

Computational analysis of HSP-60 protein with structural insights into chaperonin containing TCP-1 subunit 5 in *Bos taurus*

Abstract

Heat shock protein 60kDa (HSP60) is a crucial chaperone that guides appropriate folding of denatured proteins under heat stress conditions. HSP60 provides assistance in correct folding of a multitude of denatured proteins. The Group II eukaryotic chaperonin TCP-1 ring complex is (TRiC or CCT) important cytosolic chaperonin which plays important role in folding of essential subset of cytosolic proteins and is believed to be highly conserved among different eukaryotic species. In this study, an extensive *in silico* analysis has been performed ranging from sequence comparison among species to homology modeling of *Bos taurus* Cct5 protein and determination of Hsp60 interacting partners. The comparative analysis of the protein revealed certain significant variations in *Bos taurus*. The experimental protein structure for Cct5 in *Bos taurus* is unknown till date, therefore the putative protein structure was determined using homology modeling. The stereo chemical properties of protein were assessed by utilizing several scrutiny tools such as PROCHECK, VERIFY3D and RAMPAGE to ensure model accuracy. The mode of interaction between HSF1 and Cct5 protein was evaluated by molecular docking studies.

Keywords: heat shock protein, cct5, *bos taurus*, homology modeling, docking

Volume 6 Issue 1 - 2017

Ravinder Singh,¹ Vikash Kumar,² C Rajesh,¹ Ankita Gurao,³ Anurag Kulshrestha,⁴ Manika Sehgal,⁴ Aman Kaushik,⁵ Priyanka Sharma,⁴ Shailendra Kumar Mishra,⁴ Ranjit Singh Kataria⁴

¹Department of Biotechnology, Sri Guru Granth Sahib World University, India

²Department of Molecular Biology and Biochemistry, Guru Nanak Dev University, India

³Department of Veterinary Microbiology and Biotechnology, Rajasthan University of Veterinary, India

⁴National Bureau of Animal Genetic Resources, India

⁵School of Biotechnology, Gautam Buddha University, India

Correspondence: Ranjit Singh Kataria, National Bureau of Animal Genetic Resources, Haryana, India, Fax 01842267654 (Office), Tel 01842267153, Email katariaranji@yahoo.co.in

Received: May 14, 2017 | **Published:** August 09, 2017

Abbreviations: HSP, heat shock protein; CCT, chaperonin containing TCP-1 ring complex; TRiC, TCP-1 ring complex; chaperonin containing TCP-1 subunit 5

Introduction

Proteins being structural and functional blocks of all significant biological processes are synthesized as polypeptides comprising of amino acids. The molecular activities and functions are believed to be carried out via the protein machinery and other accessory components *in vivo*. The catalytic activity of an enzyme typically relies on its active site and the binding affinity to a compound. These sites permit the interactions among different proteins, their subunits or general structures which are devised for specific purposes. In order to retain the activity of active sites in a protein, correct folding of that particular protein is a vital process. In rare cases, the correct folding of these proteins is accomplished spontaneously, but frequently it requires the protein to interact with other proteins while it is being synthesized or while they are denatured under stress conditions. So these proteins which guide the appropriate folding of other proteins are called chaperones. The chaperone's machinery basically locates those features in proteins that are not exposed in matured confirmations, typically detecting hydrophobic stretches/regions that usually aggregate in the center of mature protein structures but are exposed during unfolding or under denatured conditions.¹

After translation, the newly synthesized proteins leave ribosomes as linear amino acid chains and then chaperone recognize region those which has ability to aggregate the chain. By binding to these interacting regions, chaperones prevent the inappropriate interactions

within different chains of proteins and monitor for the correct folding confirmation. The chaperonin folds the substrate in a closed central cavity, thereby providing the specific microenvironment favoring the folding reaction towards the phase where potential of folding is maximized and entropy gets minimized for the unfolded substrate. In eukaryotes the TCP-1/tcp-1 ring complex (CCT/TRiC) is one of the most important cytosolic subset of proteins required for living cells for maintaining their viability. This complex interacts with approximately ten percent of the freshly synthesized proteins and is involved in the folding of both skeleton and non-skeletal proteins.² It possesses a unique ability to fold some specific proteins such as actin, which are not folded by any other chaperonin. With the advent of proteomics studies, it was evident that CCT substrates include a number of proteins containing Tryptophan-Aspartic acid repeats involved in the formation of β -propeller domain. The size of CCT/TRiC substrates varies drastically ranging from small to very large and even some with molecular weight higher than 100kDa.³ These large proteins are folded by a mechanism of domain-wise folding where a substrate can be folded outside the cavity and thus, preventing encapsulation of the whole substrate.

Structurally, chaperonins are classified mainly into two groups: Group I and Group II chaperonins. The subunits of both groups share a similar basic structural pattern and consist of three domains: a) an apical domain which is related to the binding of substrate, b) equatorial domain, and function related to ATP binding, c) hinge domain which enables the connection between rests of the two domains.⁴ Endosymbiotic organelles and prokaryotic cells were found to contain in Group I chaperonin and Group II chaperonin in Eukaryotes and Archae.⁵

It was observed that CCT interacts with some very important proteins like actin, tubulin and von Hippel-Lindau tumor suppressor proteins etc. in subunit-specific and geometry dependent-manner.⁶ By using electron microscopy, three dimensional reconstruction of alpha helix showed the interaction of both small and large domain of alpha actin with different subunits of CCT, whereas Cct5p subunit found to interact with the large domain. Furthermore, electron microscopic studies and immunoprecipitation experiments of actin CCT complexes has revealed that Cct5 has highest substrate affinity than other subunits.⁷ On the other hand, tubulin comprises of total eight binding sites with CCT, three binding sites are located at N-terminal and five binding sites at C-terminal, among the five binding sites at C-terminal two were found to interact with Cct5.⁸

Heat shock transcription factors play important role in regulation of expression of molecular chaperones that are expressed during heat stress. The heat shock factors are the family of some transcription factors which controls the regulation of gene expression of proteins involved in folding of damaged or improper folded proteins during stress conditions. Among various transcription factors, one of the most important transcription factors is heat shock factor 1.⁹ Under normal conditions heat shock proteins bind with HSF1 and act as its repressor and in addition, maintaining its nonactive and monomeric confirmation.¹⁰ Under stress conditions, these heat shock proteins start to bind with damaged and denatured proteins thus releasing the HSF1, which undergoes homotrimerization (active form) and now HSF1 is able to translocate to the nucleus for DNA binding to up regulate the gene expression of heat shock proteins. These interactions between HSF1 and Cct5 have been extensively analyzed in the current study using molecular docking studies.

Cattle, being an indispensable livestock of India is reared in most of the areas bearing hot climatic conditions and in response to such extreme conditions, these animals express many heat shock proteins as a part of general adaptation phenomenon. Considering the fact, present study attempts to describe the 3D model of chaperonin containing TCP-1 subunit 5 of *Bos taurus* which provides a rational opportunity to obtain a reasonable 3D model in the absence of an experimentally determined crystal structure. The study also focuses on the physiochemical properties of HSP 60 protein and crucial interactions observed in *Bos taurus*.

Materials and methods

Categorization of Hsp60 Proteins

All the 20 available hits for Hsp60 proteins from HSPiR database¹¹ were compiled and tabulated on the basis of different selection criterions. These include the type of chaperones, absence/presence of domain, motif and signal peptide, their sub cellular location, their involvement in the biological processes and their molecular function.

Retrieval and Characterization of target sequence

The 541 amino acid long sequence of Cct5 with Accession number (NP_001029767) was obtained from the NCBI protein reference database. The physiochemical properties of the Cct5 protein were then predicted. The transmembrane helices of Cct5 were also obtained by TMHMM 2.0,¹² while the molecular weight, total number of negatively (Asp+Glu) and positively charged residues (Arg+Lys), grand average of hydropathicity (GRAVY), aliphatic index, isoelectric point, instability index and extinction coefficient of Cct5 were computed using ExPASy's ProtParam proteomics server.¹³

Comparative sequence analysis for Cct5 protein

The homologous protein sequences of Cct5 from 10 species i.e. B.T (*Bos taurus*), R.N (*Rattus norvegicus*), M.M (*Mus musculus*), H.S (*Homo sapiens*), B.B (*Bubalus bubalis*), O.A (*Ovis aries*), C.H (*Capra hircus*), S.S (*Sus scrofa*), E.C (*Equus caballus*), C.F (*Camelus ferus*) were retrieved from NCBI protein reference database. The comparative analysis for the protein was carried out using MAFFT multiple sequence alignment tools,¹⁴ which revealed important variations in *Bos taurus* Cct5 protein.

Model quality assessment

The information on beta strands and alpha helix constituting the secondary structure of Cct5 was retrieved via Phyre2 server, this server is based on Hidden Markov model.¹⁵ The protein sequence of Cct5 deduced the secondary structure and domain composition. The proteins interact with one another and perform various cellular functions in tertiary form. The tertiary structure for Cct5 protein in *Bos taurus* is unknown till date so an attempt was made to generate a homology model for the same in the study. The template (PDB ID- 3iygE) was used for comparative model building of Cct5 in *Bos taurus* using Phyre2 server. The model after being constructed using Phyre2 server was subjected to refinement and energy minimization via ModRefiner program. Energy minimization is an important aspect, as the comparison, determination and the evaluation of structure can be done by the criteria of energy minimization, even if there is unavailability of experimental structure.¹⁶ The accuracy and stereo chemical quality of predicted model was observed from RAMPAGE, QMEAN value and verify 3D score. Additionally, the overall G-factor was estimated using PROCHECK program from SAVES server.¹⁷ Finally, the protein was visualized in Swiss-PDB viewer.¹⁸

Elucidation of interacting partners of Hsp 60 and docking studies

For determining the vital interactions demonstrated by Hsp60 in *Bos taurus* for performing various cellular activities, STRING database¹⁹ was utilized and information on protein-protein interactions was deduced. In the present study, efforts are made to investigate the possible mode of interaction between HSF1 and CCT5. For the same, HSF1 was docked within the CCT5 homology model using Clus Pro 2.0 docking server where CCT5 was used as a receptor and HSF1 was used as a ligand. Clus Pro 2.0 is an automated server based on algorithm which can discriminate thousands of confirmations (2000 confirmations) of two proteins on the basis of different electrostatic and desolvation potential. The latter can be further discriminated via clustering and finally give the best fit structure between two proteins which was generally found closest to native structure from x-ray crystallography results.²⁰

Results and discussion

Since, CCT5 protein form an integral component of CCT which are involved in folding of denatured proteins under stress conditions in *Bos taurus* therefore a comprehensive *in silico* structure elucidation & stereo chemical properties evaluation of Cct5 protein was performed. In addition, classification & protein-protein interaction of Hsp 60 protein was studied. Table 1 depicts all the reported Hsp 60 proteins along with their types, domains, cellular locations and functions. The majority of Hsp60 proteins belongs to Group II Chaperonin class and only six from Group I Chaperonin class and is implicated mainly in ATP binding.

Table 1 HSP 60 proteins along with their types, domains, cellular locations and functions

S. no	Heat shock proteins	Type of chaperone	Domain	Motif	Signal peptide	Sub cellular location	Biological process	Molecular function
1	HSP60_I329	Group I chaperonin	-	-	Yes	Mitochondrion	Protein Folding	ATP binding
2	HSP60_I330	Group I chaperonin	-	-	Yes	Mitochondrion	Cellular metabolic process	ATP binding
3	HSP60_I331	Group I chaperonin	-	-	No	Nucleus	Cellular metabolic process	ATP binding
4	HSP60_I332	Group I chaperonin	-	-	No	Cytoplasm	Cellular metabolic process	ATP binding
5	HSP60_I333, I335, I336 and I337	Group II chaperonin	Chaperonins TCP	-	No	Cytoplasm	Protein Folding	ATP binding
9	HSP60_I338	Group II chaperonin	Chaperonins TCP	-	No	Melanosomes	Protein Folding	ATP binding
10	HSP60_I339	Group II chaperonin	Chaperonins TCP	-	No	Cytoplasm	Protein Folding	ATP binding
11	HSP60_I340	Group II chaperonin	Chaperonins TCP	-	No	Cytoplasm	-	-
12	HSP60_I341, I342, I343, I345	Group II chaperonin	Chaperonins TCP	-	No	Cytoplasm	Protein Folding	ATP binding
16	HSP60_I346	Group II chaperonin	Chaperonins TCP	-	No	Microtubule	Protein Folding	ATP binding
17	HSP60_I347	Group II chaperonin	-	-	No	Cytoplasm	Protein Folding	ATP binding
18	HSP60_I348	Group I chaperonin	-	-	No	Cytoplasm	-	-
19	HSP60_I349	Group II chaperonin	DEP domain, Phosphatidylinositol, phosphate kinase, (PIPK) domain	Five finger FYVE/ FYVE	No	Nucleus	Cellular metabolic process	Phosphatase Kinase activity
20	HSP60_I350	Group I chaperonin	-	-	No	Mitochondrial matrix	Protein Folding	ATP binding

Comparative analysis and characterization of target sequence

The multiple sequence alignment among the 5 species has been presented in Figure 1 where most of the sequence was found to be conserved with minor variations in context to *B.T* (*Bos taurus*), *R.N* (*Rattus norvegicus*), *M.M* (*Mus musculus*), *H.S* (*Homo sapiens*), *B.B* (*Bubalus bubalis*), *O.A* (*Ovis aries*), *C.H* (*Capra hircus*), *S.S* (*Sus scrofa*), *E.C* (*Equus caballus*), *C.F* (*Camelus ferus*). The observed variations in CCT5 have been highlighted in red color in the alignment. The primary structure of CCT5 as estimated from ExPASy's ProtParam server showed that the subunit has a total of 541 amino acids with molecular weight of 59615.0. Total number of negatively charged residues (Asp+Glu) in the sequence is 81, whereas 69 positively charged residues (Arg+Lys) were found. The grand average of hydropathicity (GRAVY) is -0.175 which is calculated by summing up of all the hydropathy values of amino acids and then dividing them by number of residues in the whole protein. Thus, the lower value of GRAVY indicates the better interaction of CCT5 protein with water

i.e., more hydrophilic nature.²¹ Interactions in solution and also help in calculating the concentration of a particular protein. TMHMM server reported that no region of CCT5 located in transmembrane (Figure 2).

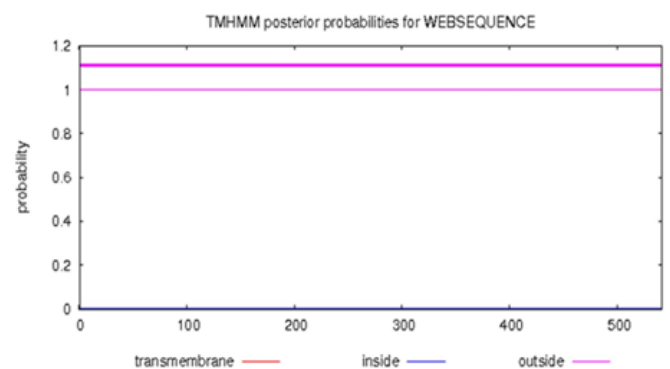


Figure 2 TMHMM server prediction for the region of CCT5 located in transmembrane.

Determination of CCT5 tertiary structure

The protein sequence of CCT5 was used to deduce the secondary structure and disorder prediction using Phyre2 program. The results showed the alpha helix region dominating other secondary structure (50%) followed by beta strand (12%) and disorder region is found to be very less (10%) in the protein. The high percentage of alpha helix region in protein indicates the stability and conservation of protein structure.

The three dimensional (3D) model for CCT5 protein (Figure 3) was generated by Phyre2 server and the model refinement and energy minimization was done by ModRefiner program that utilizes an algorithm for atomic level to develop and refine protein structures, critically based on a two-step, atomic-level energy minimization.²² Ramachandran plot of model as illustrated in Figure 4 was analyzed by RAMPAGE server and overall G factor from PROCHECK. If the three dimensional profile score S for the amino acid sequence is high only then, the 3D protein model is considered correct¹⁶ and Verify 3D analysis of best model revealed that 80.96% of the residues had an average 3D-1D score <0.2, indicating that the model is compatible with its sequence (Figure 5). Overall G-factor was found to be -0.41 with 94.4% residues were present in favored region of Ramachandran plot. The numbers of residues in allowed region were 4.6%. None of the residues was present in the disallowed region of plot as, per the acceptance criteria that the quality of model is reliable and good if the 90% of residues fall under allowed region. Qualitative Model Energy Analysis termed as QMEAN analysis^{23,24} is a composite scoring function describing the major geometrical aspects of protein structures. Our study predicts this model has comparable qualities to experimental structures, its value being 0.562, as for a good model QMEAN value lies between 0-1 (Figure 6).

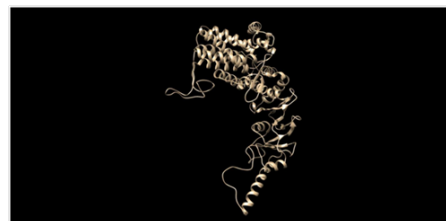


Figure 3 The 3D model for CCT5 protein (Phyre2 server).

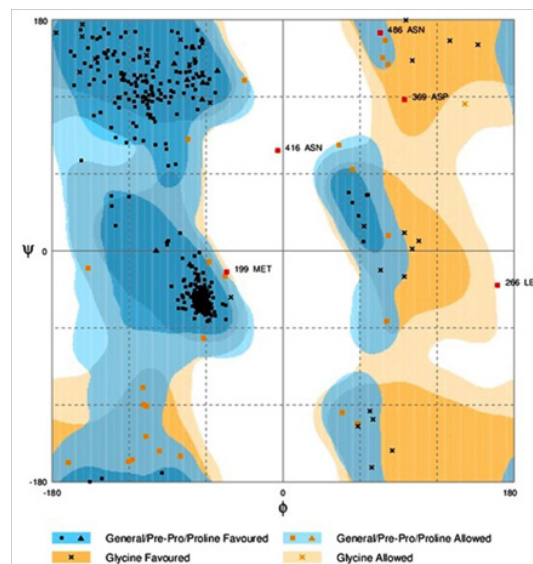


Figure 4 Ramachandran plot of CCT5 model generated (RAMPAGE and PROCHECK).

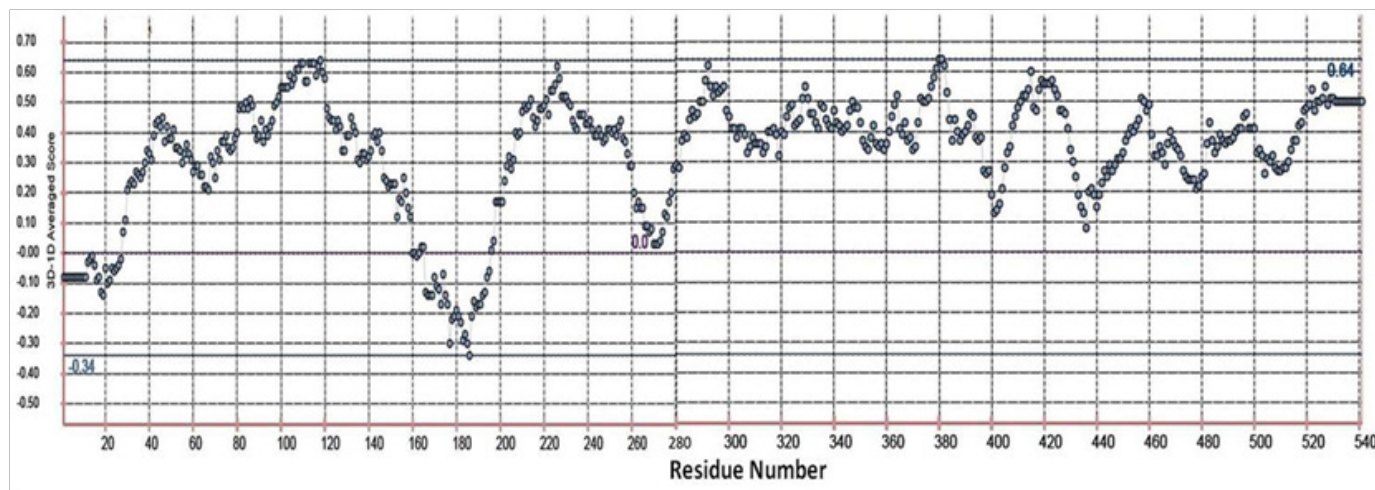


Figure 5 Verifying the 3D model by scoring the residues for 3D-1D score.

Protein- protein interactions and docking analysis

A single protein may not perform all the tasks individually therefore, these proteins interact with one another and form complexes that may initiate or regulate numerous biological processes. The HSP 60 protein interacts with the members of heat shock proteins and often other proteins to guide the folding process. These interactions are important to identify for extensive understanding of intricate biological mechanisms and pathways. The HSP 60 protein is found to be interacting with many proteins such as HSPA9, HSP90AA1,

HSP90AB1, HSPCB, TLR4, HSPA1A, HSPA8, HSPE1 and HSPA4 respectively (Figure 7). These proteins are mainly molecular chaperones and are involved in cell-cycle regulation and signal transduction (Table 2). The Figure 7 represents the evidence view from STRING database where the edges depict the type of available evidence for the interaction i.e. experimental, text-mining, co-expression, co-occurrence, gene fusion, and neighborhood or through database. Whereas, Figure 7 shows the confidence view where the thickness of the edge represents the confidence in interaction.

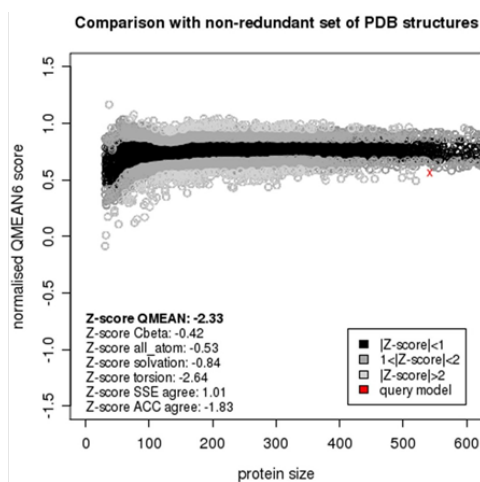


Figure 6 QMEAN score for the CCT5 3D model, a score being 0.562 indicating it as a good model.

For docking, first the model for HSF1 protein was modeled and then subjected to energy minimization. Then the modeled HSF1 protein was docked with CCT5 homology model using Clus Pro 2.0 docking environment. The molecular docking was carried out at full rotation and allowing ligand for movement while keeping receptor position fix in docked environment. Our goal of docking CCT5 and HSF1 was to find the docked confirmations which possess good surface complimentarity and this can be achieved by selecting those complexes which have lowest desolvation and electrostatic energies. Out of 2000 confirmations, top ten result was shown by Cluspro and the first model was found to possess lowest electrostatic and desolvation potential values (Lowest energy=-860.3, Electrostatic=-921.7, Hydrophobic= -1102.5, Vander waals interaction value = -278.7). The residues N, T, K, P, N, G, L, D, M, V, T, V, A, T, T, K, H, L, D, V, T, D, F, L, M, E, G, M and N of Cct5 with positions 46-47, 49, 54-58, 60, 62, 69, 70, 169, 172, 262-264, 266-269, 273-274, 277, 440, 464, 467-469 respectively were found to actively participate in docking. The lowest energy complex of HSF1 and CCT5 was finally viewed by PyMOL (Figure 8).

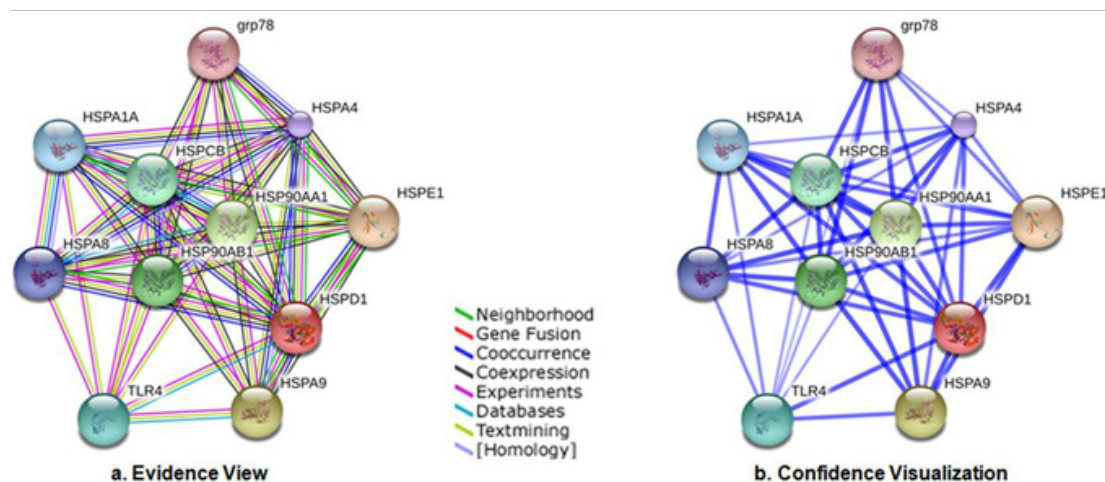


Figure 7 (a) STRING database evidences depicting various interactions (experimental, text-mining, co-expression, co-occurrence, gene fusion, and neighborhood) feasible with the predicted model. (b) STRING database representation of confidence view where the thickness of the edge represents the confidence in interaction.

Table 2 HSP 60 interacting proteins during cell-cycle regulation and signal transduction

Interacting protein	Protein size (Aa chain length)	Protein details	Combined score
HSPEI	102	10kDa heat shock protein, mitochondrial; Eukaryotic CPN10 homolog which is essential for mitochondrial protein biogenesis, together with CPN60. Binds to CPN60 in the presence of Mg-ATP and suppresses the ATPase activity of the latter	0.999
HSPA9	679	Stress-70 protein, mitochondrial; Implicated in the control of cell proliferation and cellular aging	0.988
HSP90AAI	733	Heat shock protein HSP 90-alpha; Molecular chaperone that promotes the maturation, structural maintenance and proper regulation of specific target proteins involved for instance in cell cycle control and signal transduction. Undergoes a functional cycle that is linked to its ATPase activity. This cycle probably induces conformational changes in the client proteins, thereby causing their activation. Interacts dynamically with various co-chaperones that modulate its substrate recognition, ATPase cycle and chaperone function	0.946
HSP90ABI	724	Heat shock protein HSP 90-beta	0.941
HSPCB	719	Heat shock protein HSP 90-beta; Molecular chaperone that promotes maturation, structural maintenance and proper regulation of specific target proteins involved for instance in cell cycle control and signal transduction. Undergoes a functional cycle that is linked to its ATPase activity. This cycle probably induces conformational changes in the client proteins, thereby causing their activation. Interacts dynamically with various co-chaperones that modulate its substrate recognition, ATPase cycle and chaperone function	0.94

Table Continued....

Interacting protein	Protein size (Aa chain length)	Protein details	Combined score
TLR4	841	Toll-like receptor 4 precursor; Cooperates with LY96 and CD14 to mediate the innate immune response to bacterial lipopolysaccharide (LPS). Acts via MYD88, TIRAP and TRAF6, leading to NF-kappa-B activation, cytokine secretion and the inflammatory response	0.924
HSPA1A	641	Heat shock 70kDa protein 1A; In cooperation with other chaperones, Hsp70s stabilize pre-existent proteins against aggregation and mediate the folding of newly translated polypeptides in the cytosol as well as within organelles. These chaperones participate in all these processes through their ability to recognize non-native conformations of other proteins. They bind extended peptide segments with a net hydrophobic character exposed by polypeptides during translation and membrane translocation, or following stress-induced damage	0.913
HSPA8	650	Heat shock cognate 71kDa protein; Acts as a repressor of transcriptional activation. Inhibits the transcriptional coactivator activity of CITED1 on Smad-mediated transcription. Chaperone. Component of the PRP19-CDC5L complex that forms an integral part of the spliceosome and is required for activating pre-mRNA splicing. May have a scaffolding role in the spliceosome assembly as it contacts all other components of the core complex	0.894
HSPA4	840	Heat shock 70kDa protein 4	0.869
Grb78	600	Uncharacterized protein; Probably plays a role in facilitating the assembly of multimeric protein complexes inside the ER	0.866



Figure 8 The lowest energy complex between HSF1 and Cct5 predicted from PYMOL.

Conclusion

To Study the interacting partners of HSP60 in *Bos taurus* and interaction studies of CCT5 with HSF1 bioinformatics web tools were used. Secondary structure prediction, comparative modeling, physiochemical properties, protein-protein interaction studies were done and predicted model evaluated with PROCHECK, VERIFY3D and RAMPAGE for its accuracy. The comparative analysis of the protein revealed certain significant variations in *Bos taurus* as compare to *Mus musculus*, *Rattus norvegicus*, *Homo sapiens*, *Bubalus bubalis*, *Ovis aries*, *Capra hircus*, *Sus scrofa*, *Equus caballus* and *Camelus ferus*. At the end we found several numbers of amino acid residues of CCT5 that are involved in interaction with HSF1.

Acknowledgements

Authors are thankful to Indian Council of Agriculture Research-National Bureau of Animal Genetic Resources, Karnal, India for providing all the necessary facilities for carrying out the work.

Conflict of interest

The author declares no conflict of interest.

References

- Ellis RJ. Molecular chaperones: avoiding the crowd. *Curr Biol*. 1997;7(9):531–533.
- Thulasiraman V, Yang CF, Frydman J. *In vivo* newly translated polypeptides are sequestered in a protected folding environment. *EMBO J*. 1999;18(1):85–95.
- Ditzel L, Löwe J, Stock D, et al. Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. *Cell*. 1998;93(1):125–138.
- Kim S, Willison KR, Horwich AL. Cytosolic chaperonin subunits have a conserved ATPase domain but diverged polypeptide-binding domains. *Trends in Biochemical Sciences*. 1994;19(12):543–548.
- Spieß C, Meyer AS, Reissmann S, et al. Mechanism of the eukaryotic chaperonin: protein folding in the chamber of secrets. *Trends Cell Biol*. 2004;14(11):598–604.
- Llorca O, McCormack EA, Hynes G, et al. Eukaryotic type II chaperonin CCT interacts with actin through specific subunits. *Nature*. 1999;402(6762):693–696.
- Hynes GM, Willison KR. Individual subunits of the eukaryotic cytosolic chaperonin mediate interactions with binding sites located on subdomains of beta-actin. *J Biol Chem*. 2000;275(25):18985–18994.
- Llorca O, Martín-Benito J, Ritco-Vonsovici M, et al. Eukaryotic chaperonin CCT stabilizes actin and tubulin folding intermediates in open quasi-native conformations. *EMBO J*. 2000;19:5971–5979.
- Christians ES, Yan LJ, Benjamin IJ. Heat shock factor 1 and heat shock proteins: critical partners in protection against acute cell injury. *Crit Care Med*. 2002;30:S43–S50.
- Baler R, Dahl G, Voellmy R. Activation of human heat shock genes is accompanied by oligomerization, modification, and rapid translocation of heat shock transcription factor HSF1. *Mol Cell Biol*. 1993;13(4):2486–2496.
- Ratheesh Kumar R, Nagarajan NS, Arunraj SP, et al. HSPiR: a manually annotated heat shock protein information resource. *Bioinformatics*. 2012;28(21):2853–5.
- Krogh A, Larsson B, von Heijne G, et al. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 2001;305(3):567–80.

13. Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy Server. In: John M Walker, editor. *The Proteomics Protocols Handbook*. Humana Press; 2005. p. 571–607.
14. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23(21):2947–2948.
15. Kelley LA, Mezulis S, Yates CM, et al. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015;10:845–58.
16. Lüthy R, Bowie JU, Eisenberg D. Assessment of protein models with three dimensional profiles. *Nature*. 1992;356:83–85.
17. Laskowski RA, MacArthur MW, Moss DS, et al. PROCHECK – a program to check the stereochemical quality of protein structures. *Journal of Applied crystallography*. 1993;26:283–291.
18. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modelling. *Electrophoresis*. 1997;18(15):27114–2723.
19. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43:D447–D452.
20. Kozakov D, Beglov D, Bohnuud T, et al. How good is automated protein docking? *Proteins: Structure, Function and Bioinformatics*. 2013;81(12):2159–2166.
21. Prabhavati MK, Ashokkumar NG, Saradha Devi KM. Homology modeling and structure prediction of thioredoxin (TRX) protein in wheat (*Triticum aestivum* L.). *International Journal of Biosciences*. 2011;1:20–32.
22. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys J*. 2011;101(10):2525–2534.
23. Benkert P, Tosatto SC, Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins*. 2008;71(1):261–277.
24. Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. *Nucleic Acids Res*. 2009;37:W510–W514.