allowed regions [~a,~b,~l,~p]	2	1.2%
Residues in disallowed regions	2	1.2%
Number of non-glycine and non-proline residues	160	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	11	
Number of proline residues	5	
Total number of residues	178	

The second test was carried out using Errat web server (http://nihserver.mbi.ucla.edu/ERRATv2/) to check the quality of models. Generally, the quality factor of high resolution structures produces values around 90% or higher (Colovos and Yeates, 1993). Here, the overall quality factor of the refined model (58.462) was predicted from Errat. The evaluated final reliable model has been deposited into Protein Model Database (http://mi.caspur.it/PMDB/) and is now publicly accessible [PMDB ID: PM0075773].

CONCLUSION

In silico study of proteins and nucleic acids are helpful in almost all research fields. It not only saves money but also valuable time. This model may be further used in characterizing the protein in further wet laboratory experimentations.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol*, **215:** 403-410.

Berman, H., Henrick, K. and Nakamura, H. (2003). Announces the worldwide Protein Data Bank.*Nat Struct Biol*, **10:** 980.

Binkowski, T.A., Naghibzadeh, S. and Liang, J. (2003). CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic Acids Res*, **31:** 3352-3355.

Clamp, M., Cuff, *J. Searle, S.M. and Barton*, G.J. (2004). The Jalview-Java Alignment Editor. *Bioinformatics*, **20:** 426-427

De Maio, A. (1999). "Heat shock proteins: facts, thoughts, and dreams". Shock **11**(1)**:** 1–12.

Eswar, N., Eramian, D., Webb, B., Shen, M.Y. and Sali, A. (2008). Protein Structure Modelling with MODELLER. *Methods Mol Biol*, **426:** 145-159.

Frederich, M. O., Rourke, M. R., Furey, N. B. and Jost, J. A. (2009). AMP- activated protein kinase (AMPK) in the rock crab, *Cancer irroratus*: an early indicator of temperature stress. *J. Exp. Biol*, **212:** 722-730.

Gosner, K. (1978). *A Field Guide to the Atlantic Seashore*. Boston: Houghton Mifflin.

Kaplan, W. and Littlejohn, T.G. (2001). Swiss-PDB Viewer (Deep View). *Brief Bioinform*, **2:** 195-197.

Li, Z. and Srivastava, P. (2004). *Heat-shock proteins*. (Ed. John E. Coligan) Current protocols in immunology / Appendix 1T. John Wiley and Sons Inc, Malden, MA, USA.

Lindquist, S. (1986). The heat-shock response. *Annu. Rev. Biochem*, **55:** 1151–1192.

Lindquist, S. and Craig, E. A. (1988). The heat-shock proteins. *Annu. Rev. Genet*, **22:** 631–678.

Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I. and Word, J.M. (2003). Structure validation by C alpha geometry: phi, psi and C beta deviation. *Proteins*, **50:** 437-450.

Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F. and Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*, **29:** 291-325.

Parsell, D. A. and Lindquist, S. (1994). *Heat shock proteins and stress tolerance*. (Eds.Morimoto, R.I., Tissieres, A. and Georgopoulos). *The Biology of Heat Shock Proteins and Molecular Chaperones*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY., Vol **26:** p.457-494.

Prahlad, V. and Morimoto, R.I. (2008). Integrating the stress response: lessons for neurodegenerative diseases from *C. elegans*. *Trends in Cell Biology*, **19**(2)**:** 52–61.

Ristvey, Rebach, Andrew and Steve. (1999). Enhancement of the Response of Rock Crabs, Cancer irroratus, to Prey Odors following Feeding Experience. *The Biological Bulletein*.