

Computational Comparative Homology Based 3D-structure Modelling of the HSp70 Protein from GWD

Mishra S* and Gomase VS

Department of Biotechnology, Mewar University, Chittorgarh, India

Abstract

Heat shock protein participates in several broad ranges of protein folding processes and control activity of regulatory proteins. Nearly all cells react in a similar manner against the external environmental abrupt changes like heat, chemicals such as analogues of amino acid, ethanol, arsenite, several heavy metals, mitochondrial functional inhibitors, cold, UV light. These all causes several severe, rapid changes in at the level of gene expression. The major significant activity of HSP protein to stabilize newly synthesized protein by securing its right folding, which generally undergone any damaged under cell stress condition. Through overexpression, this protein provides protection to cells against apoptotic cell death. The hsp70 chaperonis targeted as novel anticancer therapeutic candidate for a broad spectrum of tumor types due to its diverse functional role. Homology modeling methodology allows modeling the 3D structure of the protein using the experimentally determined three dimensional structure of related homologues protein as template. In the current investigation, we have taken the Hsp70 protein and conducted the 3-dimensional structure prediction through Swiss-Model by using the crystal structure as the template. Template search of the desired protein were conducted with Blast and HHBlits against the SMTL and keyed out the highest sequence identity, comparatively, above the other obtained template sequences. The homology modelling was performed and the model protein was evaluated by using the PSVS.

Keywords: HSP70; Homology modelling; Protein structure; Molecular modelling

Abbreviations

GWD: Guinea Worm Disease; HLA-DR: Human Leukocyte Antigen-Antigen D Related; MHC: Major Histocompatibility Complex; HSp70: Heat Shock Protein 70; BLAST: Basic Local Alignment Search Tool; HHBlits: Lightning-Fast Iterative Protein Sequence Searching by HMM-HMM Alignment; SMTL: SWISS-MODEL Template Library; QMEAN: Qualitative Model Energy Analysis; PSVS: The Protein Structure Validation Software Suite; PDB: Protein Data Bank

Introduction

Heat shock protein participates in several broad ranges of protein folding processes and control activity of regulatory proteins. Nearly all cells react in a similar manner against the external environmental abrupt changes like heat, chemicals (like amino acid analogues, ethanol, arsenite, several heavy metals, and certain mitochondrial functional inhibitors, cold [1-2], UV light [3] and during wound healing or tissue remodeling [4] which in turns results inspeedy changes in gene expression levels. The 70-kD class of stress proteins (HSP) draws a major attention because of its diverse range of cellular function. In most of the species it has been found that, this protein is purely stress-inducible, the notable exception being primate cells [5]. In human cells, HSP70 is cell cycle regulated [6]. The virtual abundance of the HSPs can be observed in all living organisms starting from bacteria to humans. HSP-70 protein are named according to its molecular weight. Hsp70s (70-kDa) proteins provide assistance in widely distributed protein folding operations, assembling of freshly formed proteins its folding and re-folding action for misfolded protein and aggregated proteins, and controlling the regulatory proteins activity [7-13]. This protein has also performed functions of housekeeping in the cell. In an ATP controlled fashion this protein interacts with hydrophobic peptide segments. A peptide binding domain (PBD) and the amino-terminal ATPase domain (ABD) is the two distinct functional regions of this protein. For the neutral and hydrophobic amino acid residues affinity the PBD region of HSP70 holds a groove, whereas, the C-terminal or

ABD is rich in alpha helical structure which behaves as a 'lid' for the substrate binding domain. When ATP is bounded with protein the lid remains open and allows peptides to bind and release rapidly but when ADP is bound, the lid remains closed, and peptides are strongly bound to the substrate binding domain. In malignant melanoma this protein is over-expressed in cell [14] but under expression is also seen in renal cell cancer [15]. The extracellular hsp perform a potent route for sending the danger signals to the host immune system in order to respond to an infection. The peptide complexes of hsp protein are also involve in the restricted antigen presentation of MHC classes (I&II) and enable the enhanced activation of T cells. The specific interaction of the mammalian cytosolic Hsp70 molecules with HLA-DR molecules, signifies the possibility of transferring the of bond antigenic peptides in the ternary complex into the binding groove of HLA-DR molecules by Hsp70 molecules. Rohrer et al., study suggest that the interaction of Hsp70 and HLA-DR takes place outside the peptide binding groove and is assigned to ATPase domain of heat shock protein 70 molecule, which enable the enhanced presentation of peptide to the antigenic presenting cells and improved T cells proliferations [16]. The computational method approach is the one of the reliable methodology to generate amino acid sequence into 3-D structure models [17] and a wide range of the approaches are routinely applied for such prediction for many biological applications. Homology modeling is based on the sensible assumption of two homologous proteins shares very standardized alike structures. The name homology modelling conveys exactly what this procedure is about; modeling a structure using homologous model as template (which is usually an exact X-ray or NMR-determined

*Corresponding author: Mishra S, Department of Biotechnology, Mewar University, Chittorgarh, India, Tel: 9560808369; E-mail: sonumishra1014@gmail.com

Received April 29, 2016; Accepted June 16, 2016; Published June 23, 2016

Citation: Mishra S, Gomase VS (2016) Computational Comparative Homology Based 3D-structure Modelling of the HSp70 Protein from GWD. J Health Med Informat 7: 233. doi:10.4172/2157-7420.1000233

Copyright: © 2016 Mishra S, 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.

structure). In homology modeling it is important that modeler finds a template structure with the highest possible sequence-identity. If the identity between the input sequence and the template structure falls below 40%, the output model is likely to be implausible. Earlier study reveals that about the conservation of the protein structures than protein sequences in the midst of homologues [18].

Methodology

Retrieval of the protein sequence

For the present study heat shock protein 70 [*Dracunculus medinensis*] the protein sequence was retrieved from the NCBI having 647 amino acids sequence [AD154942.1] (Table1) [19].

Template search and selection

For the target protein, the templates were search and identified by performing BLAST and HHblits against the SWISS-MODEL template library (SMTL) [20-21]. A total of 96 templates were opted. For each identified template, the quality of the template's was predicted from features of the target-template alignment. Among of them; 3c7n.1 chain B with highest sequence identity (Sequence identity: 88.45), and lowest two sequence similarity found are 2v7z.2 chain A and 2v7z.1 chain A with sequence identity: 88.40 value had then selected as template for building model.

Molecular modeling of protein

The methodological strategy used in the model building was based on the target-template alignment using Promote-II. The template were selected based on the maximum sequence similarity. The 3c7n.1 found to have the maximum sequence similarity and taken as template and alignment were performed between the both the targets and template. The coordinates which are found as conserved between the targets and templates are copied from the template to the model. Using a fragment library, insertion and deletion are remodelled and the rebuilding of the side chains was performed. Finally, the geometry of the resulting model is regularized by using a force field. The obtained model is visualized with Deep Viewer. The global and per-residue model quality has been assessed using the QMEAN4 scoring function [22].

Ligand modelling

The number of the ligands which were found to be present in the template structure (Ligands such as SO₄, BEF-ADP, ADP) are reassigned by homology in order of modelling, where the ligands shouldn't be clashing with the target protein and the residue which are in contact with the ligand were conserved among the target and the template.

The protein model statistical assessment

The protein model statistical Assessment were performed via the protein structure validation software suite (PSVS). PSVS were used for those protein structures which are generated from NMR, X-ray crystallographic and homology modelling methods. PSVS incorporates analyses from numerous widely-used structure quality evaluation tools, including RPF, PROCHECK, MolProbity, Verified 3D, Prosa II, the PDB validation software and various structure validation tools [23]. PSVS provides a standard constraint analyses, statistics on the PDB validation goodness-of-fit between structures and experimental data, Z-score values and knowledge-based structure quality scores in a standardized format suitable for database integration.

Visualization of 3D model

The generated model was visualized in 3D using the RasMOL molecular 3D viewer. The RasMol generated model information regarding chain, atoms, groups and bonds of 3D model were extracted and analysed [24].

Results and Interpretation

For the recognized 3-D protein structure model, the complete assessment were performed, analyzed and described into the following leads

Template validation: For each identified template, the template's quality has been predicted from the features of the target-template alignment as shown in Table 2. The templates with the highest quality have been selected for the model building. Template 3c7n.1.B validations for the generated model of Hsp70 protein was analyzed and were employed structure analysis for template chain. The multiple sequence alignment result of target protein sequence and the template 3c7n.1 chain B is found to be 88.45% identity by HHblits as shown in the Figure 1.

Model analysis: HSP70 protein model structure was prepared for the target-template alignment, because the template 3c7n.1 has the highest quality and alignment with the target (Table 3), therefore selected for modelling as shown in Figure 2.

The result of PSVS provides the stereo-chemical property of the model. The molecular weight of the model is 131894. RMS deviation of the bond angle is 1.4°. Number of close contacts (within 2.2 Å): 0 and the bond lengths is 0.011 Å. With respect to mean and standard deviation for a set of 252 X-ray structures <500 residues, of resolution ≤ 1.80 Å, R-factor ≤ 0.25 and R-free ≤ 0.28; a positive value indicates

```
MAKHNAVGLDGTYSVGVFMHGKVEIANDQGNRTTPSYVAFTDTERLIGDAAKNQVAMNPNNTV
FDAKRLIGRRFDDPAVQADMKHWPFKVINAECSKPKVQVEYKGETKTFTPEEISSMVLLKMKETA EAF
LGSTVKDAVVTVPAYFNDSQRQATKDAGAIAGLNVLRIINEPTAAAIAYGLDKKGHGERNVLI FDLGG
GTFDVSILTIEDGIFEVKSTAGDTHLGGEDFDNRMVNHVFAEFKRKHKKDLSTNPRALRRLRTACERAK
RTLSSSSQASIEIDSLFEGIDFYTNITRARFEELCADLFRSTMDPVEKALRDAKMDKSQMHDIVLVGGST
RIPKVQKLLSDFSGKELNKSINPDEAVAYGAAVQAAILSGDKSEAVQDLLLLLDVAPLSLGIETAGGVM
TALIKRNTTIPTKTAQTFTTYSNQPGLIQQVFEGERAMTKDNNLLGKFELSGIPPAPRGVVPQIEVTFDID
ANGILNVSAQDKSTGKQNKITITNDKGRLSKDEIERMVQEAKEYKADDEA QKDRIA AKNALESYAFN
MKQTIDDEKLKDKLSADDRKKIEDKCDEIIKWLDRNQTAEKDEFEHHQKELEAVCNPIITKMYQSAGG
MPGNPGGFPGGGAPGGGHQGGGGPTIEEVD
```

Table 1: Primary amino acid sequences of target protein for which template were searched for models building.

a 'better' score Selected residues: 3A-547B. The model protein 3D structure is shown in the Figure 3.

Ramachandran plot statistics: The Ramachandran plot displays the psi and phi backbone conformational angles for the each residue in the target protein Hsp70 as shown in the Figure 4. The displayed darkest region as 'red' in colour and correspond to the "core" region and represent the most favourable combination of phi-psi values. Few residues found in allowed region. The percentage of residue in the core regions are described as follows in Table 4.

PROCHECK output: The protein stereochemical quality was analysed. The PDB structure of HSp70 protein was examined by PROCHECK tool. Procheck G-factor evaluated (Figure 5) probability of dihedral angles of a residue types to be within a given range as below:

(a) Procheck G factor for phi-psi for ordered residue overall is -0.659.

(b) Procheck g-factor for all dihedral angles for ordered residue overall is -0.619.

Output from MolProbity

The MolProbity server (in PSVS server) is a one of the worthwhile structure validation tool in the final stage of structure refinement. VDW violations from MAGE calculate MAGE VdW clashscore: Mean 38.49; SD: 0.0000. MolProbity clash score and visualize atomic overlap and beta position deviations (Figure 6).

PDB validation software output

After the PDB validation software analysis 3.5 Angstroms are consider for hydrogen bonding in the asymmetric unit and 2.2 Angstroms are considered as close contact for heavy atoms in same asymmetric unit. Distances smaller than 2.2 Angstroms are considered as close contacts. The RMS deviation for covalent bonds relative to the standard dictionary is 0.011 Angstroms (Tables 5 and 6).

The following table contains a list of the covalent bonds greater than 6.0**RMSD*.

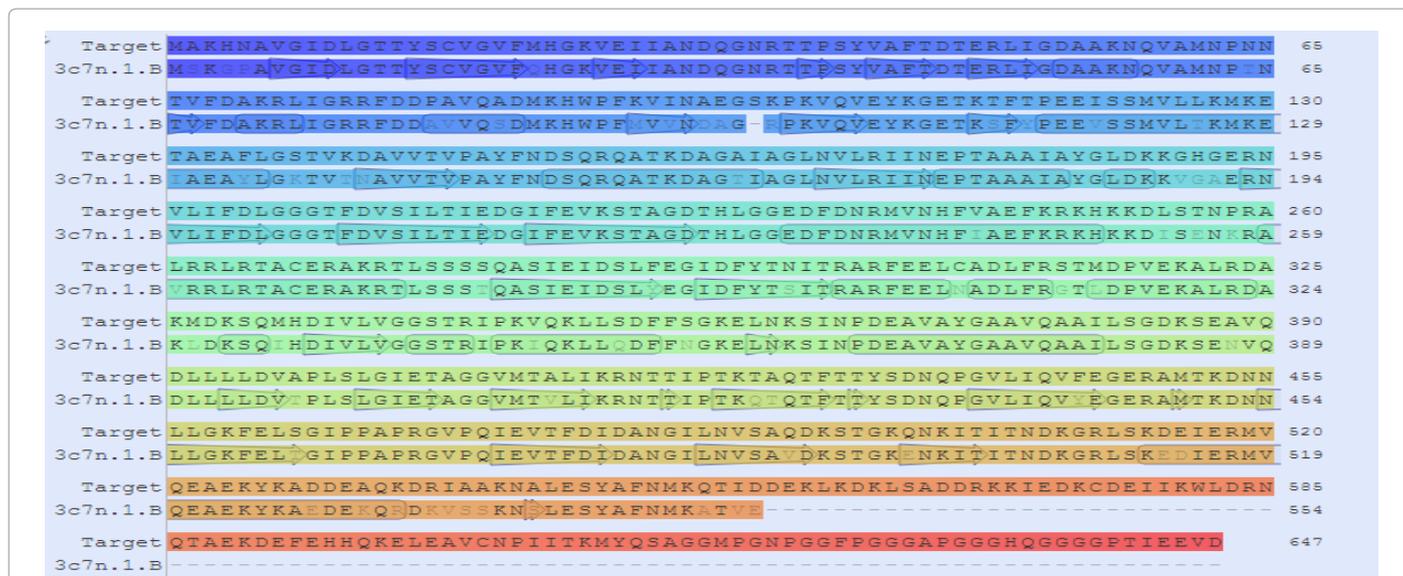


Figure 1: Hsp70 protein sequence and template 3c7n.1 chain B multiple sequence alignment result with 89.06% of identity obtained through HHblits.

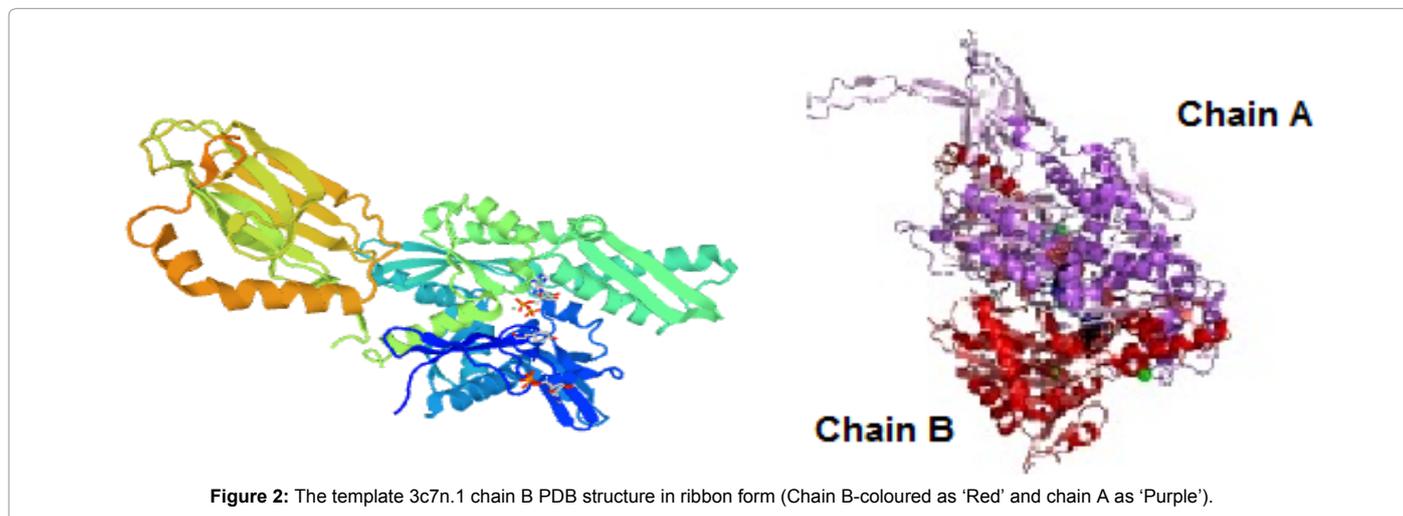


Figure 2: The template 3c7n.1 chain B PDB structure in ribbon form (Chain B-coloured as 'Red' and chain A as 'Purple').

Template	Seq Identity	Oligo-State	Found by	Method	Resolution	Sequence similarity	Range	Coverage	Description
3c7n.1.B	88.45	Hetero-oligomer	HHblits	X-ray	3.12Å	0.57	1-555	0.86	Heat shock cognate

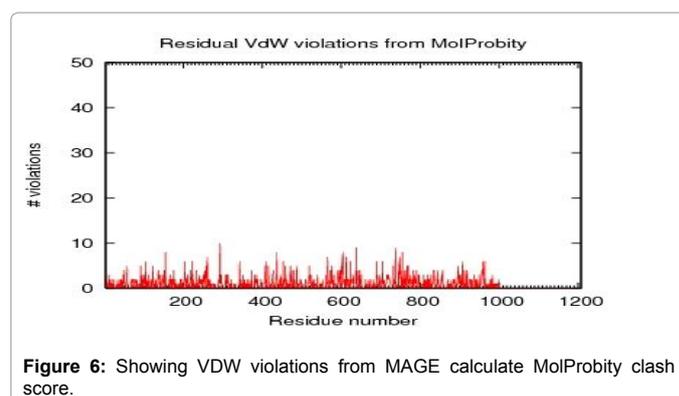
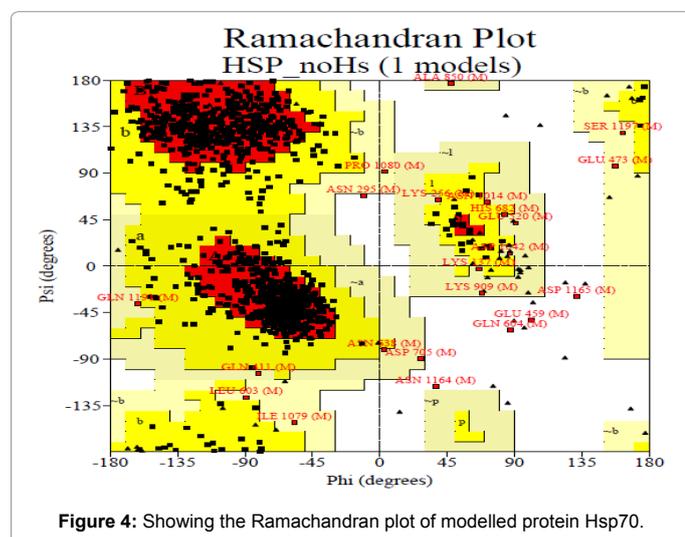
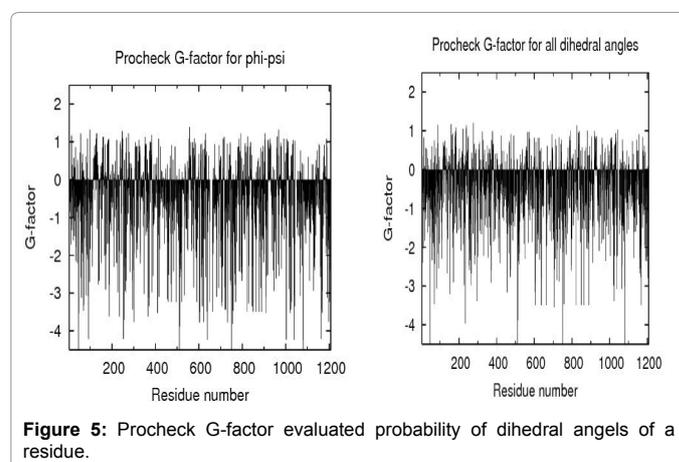
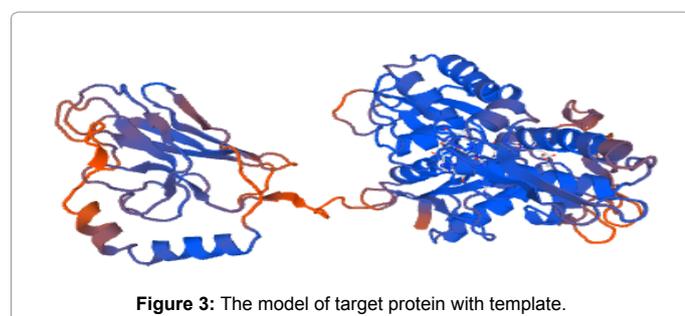
Table 2: 3c7n.1 chain B with highest sequence similarity chosen as the template for the protein modeling.

Model Properties	Data
Number of Atoms	4403
Number of Bonds	4343

Table 3: The template model important information.

	From PROCHECK	From Richardson Lab's Molprobrity % target	
Most favoured regions	81.40%	85.8%	-
Additional allowed regions	16.60%	10.9%	-
Generously allowed region	1.50%	N/A	-
Disallowed regions	0.50%	3.3%	-
Structure quality-factors-overall statistics	Mean Score	SD	Z-Scores
Procheck G-factor (phi/psi only)	-0.66	N/A	-2.28
Procheck G-factor (phi/psi only)	-0.62	N/A	-3.67
MolProbityClashscore	38.49	0.0000	-5.08

Table 4: The percentage of residue in the core regions.



Deviation	Residue Name	Chain ID	Sequence Number	AT1 - AT2 Distance	Bond Value	Dictionary
0.094	LYS	A	197	CD - CE	1.614	1.52
0.068	PRO	A	512	CB - CG	1.56	1.492
0.141	ARG	B	301	CG - CD	1.661	1.52

Table 5: List of the covalent bonds greater than 6.0° RMSD. The RMS deviation for covalent angles relative to the standard dictionary is 1.4 degrees.

Deviation	Residue Name	Chain ID	Sequence Number	AT1 - AT2 - AT3 Distance	Bond Value	Dictionary
8.4	THR	A	189	N - CA - C	119.6	111.2
-9.8	LYS	A	219	N - CA - C	101.4	111.2
-9.1	LYS	A	431	N - CA - C	102.1	111.2
9.9	GLY	B	224	N - CA - C	122.4	112.5

Table 6: List of the covalent bond angles greater than 6.0° RMSD.

Chirality

The chirality checking was performed and it was found that, that there are no incorrect carbon chiral centers.

Conclusion

In this current study, we have modelled a 3D structure of the hsp70 protein by homology modeling and visualized with the help of online computational tools. 3-D models of hsp70 protein shows significant amino acid sequence similarity with the target sequence. Homology modelling suggested the similarity between targets: Template sequences. In this model, a template is a homologous protein that can be identified by a sequence similarity with target, and 89.45% identity was identified. Protein validation prediction indicates about the region where residues are present. Ramachandran Plot analysis from PROCHECK which indicated the maximum of the residues present in most favoured region, i.e., 81.40% of the residues were found in the most favoured region and from Richardson's lab Molprobitry present 85.8% favoured regions for the selected residues. 3D structure of nematode protein-hsp70 protein structure reported extremely in any cases in GWD. In the present current study we used GWD-HSP70 protein sequence as target with 3c7n.1 template for the modeling of the target protein sequence, because at the point of the template selection procedure, we found that the target protein template provides more sequence identity in comparison to other founded template sequences. This study may provide the future prospects to illustrate consideration towards a computational approach for 3D molecular modelling and computer generated model are expected to be the most accurate model but it can't be a substitute of a crystal structure. The HSP70 3-D model representation will prove to be a useful model for exploiting in the dracunculiasis disease outbreak database and residue or derivatives relationship and experimental verification in future and could be helpful in drug designing and development.

References

- Matz JM, Blake MJ, Tatelman HM, Lavoie KP, Holbrook NJ (1995) Characterization and regulation of cold-induced heat shock protein expression in mouse brown adipose tissue. *Am J Physiol* 269: R38-R47.
- Wong HR (1999) Heat shock proteins. Facts, thoughts, and dreams. *A. De Maio. Shock* 11:1-12, 1999. *Shock* 12: 323-325.
- Cao Y, Ohwatori N, Matsumoto T, Kosaka M, Ohtsuru A, et al. (1999) TGF-beta1 mediates 70-kDa heat shock protein induction due to ultraviolet irradiation in human skin fibroblasts. *Pflugers Arch* 438: 239-244.
- Laplante AF, Moulin V, Auger FA, Landry J, Li H, et al. (1998) Expression of heat shock proteins in mouse skin during wound healing. *J Histochem Cytochem* 46: 1291-1301.
- Welch WJ, Suhan JP (1986) Cellular and biochemical events in mammalian cells during and after recovery from physiological stress. *J Cell Biol* 103: 2035-2052.
- Milarski KL, Morimoto RI (1986) Expression of human HSP70 during the synthetic phase of the cell cycle. *Proc Natl Acad Sci U S A* 83: 9517-9521.
- Bukau B, Deuerling E, Pfund C, Craig EA (2000) Getting newly synthesized proteins into shape. *Cell* 101: 119-122.
- Hart FU, Hart HM (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* 295: 1852-1858.
- Young JC, Barral JM, Ulrich Hartl F (2003) More than folding: localized functions of cytosolic chaperones. *Trends Biochem Sci* 28: 541-547.
- Neupert W, Brunner M (2002) The protein import motor of mitochondria. *Nat Rev Mol Cell Biol* 3: 555-565.
- Ryan MT, Pfanner N (2001) Hsp70 proteins in protein translocation. *Adv Protein Chem* 59: 223-242.
- Pratt WB, Toft DO (2003) Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery. *Exp Biol Med (Maywood)* 228: 111-133.
- Toft DO (1999) Control of hormone receptor function by molecular chaperones and folding catalysts. In: *Molecular Chaperones and Folding Catalysts. Regulation, Cellular Function and Mechanism*. Bukau B, Harwood Academic Publishers, Amsterdam, pp: 313-327.
- Ricaniadis N, Katakaki A, Agnantis N, Androulakis G, Karakousis CP (2001) Long-term prognostic significance of HSP-70, c-myc and HLA-DR expression in patients with malignant melanoma. *Eur J Surg Oncol* 27: 88-93.
- Ramp U, Mahotka C, Heikaus S, Shibata T, Grimm MO, et al. (2007) Expression of heat shock protein 70 in renal cell carcinoma and its relation to tumor progression and prognosis. *Histol Histopathol* 22: 1099-1107.
- Rohrer KM, Haug M, Schwörer D, Kalbacher H, Holzer U (2014) Mutations in the substrate binding site of human heat-shock protein 70 indicate specific interaction with HLA-DR outside the peptide binding groove. *Immunology* 142: 237-247.
- Bagchi P (2014) "Ayurinformatics : Selecting the best ligand for schizophrenic ProteinD (3) Dopamine Receptor (DRD3) from the Compound of Valerianawallichii Asparagus racemosus and Acoruscalamus". *Int J Pharm Bio Sci* 1: 1-10.
- Chothia C, Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J* 5: 823-826.
- <http://www.ncbi.nlm.nih.gov/>

20. <http://swissmodel.expasy.org/> quality of individual protein structure models. *Bioinformatics* 27: 343-350.
21. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Res* 25: 3389-3402.
22. Benkert P, Biasini M, Schwede T (2011) Toward the estimation of the absolute
23. http://psvs-1_5-dev.nesg.org/results/HSP/OUTPUT.html
24. <http://rasmol.org/>