International Journal of Bioinformatics Research
ISSN: 0975-3087 & E-ISSN: 0975-9115, Volume 7, Issue 1, 2015, pp.-313-320.
Available online at http://www.bioinfopublication.org/jouarchive.php?opt=&jouid=BIJ0000001

IDENTIFICATION OF DISTANT STRUCTURAL ORTHOLOG AND A POSSIBLE EVOLUTIONARY LINKAGE OF HSP60 - A FOLD BASED APPROACH

PADMADAS N.[1]*, CHELLASAMY S.[1] AND DURAIRAJ S.[2]

[1]School of Biotechnology and Bioinformatics, D.Y. Patil University, Navi Mumbai - 400 614, MS, India.
[2]Department of Advanced Zoology & Biotechnology, Loyola College, Chennai - 600 034, TN, India.
*Corresponding Author: Email- naveenpadmadas@gmail.com

Received: March 14, 2015; Revised: April 28, 2015; Accepted: May 04, 2015

Abstract- Correct folding is imperative for a protein to perform its function. Failing to fold correctly can result in a misfolded or inactive protein. Along with several other factors, molecular chaperones play an integral role in folding. Heat shock proteins (HSPs) are a specialized group of chaperone proteins synthesized in all living organisms in response to stress. Among different Hsps, Hsp60 forms the most conserved chaperone present in eukaryotes and eubacteria. Homology is one of the most important concepts in Evolutionary biology and proteomics. Identifying a distant structural ortholog designates a connection of common descent between entities. It will be great evolutionary significance to explore the distant homologs of this protein. Different biological databases were searched. With no reported valid structural ortholog, Fold based method was used. PGenTHREADER was intensively used to identify different templates sequentially. Hsp60 structure was modeled and validated using various servers. Phylogenetic analysis using ClustalW2 and Mega 5 was carried out to find the most relevant distant homologs from the PGenTHREADER templates. Sub structural comparative study was carried out between Hsp60 structure and the newly found template using UCSF-CHIMERA. The first equatorial, first intermediate and second equatorial regions of Hsp60 and Signal Recognition Particle 54kDa (SRP54) were observed to be much conserved. Some deletions have occurred in the apical and second equatorial region. This study sheds light on the highly conserved nature of the chaperonin which has not got much diversified in the course of evolution and a possible linkage with SRP54 in the light of evolution. Also, the conserved nature of Hsp60 is very much evident from their highly limited homologs reported in databases.

Keywords- Hsp60, Chaperonin, PGenThreader, Heat shock protein, ortholog, SRP54, evolution, ClustalW2, UCSF-Chimera, Mega 5

Citation: Padmadas N., Chellasamy S. and Durairaj S. (2015) Identification of Distant Structural Ortholog and a Possible Evolutionary Linkage of HSP60 - a Fold Based Approach. International Journal of Bioinformatics Research, ISSN: 0975-3087 & E-ISSN: 0975-9115, Volume 7, Issue 1, pp.-313-320.

Introduction

Protein folding is a physical process by which a polypeptide folds into its characteristic and functional three-dimensional structure from random coil [1]. This self-assembly process does not require additional cellular factors. Although some parts of functional proteins may remain unfolded, the correct three-dimensional structure is essential for a protein to perform its function [2]. A protein, failing to fold into its native structure generally produces inactive proteins or misfolded proteins.

Recent studies have reported various factors that are involved for proper folding of proteins. A protein macromolecule folds spontaneously during or after biosynthesis. This process of folding is dependent on the solvent- whether it is water or lipid bilayer, the presence of cofactors, salts concentration, the pH, the temperature, and the role of molecular chaperones [3,4]. Of these, Molecular chaperones play a critical role. These proteins assist the non-covalent folding or unfolding and the assembly or disassembly of other macromolecular structures. However, once folding is complete or before completion itself in some cases, the chaperone leaves the current protein molecule and moves to support another protein folding. Among various functions performed by molecular chaperones, one of the most important roles is to prevent the assembled subunits and newly synthesized polypeptide chains from forming into misfolded structures, resulting in non-functional protein. Most of the chaperones are reported to be Heat shock proteins (Hsp) and several of these chaperones play an important role for proper folding of some of the proteins [5].

Hsp are a group of proteins. It is synthesized when an organism is exposed to various conditions of stress, including elevated temperatures. Moreover, they are critical for cell survival both constitutively and in times of stress to ensure proper folding of non-native states of proteins [6,7]. Based on their nature of functions and molecular mass, HSPs are broadly classified into six major families, namely, Hsp40 (J-proteins), Hsp60 (chaperonins), Hsp70, Hsp90, Hsp100 (Clp proteins) and small HSPs [8,9]. A complex molecular network of these chaperone families is formed and this complex maintains the internal stability of cellular protein [8].

Hsp60 family of chaperones or chaperonins in particular plays a

pivotal role in the cellular chaperone machinery. They are a class of most conserved and an all-pervading chaperone present in eukaryotes and eubacteria. Their presence is found in cytoplasm, mitochondria and plastids. Chaperonins participates in the formation of a complex network of chaperones and ensures proper folding of proteins which are newly synthesized and proteins which are denatured by stress [6,8,10]. These specialized folding machines are required to assist the primary sequence of several polypeptides in the cell as they consist of several domains with α/β fold.

The domain structure of the chaperonin are classified into Equatorial domain harboring the ATP binding site, a substrate binding Apical domain and a middle intermediate domain which communicates between the apical and equatorial domains. Thus this protein consists of 14 subunits arranged as two stacked heptameric ring complex. This double ring structure forms a large central cavity to which the unfolded proteins bind by hydrophobic interactions. The constituent subunit of the heptamer contains the domain structure as shown in [Fig-1].

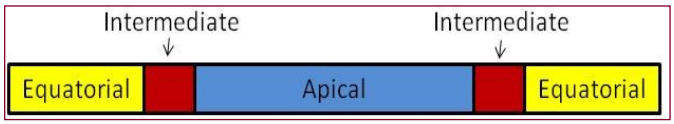

Fig. 1- Domain structure of Hsp60

The chaperonins can be grouped under two broader classifications-similar structures and quite diverse in sequence [11] and both the groups have shown two distinct evolutionary lines. Prokaryotes and endosymbiotic organelles such as mitochondria and chloroplasts made the place for group 1 and archaea and eukaryotic cytosol for group II chaperonins [8]. The chaperonins belonging to the Group I uses an attachable "lid" like structure (Hsp10) that goes and binds in a fashion like ATP- dependent while the chaperonins of group II uses a fixed protrusions like structure which may either be upward open in a protein-accepting state or marginate and then closes when the ring binds ATP to produce the encapsulated state of active folding [12,13].

The sequential homology, structure and function between Hsp60 and its prokaryotic homolog groEL demonstrates this as one among the most evolutionarily conserved proteins. The amino acid sequence of Hsp60 bears a similarity to its homologs in plants, bacteria and humans [14]. Folding and confirmation maintenance of approximately 15-30% of cellular proteins are aided by Hsp60 [15]. It plays an important role in the transport and maintenance of mitochondrial proteins as well as the transmission and replication of mitochondrial DNA [16]. Studies have also suggested that Hsp60 plays a key role in preventing apoptosis in the cytoplasm [17]. Recent investigations have even suggested a regulatory correlation between Hsp60 and a glycolytic enzyme [18].

The chaperonin has reported to show distant lineages with gram positive group of bacteria [19]. Identifying a distant homolog, probably a structural ortholog designates a connection of common descent between entities. The main objective behind this study was to identify a potential distant ortholog of this protein and the possibility of a linkage in the course of evolution.

Materials and Methods

The Hsp60 of *Homo sapiens* sequence with Accession number P10809 having 573 amino acids bases was retrieved from UniProt [20]. Structural Classification of Protein database [21] and Superfamily 1.75 was used to retrieve more information about the protein. A search in PSI BLAST [31] resulted the distant homologs of Hsp60 spanning across the kingdom of life. We were interested in finding out a distant structural ortholog of Hsp60 and hence we tried to identify the proteins which exhibit common structural folds by using the software pGenTHREADER [22]. The listed templates were considered for further sequential analysis based on their generated scores. The results of pGenTHREADER with high Net scores were aligned along with the query sequence using ClustalW2, [23] and further confirmed using MEGA 5 [24].

After sequential analysis, the human Hsp60 was further investigated to understand the structural aspects. With no reported human Hsp60 structure in Protein Data Bank [25], homology modeling was carried out. A BLAST [26] against PDB database was carried out using the Human Hsp60 as query sequence. The potential template was selected from the results after investigating their missing residues and resolution. The selected template with PDB id 1IOK was used in for homology modeling using SWISS-MODEL server using automatic mode [27,28].

The generated model was validated using different online and offline tools- Protein Structure Analysis (ProSA-web) tool [29,30] and Structural Analysis and Verification Server (SAVES- Ramachandran Plot) to understand the residual clashes [32]. The structure of the most closely related protein to human Hsp60, which was obtained as a result of phylogenetic analysis was downloaded from the PDB database. Both the structures were classified into first equatorial, first intermediate, apical, second intermediate and second equatorial regions using SwissPdbViewer (SPDBV) respectively [33]. Further Molecular graphics and analyses were performed with the UCSF Chimera package [34]. The number of alpha helix and Beta sheets in the respective structures were analyzed intensely for the similarity. Search was initiated in InterPro [35] and STRING database [36] for understanding the functional partners of the proteins.

Results

The selected query sequence of Homo sapiens was considered for sequential, structural and phylogenetic analysis during the course of this study. A search was initiated in superfamily 1.75 database using the keyword "1grl" which is a bacterial chaperonin. The class, fold, superfamily and family members of this protein were retrieved. The apical region of the structure has the following SCOP classification. Generally, this protein belongs to the class of alpha and beta protein (a/b). The fold is the swiveling beta/beta/alpha domain. Their superfamily is reported to be GroEL apical domain-like and their family is GroEL like chaperone, apical domain. The next domain was of equatorial region of the structure which has the following SCOP classification. The class has all alpha proteins. Fold has GroEL equatorial domain-like, superfamily has GroEL equatorial domain like and family has GroEL chaperone, ATPase domain. The intermediate region of the structure has the class of alpha and beta proteins (a+b). The fold is of GroEL-intermediate domain like. Superfamily, it belongs to GroEL. (http://supfam.org/SUPERFAMILY/cgi-bin/search.cgi?search_field=1grl) Chaperone proteins are less subjected to mutations. Thus their overall residual conservation throws light on their functional significance. More over we intended to identify potential distant structural ortholog which match very well with this protein, so fold based method was preferred over comparative modeling. To carry out the same - PGenThreader was considered. It gave the list of templates with the score like "exact", "high",

"medium", "low" and "guesses" with the score of p>= 000.1, p>00.1,P>0.1. The results under exact was less preferred as they all belonged to the chaperones. From the rest of the results, the high scored templates were considered over the medium, low and guess for further analysis because of their high Net score as shown in [Table-1].

Table 1- PGenThreader Results with Hsp60 query

Conf.	Net Score	P- value	PairE	SolvE	Aln Score	Aln Len	Str Len	Seq Len	CATH Classification
High	52.618	0.0002	-532.9	-4.4	81	357	416	573	3DM5
High	48.878	0.0006	-498.5	-11.6	58	281	382	573	3LKB
High	48.466	0.0006	-464.2	-6.9	80	253	308	573	1VLV
High	46.631	0.001	-469.9	-10.6	51	289	419	573	3DGG

The listed templates from pGenThreader were 3dm5 (Signal recognition particle 54 kDa protein- SRP54), 3lkb (Amino acid binding protein), 1vlv (Ornithine carbamoyl transferase) and 3dqq (Putative tRNA synthase).

The four PGenThreader sequences along with human Hsp60 sequences were converted to .fasta file using MEGA 5nd was further was uploaded for analysis using Clustalw2 tool- for generating a Phylogeny tree with default tree format parameters- Distance correction 'on', Exclude gaps 'off'' and Neighbour joining clustering method. The resulted guide tree is shown in [Fig-2].

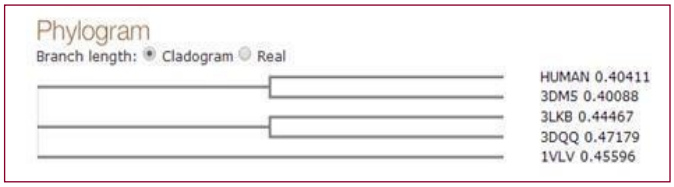

Fig. 2- Clustalw2 tree results of Human Hsp60 with PGenThreader resulted templates

The result was further confirmed using MEGA 5 phylogenetic analysis. The parameter chosen was Neighbour Joining method. The substitution model was Poisson model and the phylogeny test was carried using 500 Boot strap replications and a tree was generated as shown in [Fig-3].

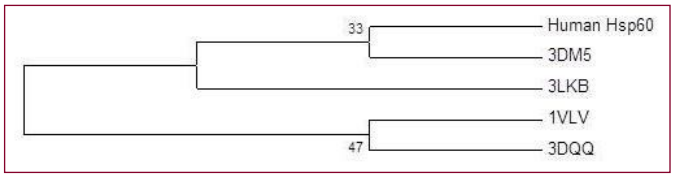

Fig. 3- MEGA 5 phylogenetic tree results of Human Hsp60 with PGenThreader resulted templates

Both the tools confirmed that human Hsp60 sequence share more close relationship with 3DM5, which is a Signal Recognition Particle 54kDa protein (SRP54). With no reported crystal structures of Hsp60 in protein Data bank, SwissModel server based model was considered for human Hsp60 structure generation [Fig-4].

The generated model was subjected to structure validation using PROSA and SAVES- Ramachandran plot. The PROSA z score of the model reported as -11.27 which was encouraging [Fig-5]. The plot of residue scores for Hsp60 shows the structure below 0 for window size 40. The PROCHECK Ramachandran Plot also showed 90.3% of residues in most favoured region, which categorise the generated structure under good quality model [Fig-6].

The modeled Hsp60 structure has five distinct regions- an equatorial followed by intermediate in the N- Terminal, then an apical region followed by an intermediate and equatorial at the C- Terminal. Because the listed templates were based on fold based methods and not on homology based method, a sub structure based comparison instead of whole structure comparison using CHIMERA with the listed templates was carried out. As we were keen to align the local regions instead of global substructures, the alignment algorithm selected was of Smith-Waterman. Same was followed for the selected structure of SRP54. Next, the substitution matrix was set to BLOSUM30. This was mainly due to partially conserved domains observed in these fold based method.

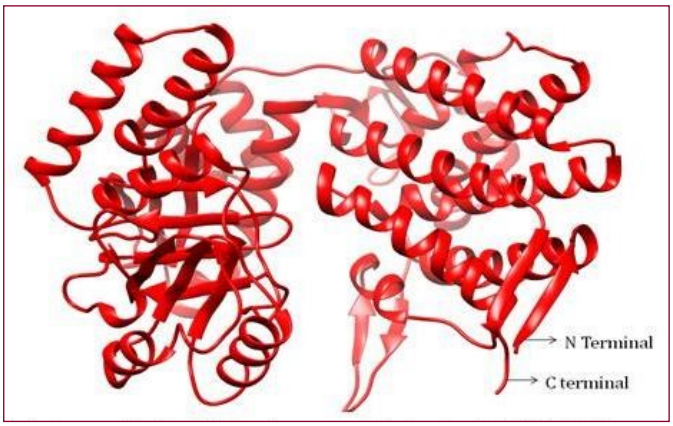

Fig. 4- Human Hsp60 full structure - modelled using SWISS-MODEL server

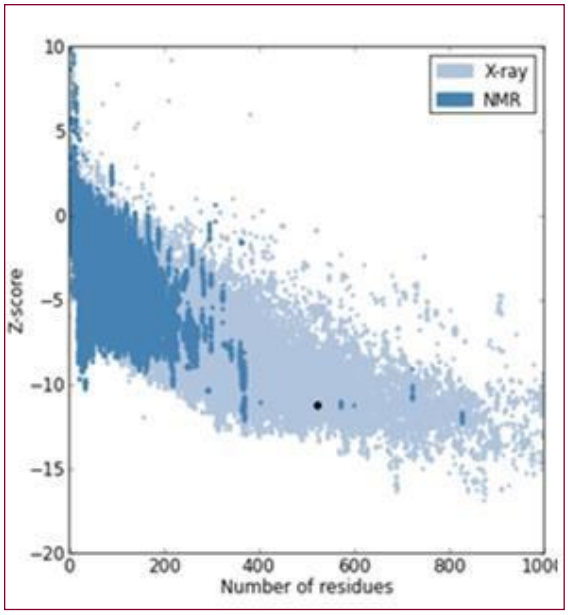

Fig. 5- PROSA Z score result

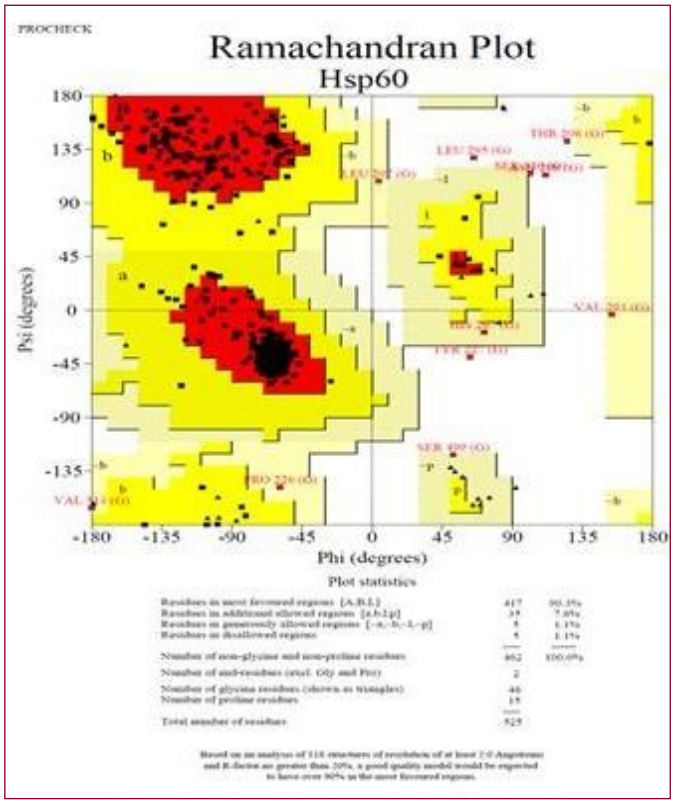

Fig. 6- PROCHEK Ramachandran Plot

Structural comparison was initiated with an aim to understand the level of secondary structural relatedness between the structures reported by PGenThreader with a highest score- SRP54 against the structure of Hsp60. To begin with, the first equatorial region, which is the ATPase binding region of Hsp60, has secondary structural elements like helices H1, H2, H3, H4 and H5 [Fig-7] which are reported to be conserved as helices H1, H2, H3 and H4 of SRP54 [Fig-8]. All the four helices of SRP54 found a match with the Hsp60 structure [Fig-9].

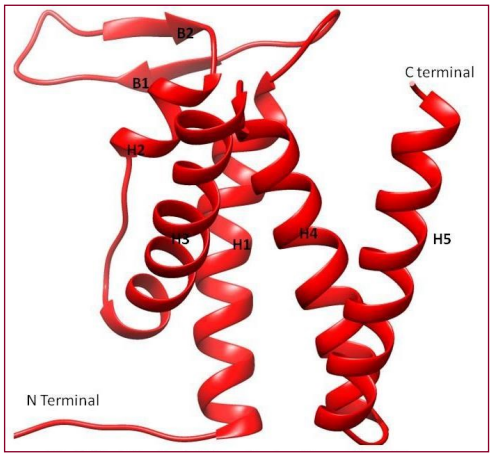

Fig. 7- First Equitorial domain structure of Hsp60

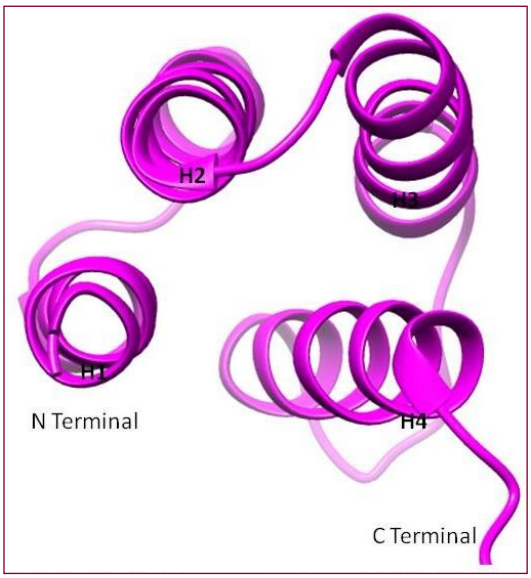

Fig. 8- First Equitorial domain structure of 3DM5 (SRP54)

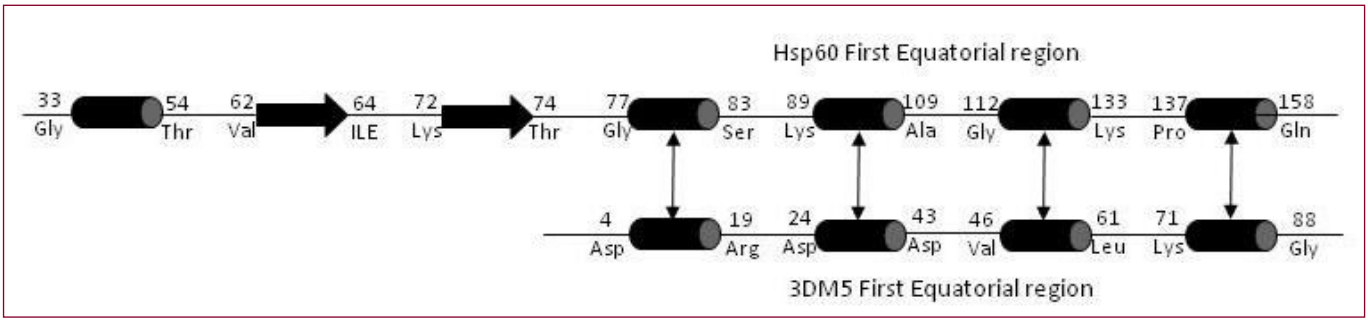

Fig. 9- Comparison of Hsp60 and 3DM5 (SRP54) Equitorial region

The first intermediate region of Hp60 has helix H6 and H7 [Fig-10] which again relates with SRP54 helix H5 and H6 [Fig-11]. Thus, this subsection is found to be well conserved between the Hsp60 and SRP54 [Fig-12]. Next, the apical region of Hsp60 [Fig-13] was compared against 3DM5 (SRP54) [Fig-14]. Here, five helices were observed in comparison to seven helices in 3DM5. Not much of conservation was observed in the apical region. A variation in the number of Beta strand was also observed [Fig-15].

In the second intermediate region there is report of a single helix region in Hsp60 [Fig-16] which matches with a single helix region of helix from 3DM5 (SRP54) [Fig-17], [Fig-18].

Next the second equatorial region shows four helices in Hsp60 namely H14, H15, H16, H17 [Fig-19] and 3DM5 (SRP54) has three of them H15, H17 and H18 [Fig-20]. A comparison diagram is shown in [Fig-21]. In the apical region of superimposed structure of 3dm5 (SRP54) against the modeled human Hsp60, we observed two well conserved parallel beta strands with an RMSD reported value of 1.93A^0 [Fig-22](a-c).

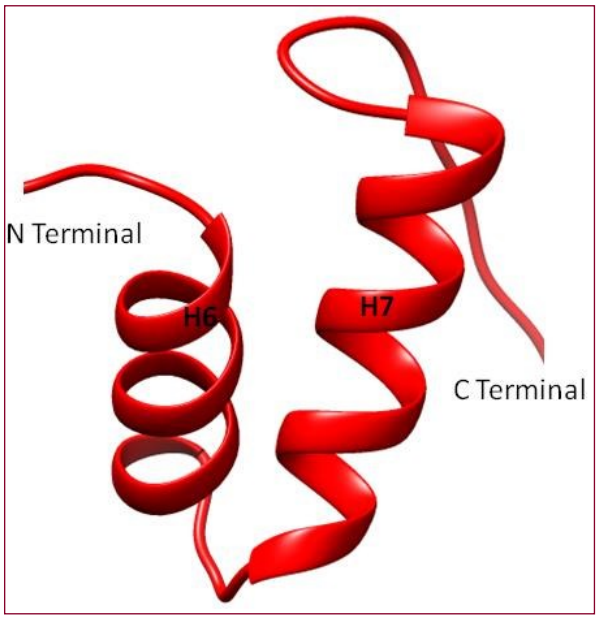

Fig. 10- First Intermediate domain structure of Hsp60

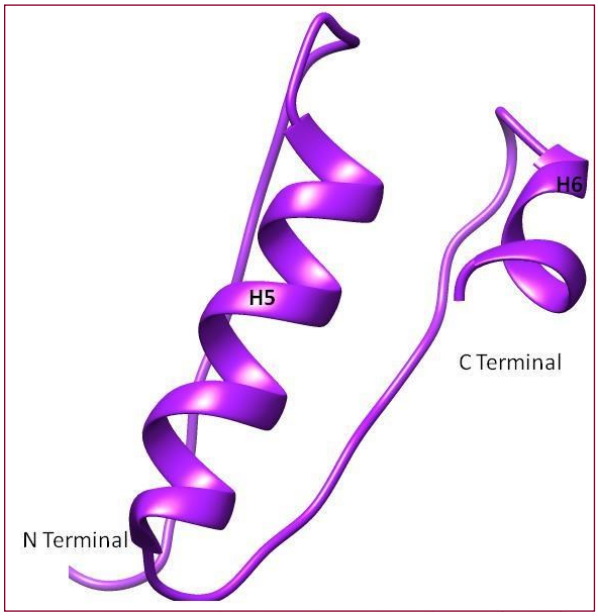

Fig. 11- First Intermediate domain structure of 3DM5 (SRP54)

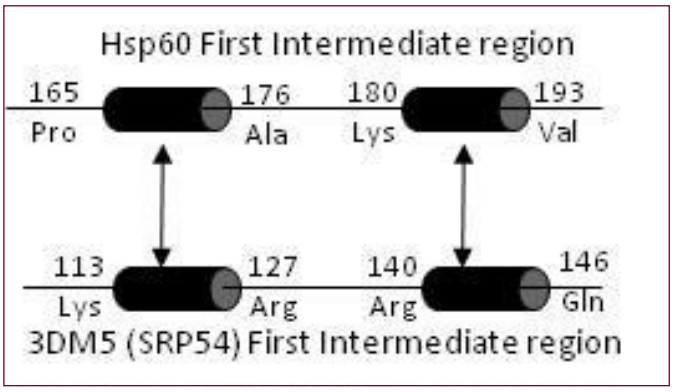

Fig. 12- Comparison of Hsp60 and 3DM5 (SRP54) Intermediate region

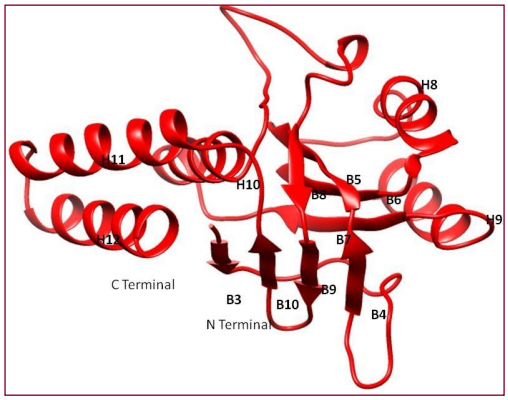

Fig. 13- Apical domain structure of Hsp60

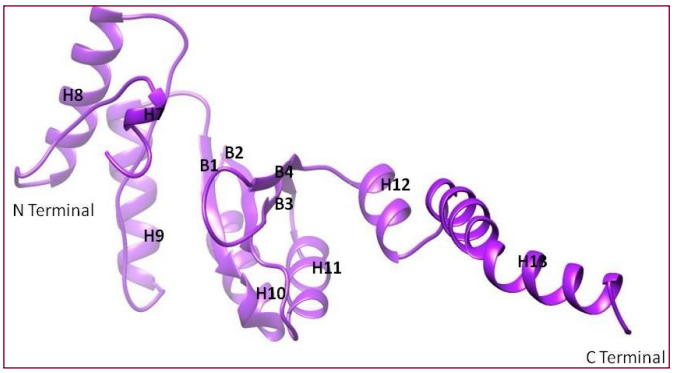

Fig. 14- Apical domain structure of 3DM5 (SRP54)

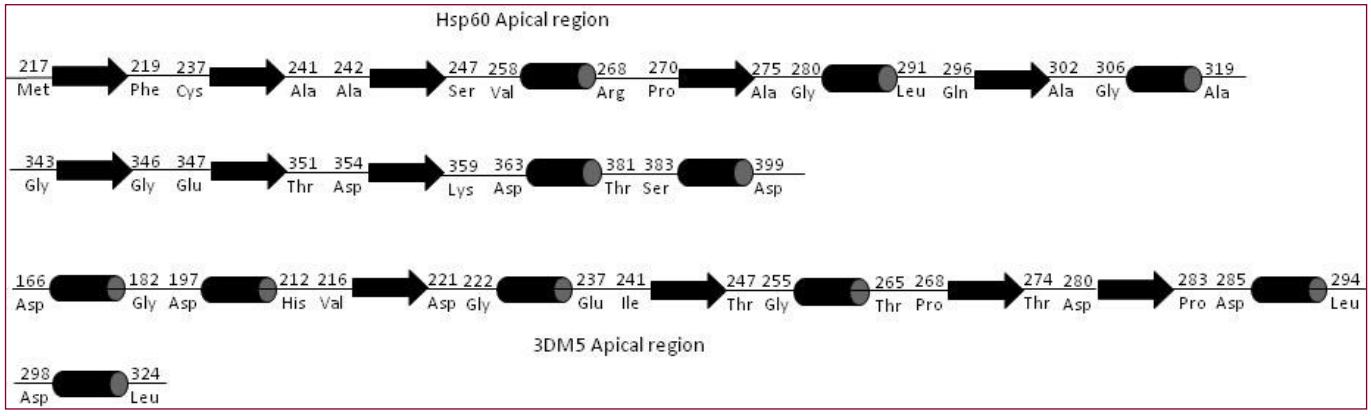

Fig. 15- Comparison of Hsp60 and 3DM5 (SRP54) Apical region

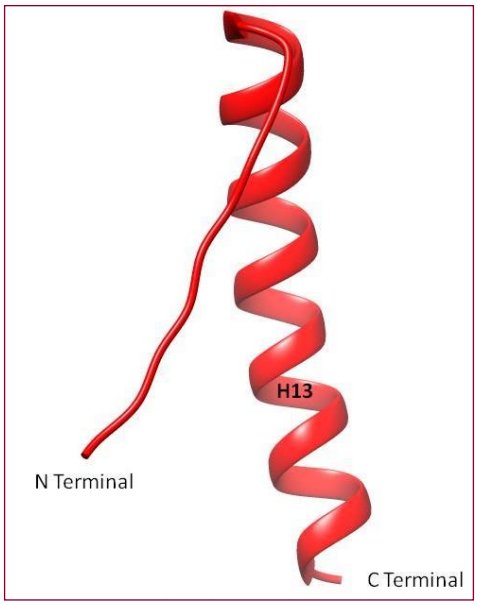

Fig. 16- Second Intermediate domain structure of Hsp60

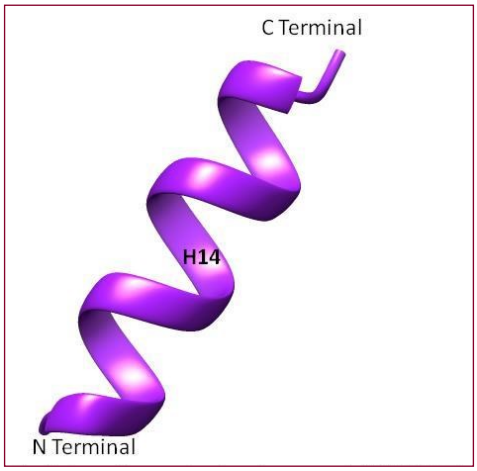

Fig. 17- Second Intermediate domain structure of 3DM5 (SRP54)

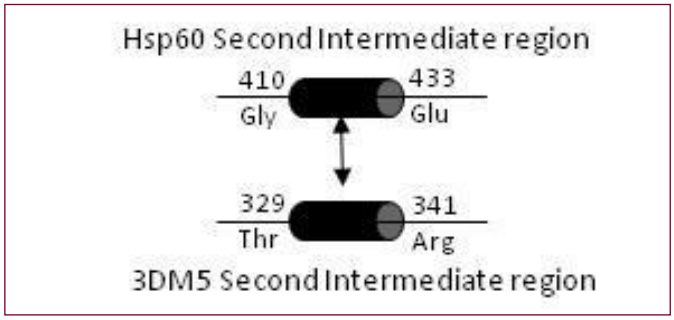

Fig. 18- Comparison of Hsp60 and 3DM5 (SRP54) second Intermediate region

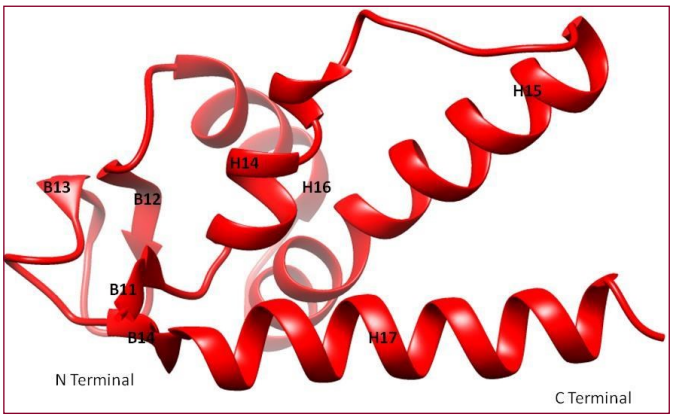

Fig. 19- Second Equitorial domain structure of Hsp60

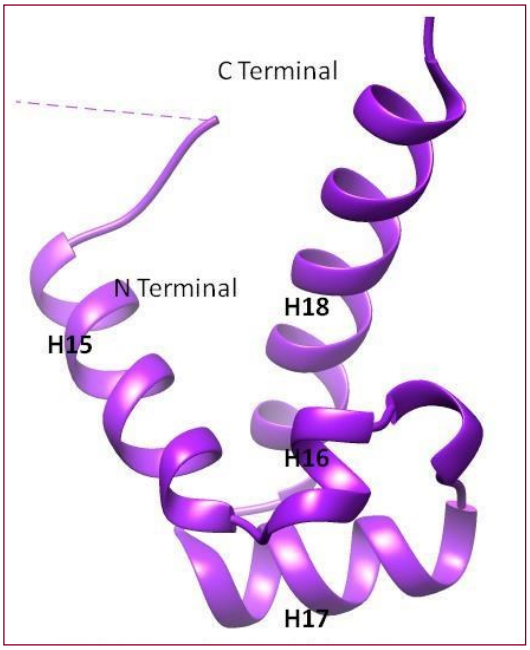

Fig. 20- Second Equitorial domain structure of 3DM5 (SRP54)

Discussion

As we observed reasonable structural conservation between these two distantly related proteins, a search in STRING database was initiated to find out the functional partners. To begin with, the bacterial sequence (*Pyrococcus furiosus*) of predicted Signal Recognition Particle 54 kDa was searched against STRING database. However, during this search, the listed hits were of orthologous sequences with no indication of chaperonine listed. So, we extended the search with predicted SRP54 of *Bos taurus* in string database based on literature survey. Bos *taurus was* sequentially relating themselves with the *Pyrococcus furiosus with an* identity score of 43.79 based on clustalw2.

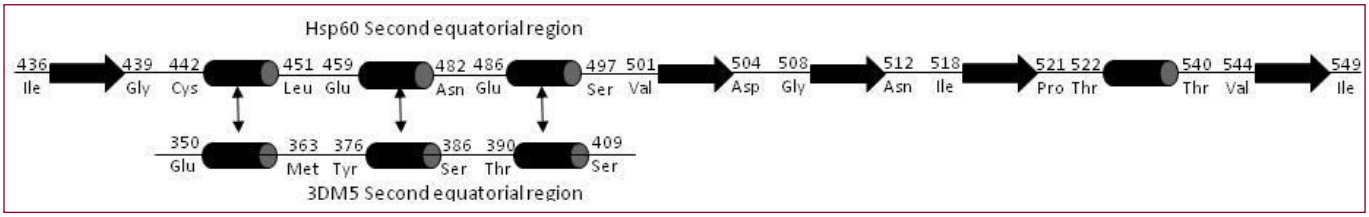

Fig. 21- Comparison of Hsp60 and 3DM5 (SRP54) second Intermediate region

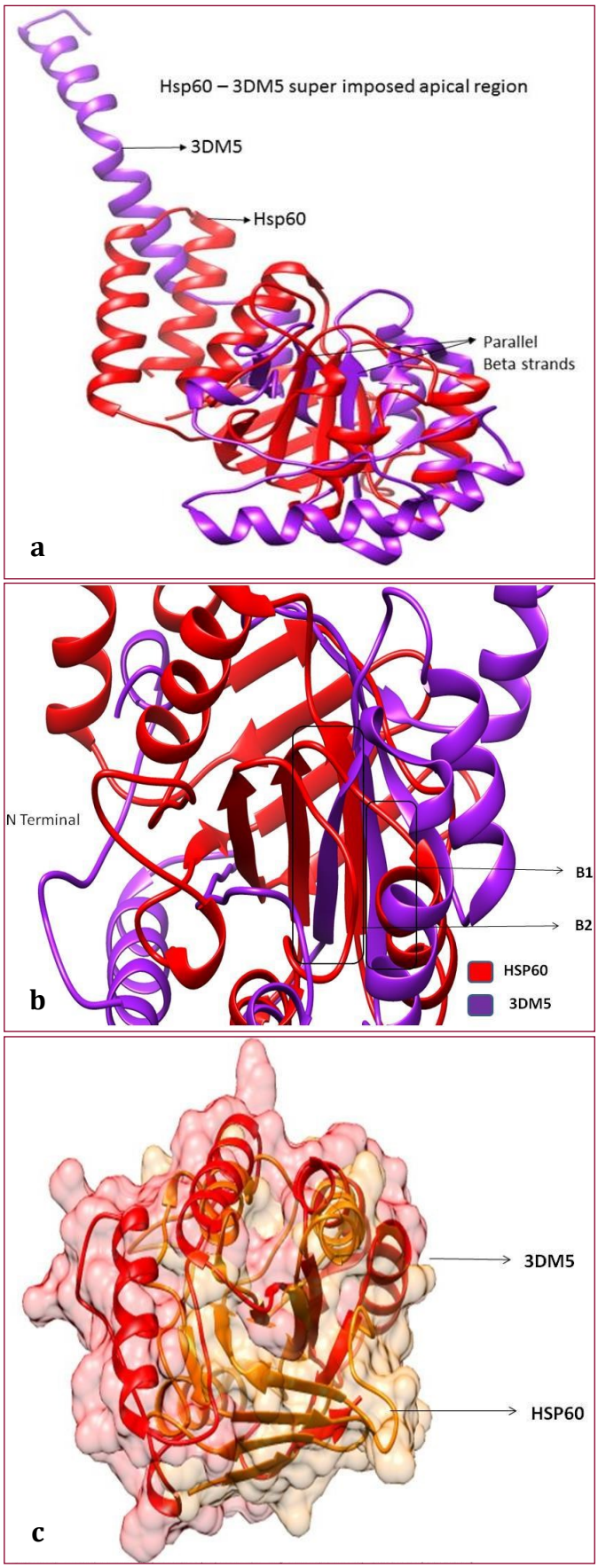

Fig. 22a- Superimposed apical domain of Hsp60 & 3DM5 (SRP54) **(a);** Hsp60 & 3DM5 (SRP54) **(b)**; Hsp60 & 3DM5 (SRP54) **(c).**

Interestingly, the Hsp60 chaperonin and predicted SRP54 was reported to be functional partners with a score of 0.770 and 0.751 respectively. Through this study, we could confirm that as such both these proteins are not exactly matching from structural perspective, but imperatively, there are considerable regions in the first equatorial , first intermediate, second intermediate and second equatorial regions of the protein . Apart from this the nucleotide binding sites were conserved in both these proteins. So a better understanding of this protein is much needed from docking, dynamics and protein-protein interactions.

Conclusion

Chaperonin is critically involved in protein folding and a search for their distant structural homologs could not give us better structures. But still there are many proteins reported to have the nucleotide binding sites intact. However, not all of them can be related to this protein. Thus only fold based method was able to give us both a sequential and structurally related ortholog structure in SRP54, which has the regions like equatorial and the intermediate regions. However, their functional relations are yet to be established. Thus this study sheds light on the highly conserved nature of the chaperonin which has not got much diversified during the course of evolution and a possible linkage between Hsp60 and SRP54 kDa in its origin state.

Acknowledgement: The authors wish to thank D Y Patil University-School of Biotechnology and Bioinformatics for providing lab infrastructure to carry out this research.

Conflict of Interest: The authors confirm that this article content has no conflicts of interest.

References

[1] Alberts B., Alexander J., Julian L., Martin R., Keith R. & Peter W. (2002) *The Shape and Structure of Proteins*, Molecular biology of the cell, Garland Science, New York and London.

[2] Berg J.M., Tymoczko J.L., Stryer L. & Clarke N.D. (2002) *Biochemistry*, W.H. Freeman and Company, San Francisco.

[3] Van den Berg B., Wain R., Dobson C.M. & Ellis R.J. (2000) *EMBO Journal*, 19(15), 3870-3875.

[4] Cooper G.M. (2000) *Protein Folding and Processing*, The Cell: A Molecular Approach, 2nd ed.

[5] Ellis R.J. & Vies S.M. (1991) *Annual Review of Biochemistry*, 60, 321-347.

[6] Bernd B., Jonathan W. & Arthur H. (2006) *Cell*, 125(3), 443-451.

[7] Parsell D.A. & Lindquist S. (1993) *Annu. Rev. Genet.*, 27, 437-496.

[8] Lund P. (2001) *Molecular Chaperones in the Cell*, Oxford University Press, Oxford.

[9] Kampinga H.H. & Craig E.A. (2010) *Nat. Rev. Mol. Cell Biol.*, 11, 579-592.

[10] Martin J., Horwich A.L. & Hartl F.U. (1992) *Science*, 258, 995-998.

[11] Horwich A.L., Fenton W.A., Chapman E. & Farr G.W. (2007) *Annu. Rev. Cell Dev. Biol.*, 23, 115-145.

[12]Stoldt V., Rademacher F., Kehren V., Ernst J.F., Pearce D.A. & Sherman F. (1996) *Yeast*, 12(6), 523-529.

[13]Christoph S., Anne M.S., Stefanie R. & Judith F. (2004) *Trends Cell Biol*., 14(11), 598-604.

[14]Johnson R.B., Fearon K., Mason T. & Jindal S. (1989) *Gene*, 84 (2), 295-302.

[15]Ranford J.C., Coates A.R. & Henderson B. (2000) *Expert Rev. Mol. Med*., 2(8), 1-17.

[16]Koll H., Guiard B., Rassow J., Ostermann J., Horwich A.L., Neupert W. & Hartl F.U. (1992) *Cell*, 68, 1163-1175.

[17]Itoh H., Komatsuda A., Ohtani H., Wakui H., Imai H., Sawada K., Otaka M., Ogura M., Suzuki A. & Hamada F. (2002) *Eur. J. Biochem*., 269(23), 5931-5938.

[18]Kaufman B.A., Kolesar J.E., Perlman P.S. & Butow R.A. (2003) *J. Cell Biol*., 163, 457-461.

[19]Gupta R.S. & Golding G.B. (1993) *J. Mol. Evol*., 37, 573-582.

[20]UniProt C. (2009) *Nucleic Acids Research*, 38, D142-D148.

[21]Hubbard T.J., Ailey B., Brenner S.E., Murzin A.G. & Chothia C. (1999). *Nucleic Acids Research*, 27(1), 254-256.

[22]Lobley A., Sadowski M.I. & Jones D.T. (2009) *Bioinformatics*, 25, 1761-1767.

[23]Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. & Higgins D. G. (2007). *Bioinformatics*, 23(21), 2947-2948.

[24]Tamura K., Peterson D., Peterson N., Stecher G., Nei M. & Kumar S. (2011) *Molecular Biology and Evolution*, 28, 2731-2739.

[25]Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. & Bourne P.E. (2000) *Nucleic Acids Res*., 28(1), 235-242.

[26]Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. (1990) *J. Mol. Biol*., 215, 403-410.

[27]Schwede T., Kopp J., Guex N. & Peitsch M.C. (2003) *Nucleic Acids Research*, 31(13), 3381-3385.

[28]Arnold K., Bordoli L., Kopp J. & Schwede T. (2009) *Bioinformatics*, 22(2), 195-201.

[29]Wiederstein M. & Sippl M.J. (2007) *Nucleic Acids Research*, 35, W407-W410

[30]Sippl M.J. (1993) *Proteins*, 17, 355-362.

[31]Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W. & Lipman D.J. (1997) *Nucleic Acids Research*, 25(17), 3389-3402.

[32]Syed R., Rani R., Sabeena T.A., Masoodi G., Shafi K. & Alharbi K. (2012) *Bioinformation*, 8(4), 175-180.

[33]Guex N. & Peitsch M.C. (1997) *Electrophoresis*, 18, 2714-2723.

[34]Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblat D.M., Meng E.C. & Ferrin T.E. (2004) *J. Comput. Chem.*, 25(13), 1605-1612.

[35]Jones P., Binns D., Chang H.Y., Fraser M., Li W., McAnulla C., McWilliam H., Maslen J., Mitchell A., Nuka G., Pesseat S., Quinn A.F., Sangrador-Vegas A., Scheremetjew M., Yong S.Y., Lopez R. & Hunter S. (2014) *Bioinformatics*, 30(9), 1236-1240.

[36]Szklarczyk D., Franceschini A., Kuhn M., Simonovic M., Roth A., Minguez P., Doerks T., Stark M., Muller J., Bork P., Jensen L.J. & Mering C.V. (2011) *Nucleic Acids Research*, 39, D561-D568.