Insights into archaeal chaperone machinery: a network-based approach

Shikha Rani¹ · Ankush Sharma^{2,3} · Manisha Goel¹

Received: 14 June 2018 / Revised: 3 August 2018 / Accepted: 20 August 2018 / Published online: 3 September 2018
© Cell Stress Society International 2018

Abstract

Molecular chaperones are a diverse group of proteins that ensure proteome integrity by helping the proteins fold correctly and maintain their native state, thus preventing their misfolding and subsequent aggregation. The chaperone machinery of archaeal organisms has been thought to closely resemble that found in humans, at least in terms of constituent players. Very few studies have been ventured into system-level analysis of chaperones and their functioning in archaeal cells. In this study, we attempted such an analysis of chaperone-assisted protein folding in archaeal organisms through network approach using *Picrophilus torridus* as model system. The study revealed that DnaK protein of Hsp70 system acts as hub in protein-protein interaction network. However, DnaK protein was present only in a subset of archaeal organisms and absent from many archaea, especially members of Crenarchaeota phylum. Therefore, a similar network was created for another archaeal organism, *Sulfolobus solfataricus*, a member of Crenarchaeota. The chaperone network of *S. solfataricus* suggested that thermosomes played an integral part of hub proteins in archaeal organisms, where DnaK was absent. We further compared the chaperone network of archaea with that found in eukaryotic systems, by creating a similar network for *Homo sapiens*. In the human chaperone network, the UBC protein, a part of ubiquitination system, was the most important module, and interestingly, this system is known to be absent in archaeal organisms. Comprehensive comparison of these networks leads to several interesting conclusions regarding similarities and differences within archaeal chaperone machinery in comparison to humans.

Keywords Chaperones · Archaea · Networks · Protein-protein interactions · Protein folding

Introduction

Knowledge of interactions between various constituent cellular molecules is a fundamental requirement for better understanding of pathways at system level (Raman 2010). Protein-protein interaction networks can not only help in understanding the role of individual proteins in a pathway but also often bring to fore some novel hypotheses about the cellular

processes under investigation, which can be further tested in laboratories (Zak et al. 2014; Juhas et al. 2011; Zhan and Boutros 2016). Networks can also help in identification of key and essential nodes that are driving specific biological processes or even specific characteristics in organisms such as adaptability of extremophiles. Protein-protein interaction networks mostly use data derived from experimental evidences like co-expression, gene neighborhood, gene fusion, co-occurrence, and text mining, present in various databases (Marcotte et al. 1999; Pellegrini et al. 1999; Pazos and Valencia 2008). The topological analysis of networks based on centrality statistics using measures such as degree distribution, betweenness centrality, and bottleneck score can reveal hub and essential proteins as well as modular organization of the system (Przulj et al. 2004; Yu et al. 2007; Khuri and Wuchty 2015). Detailed analysis of such networks can additionally provide insights on robustness and efficiency of the system (Barabasi and Oltvai 2004; Sharma et al. 2012, 2013).

Molecular chaperones appear to be central components of living cells because of their interactions with a large number of proteins, while they facilitate the acquisition of the native state structure of these target proteins. Chaperone proteins do not

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12192-018-0933-y>) contains supplementary material, which is available to authorized users.

✉ Manisha Goel
manishagoel@south.du.ac.in

¹ Department of Biophysics, University of Delhi South Campus, Benito Jurarez Road, New Delhi 110021, India

² Department of Molecular Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands

³ Center for Computational Science, University of Miami, Miami, FL, USA

operate independently but often act as parts of complex functional networks of interacting molecules (Kampinga and Craig 2010). These chaperones interact with other proteins and act as mastermind of the cells to make efficient protein folding machinery. Chaperones often act as integrators and perform a regulatory role while overlapping in other network modules (Csermely et al. 2008). Gong et al. (2009) had performed the systematic analysis of chaperones found in *Saccharomyces cerevisiae* and elucidated the hot spot as well as the presence of multicomponent modules in this organism. The chaperone interaction networks for *Plasmodium falciparum* and its interconnection with human proteins predicted the involvement of chaperones in various cellular functions (Pavithra et al. 2007). However, there is no studies available vis-à-vis the organization of the chaperone machinery in archaeal organisms. It is currently believed that protein folding pathways or the *proteostasis machinery* of the archaeal organisms closely resembles that found in eukaryotes, rather than other prokaryotes like bacteria, at least in terms of constituent members of this system (Laksanalamai et al. 2004). Although the structure and function of individual archaeal chaperones have been studied and compared to eukaryotic and bacterial counterparts, how multiple chaperones work together in systematic form has not been addressed yet. Network studies involving archaeal chaperones could generate novel hypothesis regarding the modularity of protein folding machinery and also throw light on the evolution of these systems from prokaryotes to eukaryotes. In the present study, we have endeavored towards this end and, here, we report some interesting observations gathered during this analysis.

A manually curated database of chaperone repertoire in archaeal genomes (CrAgDb) has been developed (<http://proteininformatics.org/mkumar/cragdb>) (Rani et al. 2016). The preliminary analysis of the CrAgDb data suggested that *Picrophilus torridus* was one of archaeal organisms having chaperone representatives for a maximum number of classes among all 144 organisms studied (Rani et al. 2016). It has representatives of all chaperone families except parvulin and group I chaperonins, which are both rarely found in archaea. Therefore, we have attempted to create and analyze the protein-protein interaction network of *P. torridus* chaperones in the current study. This network was then compared to the similar network created for another archaea, *Sulfolobus solfataricus*, which is also a well-established archaeal model system (Angelov and Liebl 2006; Snijders et al. 2006; Takagi et al. 2010; Thürmer et al. 2011; Goswami et al. 2015). Thus, in the present study, protein-protein interaction networks of chaperones from two archaeal organisms (*P. torridus* and *S. solfataricus*) have been compared and subsequently paralleled to a similar network of *Homo sapiens* (Eukarya). This detailed comprehensive analysis sheds new light on the similarities and differences in the modular organization of chaperone machinery in archaea as well as in Eukarya.

Materials and methodology

Construction of chaperone networks

To construct the chaperone protein-protein interaction (PPI) networks, the data related to the presence of chaperones in organisms *P. torridus* and *S. solfataricus* was derived from CrAgDb database (<http://proteininformatics.org/mkumar/cragdb>) (Rani et al. 2016). In the predicted protein-protein interactions for *P. torridus* and *S. solfataricus*, chaperones were obtained from STRING database (<http://string-db.org/>) (von Mering et al. 2003). STRING database integrates the protein-protein interactions derived from various high-throughput experiments, gene neighborhood, gene fusion, co-occurrence, co-expression, and co-existence data from different databases (Rao et al. 2014). The interactions included in our dataset were based on predictions as well as experimental data with most confident interacting partners involved. To construct the chaperone networks, the protein-protein interactions for *P. torridus* and *S. solfataricus* were retrieved with a confidence score cutoff of 0.400 (medium level). A list of human chaperones was obtained from the Human Protein Reference Database (HPRD) (<http://www.hprd.org>). A total of 149 chaperones were retrieved for *Homo sapiens*, and their protein-protein interactions were extracted from various databases like the Biological General Repository for Interaction Datasets (BioGRID) (Stark et al. 2006), HPRD (Keshava Prasad et al. 2009), Molecular Interaction (MINT) (Chatr-Aryamontri et al. 2007), and National Cancer Institute-Pathway Interaction Database (NCI-PID) (Schaefer et al. 2009). Due to different scopes of each database, these protein interaction datasets are represented in many different formats, which make it difficult to project in a single data model capturing all necessary details on the experimental setup. This propelled a need for exchanging and integrating data to avoid secondary curation and duplication of the data. International Molecular Exchange (IMEx; available at <http://www.imexconsortium.org>) consortium was formed between groups of major public data providers which share curation efforts (Orchard et al. 2012). The data is presented in Human Proteome Organization (HUPO) Proteomics Standards Initiative format (MITAB or PSI-MI XML 2.5). HUPO PSI defines community standards for data representation in proteomics to facilitate data comparison, exchange, and verification for these resources. The protein interaction for human chaperones utilized for this analysis is following the HUPO Proteomics Standard Initiative format and was manually curated and updated by the Center for Biomedical Computing (CBC) at the University of Verona, Italy (<http://dp.univr.it/~laudanna/LCTST/index.html>) (Scardoni et al. 2009). The dataset contains only experimentally known non-redundant, undirected, no-loop, physical protein-protein binary interaction dataset from all databases following HUPO PSI

community standards. These interactions were then constructed into a network and visualized using the program Cytoscape (version 3.3.0), a powerful and user-friendly tool used for visualization of chaperone networks (Kohl et al. 2011; Saito et al. 2012). Each network was considered as an undirected graph (Zhu et al. 2007), where each node represented a protein and was connected by the interactions, known as edges although the isolated and orphan nodes were removed. The graph is called connected if the nodes are connected by edges, and the nodes which are not connected with each other are known as disconnected component as described earlier in other studies (Huang et al. 2016).

Network topology and centrality statistics

The centrality statistics and topological properties of networks were calculated with network analyzer and CentiScaPe (Assenov et al. 2008; Scardoni et al. 2009). The basic topological and centrality properties such as number of nodes, edges, clustering coefficient, average number of neighbors, characteristic path length, connected components, degree of nodes, and betweenness centrality were determined for each network using Cytoscape network analyzer software (Shannon et al. 2003). These fundamental parameters were adopted to evaluate the nodes in a network (Ran et al. 2013). The topological properties such as node degree, betweenness centrality, and bottleneck score for chaperone interaction networks were calculated. The proteins (nodes) showing high betweenness centrality and bottleneck scores were termed as hub and essential nodes. The hub proteins in a network are known to play a central role, and it has been indicated in earlier studies that these may also be evolutionarily conserved (Wuchty and Almaas 2005).

Network modules

It was well established through a multitude of earlier studies that the biological networks are scale free and these scale-free networks comprised different modules lacking well-defined boundaries (Hartwell et al. 1999). The node degree of a scale-free network follows a power law, which means these networks were robust against random perturbations (Hu et al. 2016). These interconnected modules often work in cooperation to accomplish cellular processes. In protein-protein interaction networks, a majority of proteins are known to carry out more than one function and can belong to more than one protein complex (Kühner et al. 2009). The disjoint modules can be called as non-overlapping modules and correspond to functional units, whereas modules expressing coordination with other modules in the entire system are called as overlapping modules. Community centrality measured in networks identified the most central proteins (core) in modules in protein-protein interaction networks, and core proteins of the

modules were known to be important for characterization of the molecular function of the network modules (Kovács et al. 2010).

In the present study, ModuLand framework has been used to identify overlapping modules in the network (Szalay-Beko et al. 2011), and the algorithms were based on the local maxima-based gradient hill method by means of a calculation-based function approach of LinkLand algorithm (Szalay-Beko et al. 2012). Community centrality landscape (local maxima-based hill determination) was measured to identify overlapping modules, and each node of the network was assigned to modules based on their participation in modules. Overlap values for nodes revealed the number of modules to which they were assigned, and high bridgeness values for nodes showed a larger overlap between many diverse module pairs.

Gene ontology of chaperones

Gene ontological data were mapped to nodes (proteins) in the network using the Biological Networks Gene Ontology tool (BiNGO) plugin (Maere et al. 2005). Gene ontology analysis of a network annotated each node with its known biological processes, molecular function, and cellular components where it functions. Putting all such relevant information together in terms of network can certainly help the researchers identify all components and functional relevance of this complex machinery.

Results

Network analysis of chaperones from *P. torridus*

The preliminary analysis of the *Picrophilus torridus* genome suggested the presence of genes from all major chaperone families found in archaea. A total of 19 chaperones have been identified in the *P. torridus* genome through CrAgDb database (<http://proteininformatics.org/mkumar/cragdb>). The detailed list of all 19 chaperones with chaperone name, accession no., locus tag, amino acid length, molecular weight, and isoelectric point (pI) is presented in Table 1. To analyze the functional associations between various chaperones of *P. torridus*, first, only the chaperone-chaperone interaction network (seed network) was created. The seed network created among these 19 chaperones revealed that all chaperones interact with each other except CsaA and Lon-2 (Fig. 1). This seed network of 19 nodes has 27 edges representing the interactions between various chaperones. The average clustering coefficient of this network was 0.0652. The chaperone network showed that chaperone proteins (DnaK, DnaJ, GrpE), thermosomes (Ths α , Ths β), and nascent polypeptide-associated complex (NAC) were connected with a higher

Table 1 List of chaperones identified in the *P. torridus* genome

Chaperones	Accession no.	Locus tag	Protein length	Mol. wt. (Da)	pI
DnaK	YP_023618	PTO0840	613	66,353.44	5.17
DnaJ	YP_023619	PTO0841	357	40,044.12	6.80
GrpE	YP_023617	PTO0839	180	21,236.45	7.74
Thermosome α	YP_023513	PTO0735	546	58,821.52	5.55
Thermosome β	YP_023973	PTO1195	541	58,570.93	4.93
Small Hsp	YP_023199	PTO0421	178	20,665.99	4.95
Hsp20	YP_023517	PTO0739	126	14,542.71	7.84
Prefoldin α	YP_024293	PTO1515	139	15,756.77	4.95
Prefoldin β	YP_023786	PTO1008	123	14,265.14	5.15
HtpX	YP_022938	PTO0160	307	34,744.86	8.97
NAC	YP_023312	PTO0534	108	12,205.16	6.32
CsaA	YP_023187	PTO0409	107	11,792.79	8.75
Thioredoxin	YP_022899	PTO0121	132	15,282.83	4.47
Glutaredoxin	YP_023147	PTO0369	220	24,905.15	4.66
FKBP	YP_023106	PTO0328	255	29,759.52	4.95
Cyclophilin	YP_023415	PTO0637	151	16,737.87	5.90
Lon-1	YP_024013	PTO1235	649	70,679.82	5.96
Lon-2	YP_023396	PTO0618	492	55,134.10	5.82
Cdc48/Vat	YP_023234	PTO0456	744	82,385.03	5.27

number of nodes, pointing towards their pivotal role in the protein folding machinery of *P. torridus*.

In the next step, first-order interaction network was created by obtaining interacting protein partners for 19 *P. torridus* chaperones from STRING database (Fig. 2a). This network was composed of 263 nodes and 2180 interconnecting links (edges) showing a high clustering coefficient of 0.652, with each node being connected to approximately 16 neighbors. This network had 68,902 shortest path lengths. The top seven chaperone proteins showing highest values of three centrality indices are DnaK, GrpE, Ths α , Ths β , FKBP, CDC48, and

thioredoxin (Table 2). Among these top seven, the chaperone protein DnaK was connected with the highest number of other nodes in the network, suggesting that DnaK acts as a hub protein in this network. The chaperone protein DnaK stands out to be exclusive, since it had the highest degree, highest betweenness value, and highest bottleneck score as compared to other chaperones. Thus, DnaK was expected to act as a central protein that has a number of interconnection with all the chaperones and non-chaperone proteins in the *P. torridus* network. Except for the DnaK chaperone, all other proteins exhibiting high degree are non-chaperone proteins such as ribosome assembly proteins. However, in case of betweenness centrality and bottleneck scores, various other chaperones demonstrated similar distribution as that of DnaK. This inferred about higher importance of chaperones in maintaining *P. torridus* cellular machinery by regulating some crucial biological processes discussed later.

The molecular functional analysis of *P. torridus* genes through gene ontology (GO) terminology suggested that most of the proteins of this network were involved in biological processes related to binding, nucleic acid binding, structural constituents of ribosome, structural molecular activity, RNA binding, protein binding, antioxidant activity, oxidoreductase activity, unfolded protein binding, glucosidase activity, and transfer RNA (tRNA) binding with significant *e* values (Fig. 2b). Forty-eight percent of proteins in the *P. torridus* chaperone network exhibited molecular function *related to binding* (GO ID: 0005488) with a *p* value of 0.003

Fig. 1 Chaperone-chaperone interaction map of *P. torridus*. The squares represent the individual chaperone proteins as nodes. The size of nodes (larger $\rightarrow$ smaller) and colors of nodes (blue $\rightarrow$ yellow) are represented on the basis of their degree values from higher to lower

Fig. 2 **a** First-order interaction network of chaperone proteins of *P. torridus*. The square nodes represent the chaperone proteins, with their identity highlighted in red. The interactome as visualized from higher degree nodes to lower degree nodes is graded in colors from green to

blue, respectively. **b** Molecular functions assigned to the first-order chaperone network of *P. torridus* with the significant *p* values and percentage of genes involved

(Supplementary Table 1). The results reiterated that the proteins of this network were largely involved in protein folding. The function related to nucleic acid binding (GO ID: 0003676) covers 22% of proteins with a significant *e* value of 2.84×10^{-5} . The essential proteins such as DnaK, GrpE, and thermosome with high bottleneck and betweenness values were involved in the function related to *unfolded protein binding* (GO ID: 0051082) with a significant *p* value of 0.053648. The hub nodes (proteins with a high number of interactions (degree)) of *P. torridus* networks were largely involved in the constituent assembly of ribosome, nucleic acid binding.

The modular analysis of chaperone networks was also performed to identify the functionally important proteins that cluster together. The current network exhibited modular organization and was decomposed into six overlapping modules. Each module was identified by the name of the most prominent protein. Five of these are well-annotated chaperones (DnaK, CsaA, prefoldin B, Lon, and FKBP), while the sixth protein was a hypothetical protein (labeled as HP). The DnaK protein appeared to be the hub protein in the first and most prominent module of this network, which was called as the central overlapping module of the network. The core nodes (nodes with maximum assignment value) of

Table 2 Prediction of hub and essential proteins in the *Picrophilus torridus* chaperone network on the basis of degree, betweenness centrality, and bottleneck scores

ID	Degree	ID	Bottleneck score	ID	Betweenness centrality
DnaK	82	DnaK	198	DnaK	0.46
rps7p	67	rpl24e	24	Lon	0.11
rpl2p	67	GrpE	15	FKBP	0.09
rps11p	65	ThsA	15	Cdc48	0.07
rps5p	65	ThsB	15	ThsA	0.06
rps3p	65	FKBP	15	ThsB	0.06
rps12p	64	CDC48	14	Thioredoxin	0.05

The table illustrates the top seven proteins with the highest values of centrality indices

the central overlapping module in *P. torridus* are listed in Table 3.

The core node of the first module, DnaK, from the interaction network indicated extensive interactions with smaller and larger subunits of ribosomal proteins. The protein parts of module 1 were rps7p, rpl2p, rps5p, rps11p, rps12p, rps13p, rpl15e, rpl37ae, and rpl23p. The interaction with these ribosomal proteins indicated that DnaK acts as a chaperone protein that may make contact with newly synthesized polypeptide chains, since it had a large number of interactions with the proteins of the translation machinery. The central module of the network, DnaK module, showed high overlapping values with other modules and dominates the whole network. The proteins participating in DnaK module show molecular functions (GO terms) mainly related to

structural constituents of ribosome, structural molecule activity, and RNA binding.

Table 3 reveals that CsaA (PTO0409) was a hub protein for the second module, which was interacting with the highest number of other nodes in the network, second to DnaK. Prefoldin β (PTO1008) was a hub protein of the third module which was composed of 90 proteins. The core proteins of this module were mainly involved in molecular functions (GO terms) related to metabolic processes such as transferase activity and unfolded protein binding. In the fourth module, a hypothetical protein (PTO0619) acted as the hub protein. This module was composed of 34 proteins, a majority of which showed molecular function (GO terms) related to nucleotide binding and involved in amino acid metabolism. This hypothetical protein contains a single ACT domain (ACT

Table 3 Modular organization of the *P. torridus* chaperone networks

Module ID	Number of nodes	Effective number of nodes in module	Module size (sum of module assignment values)	Central protein participating in module (nodes with maximum assignment value)
DnaK	249	134.18062	4139.1953	DnaK, rps7p, rpl2p, rps5p, rps11p, rps12p, rps13p, rpl15e, rpl37ae, rpl23p
CsaA	210	31.293447	65.47925	CsaA, metG, gltX, tRNA synthetase, tRNA synthetase, ileS, tRNA synthetase proS, tyrS, valS
Prefoldin B	90	11.577784	18.139067	Prefoldin B, hypothetical protein, prebonyltransferase, Na ⁺ symporter, ribonuclease, RNA polymerase, succinyl-CoA synthetase, psmA, hypothetical protein, proteasome B
Hypothetical protein	34	12.572951	55.92704	Hypothetical protein, LAII endonuclease, Lon, GTPase, hypothetical protein, Mov34, hypothetical protein, PmbA, DnaK
Lon	31	10.148024	32.1342	Lon, hypothetical protein, Mov34, hypothetical protein, TIFIB, PmbA, phosphate aldolase, hypothetical protein, LAII, DnaK
FKBP	26	11.228254	49.124756	FKBP, hypothetical protein, sugar transport protein, succinate dehydrogenase, hypothetical protein, PPIase, MoaA, regulatory protein, NADH, oxidoreductase, eno

The table provides information on the number of nodes present in each module with module size, which is the sum of module assignment values

Table 4 A list of chaperones annotated in the *S. solfataricus* genome

Chaperones	Accession no.	Locus tag	Protein length	Mol. wt. (Da)	pI
Prefoldin α (PfdA)	NP_341889.1	SSO0349	147	16,330.59	5.2
Prefoldin β (PfdB)	NP_342237.1	SSO0730	126	14,543.78	6.37
Thermosome α (ThsA)	NP_342362.1	SSO0862	559	59,676.21	5.34
Thermosome β (ThsB)	NP_341830.1	SSO0282	557	60,368.64	5.55
Thermosome γ (ThsC)	NP_344314.1	SSO3000	539	59,273.61	5.51
Hsp20	NP_343781.1	SSO2427	176	20,098.03	5.28
Hsp20 (1)	NP_343935.1	SSO2603	124	14,145.54	7.51
HtpX1	NP_343257.1	SSO1859	311	34,986.45	6.6
HtpX2	NP_344535.1	SSO3231	325	34,171.86	9.20
htpX-like	NP_344019.1	SSO2694	291	36,421.05	8.85
NAC	YP_005643197.1	SsoI_1362	116	12,932.95	6.36
FKBP	NP_342264.1	SSO0758	235	26,717.55	6.65
Thioredoxin (trxA-1)	NP_341908.1	SSO0368	133	15,229.51	5.02
Thioredoxin (trxA-2)	NP_343612.1	SSO2232	135	15,521.08	4.87
Glutaredoxin	NP_341748.1	SSO0192	233	26,033.69	4.69
Glutaredoxin 1	NP_342586.1	SSO1120	171	19,931.84	5.33
pan	NP_341819.1	SSO0271	393	44,143.36	5.29
ATPaseAAA	NP_341734.1	Saci_1462	369	85,809.51	5.75
ATPaseAAA1	NP_341956.1	SSO0176	769	86,348.03	6.42
ATPaseAAA2	NP_342401.1	SSOP1_0945	372	42,249.04	6.19
ATPaseAAA3	NP_343775.1	SSO2420	607	68,397.85	8.78
ATPaseAAA4	NP_344151.1	SSO2831	585	66,979.02	7.68

superfamily, CDD accession ID: cl09141). The ACT domains are known to regulate the enzymes via specific binding of an amino acid or other small ligands. This domain has been reported in a variety of proteins involved in amino acid biosynthesis, phenylalanine hydroxylation, regulation of bacterial metabolism, and transcription. Several proteins are known to have a single, double, or even four ACT domain repeats. This hypothetical protein does not appear to have a homolog in either bacteria or eukaryotes as deduced from similarity searches using BLAST. It appears to be an archaea-specific protein found only in organisms of Euryarchaeota phylum but totally absent in Crenarchaeota phylum of archaeal domain. It is possible that this protein could have hitherto unknown chaperone-like function, which needs to be validated through wet lab studies. The module with the fifth largest density of internal connections had Lon (PTO1235) as the hub protein. The core proteins of this modules showed functionality related with binding a variety of biological molecules such as DNA, ATP, metal, and proteins. Lon may act as a regulator of transcription factors, due to its interaction with transcription factor regulator proteins. FKBP (PTO0328) was the hub protein in the sixth and the smallest module, which was composed of 26 different proteins. The molecular functional analysis of this module suggested the involvement of FKBP in sugar metabolism along with other proteins showing functions related to

metal binding, transporter activity, and co-enzyme binding (Table 3).

The overall analysis thus suggested that DnaK protein forms the major hub of chaperone-related activities in *P. torridus*, which is in agreement to the observations from previously studied bacterial organisms (Calloni et al. 2012). However, this observation cannot be generalized for the whole archaeal kingdom. The DnaK protein was present in all the bacterial and eukaryotic organisms, but this important class of chaperones was known to be absent from a substantial number of archaeal organisms. Therefore, it becomes imperative to study the chaperone-chaperone and chaperone-protein interactions in another archaeal model system where DnaK was absent. We, therefore, selected another archaeal organism, *Sulfolobus solfataricus* from Crenarchaeota phylum, as a model system to understand the chaperone organization and functioning in this scenario.

Network analysis of chaperones from *S. solfataricus*

CrAgDb analysis identified 22 chaperones in the genome of *S. solfataricus*. Table 4 lists all these 22 chaperones with chaperone name, accession no., locus tag, amino acid length, molecular weight, and pI. Although the total number of chaperone proteins used to create this chaperone network for

Fig. 3 Chaperone-chaperone interaction map of *S. solfataricus*. The squares represent the individual chaperone proteins as nodes. The size of nodes (larger → smaller) and colors of nodes (blue → yellow) are represented on the basis of their degree values from higher to lower

S. solfataricus was higher than that present in *P. torridus*, it was only because there were multiple paralogs in a single chaperone family whereas several chaperone classes are totally absent from *S. solfataricus*. These chaperones were then used as seed proteins for creating the chaperone interaction network using a methodology similar to that used for *P. torridus*. All the chaperone proteins of *S. solfataricus* appeared to be interconnected with each other, the 22 nodes of the network being extensively interconnected with 83 edges (Fig. 3). The clustering coefficient of chaperone-chaperone interaction networks was calculated to be 0.280 with the average numbers of neighbors being 7.2. The size of the nodes in Fig. 3 was correlated to higher values of degree for each protein. Figure 4a represents a first-order network for *S. solfataricus* chaperones with their interacting partners, created using the data extracted from STRING database. The map was composed of 345 nodes connected by 908 links, with the clustering coefficient of 0.069. The average number of neighbors in this network was 5.15. The size of the nodes (Fig. 4a) was proportional to the degree of interactions. From Fig. 4a, it was evident that thermosomes, prefoldins, thioredoxin, and AAA⁺ATPases appeared to be nodes with highest values of degree. This network also revealed that thsA, thsB, and thsC (three paralogs of thermosomes) were hub nodes, which were connected with the highest number of other nodes. Few other chaperones, like ATPases, also exhibited similar values of degree to those of thermosomes, suggesting that most chaperones interact with a more or less similar number of partners and, therefore, the burden of protein folding appears to be equally divided among all constituent chaperones in *S. solfataricus*. The chaperone protein thermosome was the central node of *S. solfataricus* network, since it exhibited a highest degree. These hub proteins (thermosomes) also showed highest bottleneck scores, suggesting that this chaperone played an essential role in information transfer.

The gene ontology analysis of constituents of this network showed that the proteins were involved in a variety of molecular functions such as oxidoreductase activity, nucleoside binding, purine nucleotide binding, antioxidant activity, dihydrolipoyl dehydrogenase activity, DNA-directed RNA polymerase activity, ribonucleotide binding, ligase activity, and ATP binding (Fig. 4b). The analysis revealed that 24% of proteins were involved in nucleotide binding (GO ID: 0000166) with a significant *e* value of 0.00415. The majority of proteins including one of the major hubs, thsA in *S. solfataricus* network, showed molecular function related to ligase activity, forming aminoacyl-tRNA and related compounds (GO ID: 0016876, *p* value = 0.006054). The major functions of network components were enriched in purine nucleotide binding (GO ID: 0017076, *p* value = 5.14e−04) and ATP binding (GO ID: 0005524, *p* value = 0.008739). The biological processes of *S. solfataricus* network components were mainly related to translation (GO ID: 0006412, *p* value = 0.00132152) and cellular homeostasis (GO ID: 0019725, *p* value = 5.15e−07) (Supplementary Table 2). The thsA proteins were involved in a large number of biological processes. The seed proteins thsA, thsB, thsC, pan, trxA-1, trxA-2, prefoldin (pfdA and pfdB), and ATPaseAAA revealed a higher number of interactions among themselves as compared to those seen in the case of network from *P. torridus*. From the topology and centrality analysis, it was inferred that proteins thsA, thsB, thsC, pan, trxA-1, trxA-2, prefoldin (pfdA and pfdB), and ATPaseAAA had the highest degree, the betweenness centrality, and the bottleneck score in *S. solfataricus* network (Table 5).

The modular analysis was performed for *S. solfataricus* network which predicted two overlapping modules based on the numbers of edges present in the cluster, where one module was covering most of the nodes of network, while the second was composed of only a few nodes. For the first module, pfdA which acted as hub protein showed high overlapping values and was connected with 338 proteins, which dominated the whole network. The well-connected nodes of pfdA were also chaperone proteins that together composed the whole protein folding machinery of the archaeal organisms. The high bottleneck values also signified their importance in the protein folding machinery. The other module had FKBP as the hub protein, connected with 297 different proteins. Ten central proteins participating in each module showed interactions with highly conserved proteins such as RRP42 and rps19e as listed in Table 6, which played an essential role in RNA degradation pathways and large ribosomal subunit rRNA binding, respectively.

Thus, the chaperone networks from both the archaeal organisms exhibited significant differences. Could either of these act as a representative for the eukaryotic counterparts remained to be explored. We, therefore, chose to create a network for human chaperones in the next step.

a **Molecular function****b**

Fig. 4 **a** First-order chaperone interaction network of chaperone proteins from *S. solfataricus*. Chaperone proteins of the network are shown in square shape. The interactome as visualized from higher degree nodes to lower degree nodes is graded in colors from blue to orange,

respectively. **b** Molecular functions assigned to the first-order chaperone network of *S. solfataricus* with the significant *p* values and percentage of genes involved

Network analysis of chaperones from humans

Human chaperone proteins were retrieved from HPRD, and their interactions were obtained through various databases listed in “Materials and methodology” section. A total of 149 chaperones were identified, performing a wide variety of molecular function mainly related protein folding, stress response, assistance in protein degradation, and transport across membrane. The interaction network of human

chaperones (seed network) was composed of 82 chaperone proteins linked with 387 edges showing scale-free topology and small-world property (Fig. 5). The disconnected component of the network was comprised of 67 proteins linked with six edges. The human chaperone network exhibited a clustering coefficient of 0.503 with an average number of neighbors being 9.439 and the characteristic path length of 2.

To better understand the importance of chaperone proteins and their interacting partners in human proteome, the protein-

Table 5 Hub and essential proteins in the *S. solfataricus* chaperone network on the basis of centrality indices

ID	Degree	ID	Betweenness centrality	ID	Bottleneck score
ThsA	61	PfdA	0.21	ThsB	52
ThsC	61	PfdB	0.19	ThsC	46
ThsB	61	ThsB	0.16	ATPaseAAA4	44
pan	55	ThsC	0.16	Trx-2	43
TrxA-2	55	ThsA	0.14	PfdA	41
TrxA-1	55	pan	0.14	ATPaseAAA	37
PfdB	52	ATPaseAAA4	0.1	pan	31

protein interaction network of 149 human chaperones was constructed, which was composed of 3675 proteins and showed 140,953 interactions among them (Fig. 6a). This network showed a scale-free and small-world property following the paradigm of real-world networks. The clustering coefficient (0.366) of proteins in first-order human chaperone network showed lower capacity of creating tightly knit groups characterized by a relatively high density of links as compared with the seed network of human chaperones. Evaluation of the network suggested that the chaperone proteins HSPA5, HSPA8, and HSP90AA1 were highly connected with other chaperones and, therefore, act as hub proteins of the networks. Putatively important proteins of first-order interaction networks of human chaperones were detected on the basis of highest degree nodes, betweenness centrality, and bottleneck scores, and it provided information about the core skeleton of the network. Table 7 depicts that the nodes of proteins polyubiquitin-C, HSP90AA1, SUMO2, ELAVL1, HSPA5, and HSPA8 had the highest values of these topological characteristics. These nodes may act as essential proteins and would be responsible for maintaining the interaction and information flow through the network.

The first-order human chaperone interaction network contained 12.08% of proteins showing molecular function (as GO terms) related to transcription regulator activity with p -values of $1.90e-05$, and 7.07% of proteins were predicted to be involved in receptor binding with a p value of $2.37e-02$. The hubs and essential nodes in this network were comprised of heat shock proteins (HSP90AA1, HSPA1, HSPA8), eukaryotic translation factor (EEF1A1), and proteins performing post-translational modifications (SUMO1, SUMO2) for transcriptional regulation, stress response, and protein stability. The molecular function assigned to core nodes of modules shows function related to unfolded protein binding (GO ID: 51082, p value = $1.19e-03$) and transporter activity. The core nodes of the modules showed heterogeneous molecular function, indicating that many proteins with different molecular functions interacted in this case. About 41.5% of the proteins in network showed molecular function related to catalytic activity (GO ID: 3824, p value = $1.1515e-27$), and 16.7% showed function related to transferase activity (GO ID:

16740, p value = $3.60e-33$) and transcription regulator activity (GO ID: 30528, p value = $1.90e-05$) (Supplementary Table 3). The proteins were also involved in function related to heat shock-binding proteins and phosphatase binding (Fig. 6b).

The modular organization of the first-order network predicted eight overlapping modules with a core module containing interconnected hub nodes, indicating swift transmission of information across proteins. The first module had UBC as the hub protein, a polyubiquitin-C protein involved in protease binding and RNA binding, linked with the highest number of nodes (3672). The second module had TUBA1A as the hub and has 3660 interconnections. The next module, named CCT5 after its hub node, was made up of 3642 nodes. The hub protein for the next module was RAN (involved in transcriptional regulation process), with a total number of nodes being 3635, which were involved in the process. The fifth module had chaperonin protein HSPD1 as the hub protein. SRRM2 proteins were predicted as the hub for the sixth module and involved in the transcriptional activation of the molecules having a number of 3616 nodes. The seventh module had VCAM1 as the hub protein and was involved mainly in the process of signal transduction. The last and the eighth module was having CTNNB1, a catenin beta-1 protein, as the hub protein that is an important component of the signaling pathway. The hub chaperone proteins contain members of HspA and CCT family in the chaperone-chaperone network showing molecular function related to de novo post-translational protein folding, protein refolding, and chaperone-mediated protein complex assembly (Table 8).

Comparison of the three chaperone networks

The analysis of chaperone networks from *P. torridus*, *S. solfataricus*, and *H. sapiens* revealed that there are 14 chaperone proteins that were common in the protein folding machinery of *P. torridus* and *H. sapiens* whereas nine were shared between *S. solfataricus* and *H. sapiens* (Table 9). The chaperone classes represented in the *P. torridus* genome are larger as compared to those found in *S. solfataricus*. The homologs of chaperones DnaK, DnaJ, GrpE, NAC, and Lon were absent in *S. solfataricus*. Thus, the chaperone repertoire

Table 6 Modular organization of the *S. solfataricus* chaperone networks

Module ID	Number of nodes	Effective number of nodes	Module size (sum of module assignment values)	Central protein participating in module (nodes with maximum assignment value)
PfdA	338	78.85254	802,338.0	PfdA, PfdB, ThsB, YhsC, ThsA, pan, Trx-2, ATPAAA4ase, TrxA-1, ATPaseAAA
FKBP	297	70.249593	22,363.309	FKBP, psmB2, spt5, ppa, hycE, ileS, ATPaseAAA4, HtpX1, htpX-like, ATPaseAAA4

The table provides information on the number of nodes present in each module with module size, which is the sum of module assignment values. The central node of module has the maximum assignment value

of *P. torridus* was closer to that found in *H. sapiens* than that shared between *S. solfataricus* and *H. sapiens*.

The global centrality statistics of chaperone networks of *P. torridus*, *S. solfataricus*, and *H. sapiens* are compared in Table 10. The number of interacting proteins in human network was much higher (about ten times) than in case of archaea. Consequently, the numbers of interactions in the complex human proteome are also many times higher than those predicted for *P. torridus*. This is also reflected in the values for

an average number of neighbors, which was highest for human network, suggesting that each node or protein interacted with a larger number of proteins in case of humans than in archaeal organisms. The average number of neighbors for the *S. solfataricus* networks was even smaller than that found for the *P. torridus* network. The centrality statistics of chaperone networks revealed that the average clustering coefficient value of *P. torridus*, *S. solfataricus*, and *H. sapiens* are 0.0652, 0.069, and 0.366, respectively, which was again significantly higher in case of humans than in archaea, suggesting that proteins in case of human network tend to cluster together to form modules. This possibly helps in efficiently managing a multitude of different biological processes. Shorter path lengths as seen in case of human network also signify a more efficient information transfer between various nodes. However, network density was highest in *P. torridus*, meaning that this network actually makes a larger number of connections out of total possible connections than the other two networks (*S. solfataricus* and *H. sapiens*).

The *P. torridus* network was composed of six overlapping modules, whereas the *S. solfataricus* network contained only two modules. In the *P. torridus* networks, the core modules in the network were chaperones such as DnaK, CsaA, prefoldin β , hypothetical protein, Lon, and FKBP. Thus, DnaK and CsaA were the two biggest interactors and it is an interesting observation because both DnaK and CsaA are absent in *S. solfataricus*. The *S. solfataricus* network shows prefoldin α and FKBP as hub proteins of two overlapping modules. In this case, both proteins appear to be equally important, unlike in case of *P. torridus*, where DnaK tends to dominate the

Fig. 5 Chaperone interaction maps of human chaperones shows interactions in the form of nodes and edges. The size of the nodes (larger $\rightarrow$ smaller) and colors of nodes (blue $\rightarrow$ yellow) are represented on the basis of their degree values from higher to lower

Fig. 6 **a** First-order chaperone interaction network of chaperone proteins from *H. sapiens*. **b** Molecular functions assigned to the first-order chaperone network of *H. sapiens* with the significant *p* values and percentage of genes involved

Table 7 Important proteins in the human chaperone network calculated on the basis of centrality indices

ID	Degree	ID	Betweenness centrality	ID	Bottleneck score
UBC	3054	UBC	0.399	UBC	1891
HSP90AA1	1046	HSP90AA1	0.052	HSP90AA1	58
SUMO2	887	APP	0.016	ELAVL1	20
CUL3	742	NRF1	0.015	NPM1	19
NRF1	703	SUMO2	0.014	APP	19
HSPA5	684	ELAVL1	0.012	ACTB	16
HSPA8	666	HSPA8	0.010	AKT1	13
EEF1A1	639	HSPA5	0.009	KIAA0101	12
APP	630	CUL3	0.009	PTK2B	11
ELAVL1	626	NPM1	0.006	FN1	11

whole network as the hub protein. The *Homo sapiens* network was composed of a total of eight overlapping modules named UBC, TUBA1A, CCT5, RAN, HSPD1, SRRM2, VCAM1, and CTNBN1. The centrality statistics show that the UBC module was one of the most important modules of the human chaperone network system.

Discussion

The network analysis of chaperone molecules was essential to rebuild intermolecular contacts, recognize potential interaction partners in different modules, as well as identify the hub and essential proteins of this machinery. The combined analysis of such chaperone networks from *Picrophilus torridus* and *Sulfolobus solfataricus* in comparison to the network from *Homo sapiens* highlighted important differences in these machineries. It had been established previously that larger varieties of chaperones were found in organisms of phylum Euryarchaeota as compared to Crenarchaeota, as also evident from CrAgDb (Rani et al. 2016). However, the archaeal organisms are currently not as widely studied as their bacterial and eukaryotic counterparts, leading to lacunae in information required to create robust networks, which may eventually affect the biological significance of such networks. However, STRING database represents the most current and updated form of protein interaction data available. Protein interactions are inferred using various techniques such as physical binding or functional associations. There can be quirks in applicability, false positives, false negatives, etc., in each inference technique, which may result in some noisy observations. STRING assigns a confidence score between zero and one to each interaction, based on the available evidences. This score is based on different parameters such as in-house prediction of functional protein association and homology transfers. It also takes into account several other updated and maintained databases. In the present study, the data obtained for *S. solfataricus* and *P. torridus* from STRING database is with a confidence

score of 0.400 which can be considered as optimal, keeping in view the small proteome of these organisms and fewer evidenced supporting information (Szkarczyk et al. 2017).

It was evident that the chaperone DnaK functions as a central hub in the chaperone network of *P. torridus*, showing a vast number of interactions with other proteins. Jeong et al. (2001) suggested that proteins connected with a larger number of proteins (hub proteins) were more significant in the network than the proteins showing a lesser number of interactions. The nodes that have higher degree values have been considered to play a central role in the network (Vallabhajosyula et al. 2009). It has also been shown that the connected hubs are of higher functional importance (He and Zhang 2006). Interestingly, DnaK has been reported to function as a central organizer of the chaperone networks in some earlier studies also, for example in *Escherichia coli* (Deuerling et al. 2003). In fact, experimental studies have also revealed that DnaK interacts with ribosomal proteins and other common proteins, if there was loss of trigger factor (TF) in bacteria. The combined loss of DnaK and TF resulted in ribosomal degradation, defective functions of other chaperone proteins, misfolding, and aggregation of a large number of proteins in *E. coli* (Calloni et al. 2012). However, DnaK was shown to be absent in hyperthermophilic archaea, several mesophilic and thermophilic archaea, and especially, members of the Crenarchaeota (Macario and Conway de Macario 1999; Laksanalamai et al. 2004). The CrAgDb database also corroborates this earlier observation (Rani et al. 2016). Considering the diversity of biological functions that have been described to DnaK (Craig et al. 1993), it would be interesting to explore what other proteins can fulfill these functions in Crenarchaea. Only in one case, *Haloferax volcanii*, it has been shown that DnaK was not essential and dnaK mutant shows no obvious stress phenotype (Zhang et al. 2006). We, therefore, explored further by creating a network for an archaeal organism from Crenarchaeota phylum. Our interactomic analysis at system level shows that ubiquitous chaperone, the thermosomes (chaperonins of group II), functions as a central hub in the

Table 8 Modular organization of the *H. sapiens* chaperone networks

Module ID	Number of nodes	Effective number of nodes	Module size (sum of module assignment values)	Module center nodes (nodes with maximum assignment value)
UBC	3672	869.42993	5,798,632.07	UBC, HSP90AA, APP, NRF1, SUMO2, ELAVL1, HSPA5, CUL3, NPM1
TUBA1A	3660	48.295635	122,934.03	TUBA1A, TUBA1A, IFNE, NPPC, CXCR1, SUMO2, LTA, DUSP23, MME, CUL3, DFFA
CCT5	3642	23.29238	57,804.26	CCT5, THEG, RGS7, GLYATL3, DOCK5, GNB5, IER5, PPP2R5B, TBC1D17, CTTNBP2
RAN	3635	24.235872	53,409.96	RAN, NSMF, GFII1B, ASB16, ZZZ3, KIAA1377, GADD45G, VRK2, NFATC1, HSPB3
HSPD1	3647	40.3181	92,771.766	HSPD1, MT-CO1, MCL1, CYCS, DIABLO, TLR1, LINC00846, METTL20, NRG1, B3GNT9
SRRM2	3616	12.675551	35,901.344	SRRM2, CALR3, CLK4, PRNP, MMP2, BAK1, CA2, YWHAE, PPP4R2, ESR1
VCAM1	3639	24.385048	39,264.59	VCAM1, CCL22, CNOT6, HBA2, CTSG, DDAH2, HBA1, MRPL43, TALDO1, SLC2A1
CTNNB1	3647	23.239563	46,830.855	CTNNB1, CDH16, LSAMP, LGALS9, LRP2, DOT1L, CPLX1, PITX2, L1CAM, LRP5

The table provides information on the number of nodes present in each module with module size, which is the sum of module assignment values. The central node of module has the maximum assignment value

S. solfataricus network with a large number of interactions. Interestingly, it has previously been reported in bacteria that other components, particularly Hsp60s, could make up for the absence of Hsp70 (Bukau and Horwich 1998). In fact, overproduction of GroEL and GroES (the two proteins of Hsp60 in bacteria) has been shown to compensate for the loss of DnaK in *E. coli* (Vorderwülbecke et al. 2004).

The topological analysis based on centrality statistics, degree distribution, betweenness centrality measures, and bottleneck score revealed hub and essential chaperone proteins that may act as an important component of the protein folding machinery. The proteins with higher betweenness centrality were considered to be essential. The strong relationship has been observed between the essentiality and betweenness values in earlier studies (Joy et al. 2005). Various researchers have earlier reported the importance of bottleneck nodes in the protein-protein interaction network, and their correlation with protein essentiality was highly important (Przulj et al. 2004; Yu et al. 2007). Several researches have already demonstrated the importance and essentiality of bottleneck nodes in protein-protein interaction networks (Przulj et al. 2004; Yu et al. 2007). The characteristic path length and clustering coefficient demonstrate a small-world property of the network, whereas the nodes can be connected with each other through a small number of steps. The *P. torridus* and *S. solfataricus* demonstrated relatively similar average clustering coefficient indicating similarity in tendency of proteins in the networks to cluster together. However, the network heterogeneity of *S. solfataricus* is considerably higher than that of *P. torridus*. The *P. torridus* network shows higher degrees as compared to the chaperone networks of *S. solfataricus*. Betweenness centrality measure is higher for *S. solfataricus* as compared with *P. torridus*. Smaller betweenness centrality values of the *P. torridus* network indicate slower informational flow across its modules in stressed state that *P. torridus* may encounter due to highly extremophile conditions (Mihalik and Csermely 2011).

The networks presented in this paper provide valuable clues to chaperone function with GO analysis of proteins involved in the networks of *P. torridus*, *S. solfataricus*, and *Homo sapiens*. The networks analysis showed that the chaperone proteins of these organisms were mostly involved in molecular functions such as interacting with ribosomes and nucleic acid-binding proteins, ATP binding, transcriptional regulator activity stress response, and protein stability which signify that they are involved in protein folding and DNA repair activities when the cells are under stress.

The *P. torridus* network was composed of six overlapping modules. The DnaK and CsaA were two hub modules, and it was interesting to see this because both DnaK and CsaA were absent in *S. solfataricus*. The *S. solfataricus* network shows prefoldin α and FKBP as hub proteins of the two predicted modules. Prefoldin α is essentially needed to restrain

Table 9 The chaperone homologs present in *P. torridus*, *S. solfataricus*, and *H. sapiens*

S. no.	<i>P. torridus</i>	<i>S. solfataricus</i>	<i>H. sapiens</i>
1	Small Hsp	Hsp20	CRYAA
2	Hsp20	Hsp20	CRYAB
3	Lon II	—	HSP90B1
4	DnaK	—	HSPA90B
5	Nac	—	NACA
6	FKBP	FKBP	FKBP6
7	Pfd β	Pfd β	PFDN1
8	Pfd β	Pfd β	PFDN4
9	Pfd α	Pfd α	PFDN5
10	PPIase	FKBP	PPIL2
11	Thermosome α	Ths α	CCT3
12	Thermosome β	Ths β	CCT3
13	DnaJ	—	DNAJA3
14	GrpE	—	GRPEL2

denatured protein from aggregation, helping to maintain cellular viability under stress conditions, whereas FKBP's are known to speed up the folding of intracellular proteins in thermophilic organisms and may have a similar role in *S. solfataricus* (Ideno and Maruyama 2002). The two modules computed are governing the network and may be needed to contribute significantly to *S. solfataricus* in the maintenance of cellular machinery in order to perform other important indispensable functions. The *Homo sapiens* network was composed of a total of eight overlapping modules. The central hub of the *Homo sapiens* network was the ubiquitin protein UBC. The UBC protein was related to the ubiquitination system, which is totally absent from archaeal organisms. In fact, the ubiquitin system is unique to eukaryotes and is thought to target proteins for degradation, suggesting that eukaryotic cells may have evolved to prioritize the process of degradation

and removal of incorrectly folded proteins over and above the elemental protein folding process.

Conclusions

This chaperone network analysis study holds the key for understanding the association between various chaperones, helping to gain insights into the protein folding mechanism in the archaeal organisms. Our findings predicted that DnaK functions as a central hub in organisms where Hsp70 machinery was also present, which is similar to what has been seen earlier in case of *E. coli*, whereas in organisms of Crenarchaeota phylum, where DnaK protein was absent, other chaperones like thermosomes and other prefoldins act as central organizers. The modular organization of the two networks was thus different. In cases where DnaK was not present, all other chaperone constituents appear to be equally important in contrast to the monopoly exerted by DnaK in controlling proteostasis machinery of *P. torridus*. Most interestingly, neither of archaeal organisms matches the modular organization predicted for the human network. The two hub proteins predicted for the human network were part of the ubiquitination and Hsp90 systems, both of these systems are absent from all archaea. Finally, our analysis provides a theoretical basis for further experiments that can be designed for furthering the understanding of involvement of DnaK, thermosomes, and prefoldin genes in controlling the archaeal protein folding machinery. One novel hypothetical protein has been predicted to be the central protein of one of the six modules found in *P. torridus*. Whether this protein is a novel type of chaperone also remains to be validated. We believe that the chaperone networks would contribute to archaeal exploitation to extreme environmental conditions and might provide an insight into the evolution of complex protein folding machinery of the eukaryotic systems.

Table 10 Global centrality statistics of chaperone networks of *P. torridus*, *S. solfataricus*, and *H. sapiens*

Parameters	<i>P. torridus</i>	<i>S. solfataricus</i>	<i>H. sapiens</i>
Number of nodes	263	345	3675
Number of edges	2180	908	140,953
Average clustering coefficient	0.0652	0.069	0.366
Network diameter	6	5	4
Network radius	3	3	4
Network centralization	0.252	0.163	0.361
Shortest paths	68,906	118,680	13,501,950
Characteristic path length	3.020	3.11	2.164
Avg. no. of neighbors	16.578	5.154	76.709
Network density	0.063	0.015	0.021
Network heterogeneity	1.151	2.157	1.312

Acknowledgements We are thankful to the reviewers for appreciating the work we have done and for providing detailed and constructive suggestions.

Funding information Shikha Rani gratefully acknowledges the fellowship and grant received for this work from the “WOS-A” Scheme of Department of Science and Technology (DST), Government of India, New Delhi (Grant No. SR/WOS-A/LS-123/2011).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Network Glossary

$CC_n = 1 / \text{avg}(L(n, m))$ where $L(n, m)$ is the length of the shortest path between two nodes (n and m). The closeness centrality of each node is a number between 0 and 1.

Average path length Average path length is defined as the mean value of all shortest paths between all pairs of nodes. It is one of the most robust measures of topology of network and calculates the efficiency of flow of information. It indicates the expected distance between two connected nodes and is also known as the characteristic path length (Doncheva et al. 2012).

$BC_n = \sum_{st} (\sigma_{st}(n)) / \sigma_{st}$ where σ_{st} is the number of shortest path between nodes s and t and $\sigma_{st}(n)$ is the number of paths that pass through node n . Thus, the betweenness centrality of each node is a number between 0 and 1 (Yoon et al. 2006).

Betweenness centrality (BC) Betweenness centrality is an important global centrality measure which provides information about the core skeleton and suggests the node's importance to the network. The betweenness centrality BC_n of a node is defined as the number of shortest paths between every two other nodes in the network that pass through the node.

Centrality Centrality is a measure of how many connections one node has to other nodes.

Closeness centrality (CC) Defined as a measure of degree to which a node is near to all other nodes in the network and inverse of farness. It provides information on a measure of how long it will take to spread information from one particular node. The closeness centrality CC_n of a node n is the reciprocal of the average shortest path length.

Clustering coefficient Clustering coefficient is defined as a measure of degree to which a node tends to

$$C_n = 2e_n / (k_n(k_n - 1))$$

Connected component

Degree distribution

Edges

Hubs

Modules (clusters)

Network diameter

Nodes

Shortest path length

cluster together and estimates how densely a node is connected to its neighbors. In undirected networks, clustering coefficient C_n of node is defined as

where k_n is the number of degree of node and e_n is the number of connected pairs between all neighbors of n . The clustering coefficient of a node is always a number between 0 and 1 (Ravasz et al. 2002).

Defined as the nodes of a network that are pairwise connected with each other. The number of connected components is an indicator of the global connectivity of a network.

Degree is defined as the number of proteins that have a certain number of connections to other nodes in the network. The relative degree distribution is the probability distribution of these degrees over the whole network. The degree (K) of a node is the number of links (L) (interactions) associated with node, and the average degree (K) can be calculated as $2L / N$. Degree distribution is a probability distribution of the connection (degrees) that nodes have in a network (Ran et al. 2013).

Edges are defined as the ties or connections between the nodes.

Highly connected nodes having a significant and larger number of interacting partners than others are called hubs (He et al. 2006).

Scale-free networks are composed of modular structures. The nodes that are highly connected and form different modules are likely to be functionally related to each other. Modular design is the critical aspect of the robustness of any network (Newman 2006).

Defined as the measure of the maximum length of the shortest path between two nodes in the whole network. It indicates the size of the network (Doncheva et al. 2012).

Nodes can be defined as the number of proteins present in a network. Protein interaction maps consisted of a set of N nodes and various links among them. Shortest path length is defined as the shortest possible distance between two nodes in a network (Watts and Strogatz 1998).

References

- Angelov A, Liebl W (2006) Insights into extreme thermoacidophilicity based on genome analysis of *Picrophilus torridus* and other thermoacidophilic archaea. *J Biotechnol* 126(1):3–10
- Assenov Y, Ramírez F, Schelhorn SE, Lengauer T, Albrecht M (2008) Computing topological parameters of biological networks. *Bioinformatics* 24(2):282–284
- Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5:101–113
- Bukau B, Horwich AL (1998) The Hsp70 and Hsp60 chaperone machines. *Cell* 92(3):351–366
- Calloni G, Chen T, Schermann SM, Chang HC, Genevoux P, Agostini F, Tartaglia GG, Hayer-Hartl M, Hartl FU (2012) DnaK functions as a central hub in the *E. coli* chaperone network. *Cell Rep* 1(3):251–264
- Chatr-Aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, Cesareni G (2007) MINT: the Molecular INTeraction database. *Nucleic Acids Res* 35:D572–D574
- Craig EA, Gambill BD, Nelson RJ (1993) Heat shock proteins: molecular chaperones of protein biogenesis. *Microbiol Rev* 57(2):402–414
- Csermely P, Korcsmáros T, Kovács IA, Szalay MS, Soti C (2008) Systems biology of molecular chaperone networks. *Novartis Found Symp* 291:45–54
- Deuerling E, Patzelt H, Vorderwülbecke S, Rauch T, Kramer G, Schaffitzel E, Mogk A, Schulze-Specking A, Langen H, Bukau B (2003) Trigger factor and DnaK possess overlapping substrate pools and binding specificities. *Mol Microbiol* 47(5):1317–1328
- Doncheva NT, Assenov Y, Domingues FS, Albrecht M (2012) Topological analysis and interactive visualization of biological networks and protein structures. *Nat Protoc* 7(40):670–685
- Gong Y, Kakiyama Y, Krogan N, Greenblatt J, Emili A, Zhang Z, Houry WA (2009) An atlas of chaperone-protein interactions in *Saccharomyces cerevisiae*: implications to protein folding pathways in the cell. *Mol Syst Biol* 5:275
- Goswami K, Arora J, Saha S (2015) Characterization of the MCM homohexamer from the thermoacidophilic euryarchaeon *Picrophilus torridus*. *Sci Rep* 5:9057
- Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. *Nature* 402(6761 Suppl):C47–C52
- He X, Zhang J (2006) Why do hubs tend to be essential in protein networks? *PLoS Genet* 2(6):e88
- Hu M, Shen L, Zan X, Shang X, Liu W (2016) An efficient algorithm to identify the optimal one-bit perturbation based on the basin-of-state size of Boolean networks. *Sci Rep* 6:26247
- Huang L, Liao L, Wu CH (2016) Inference of protein-protein interaction networks from multiple heterogeneous data. *EURASIP J Bioinforma Syst Biol* 2016(1):8
- Ideno A, Maruyama T (2002) Expression of long- and short-type FK506 binding proteins in hyperthermophilic archaea. *Gene* 292(1–2):57–63
- Jeong H, Mason SP, Barabasi AL, Oltvai ZN (2001) Lethality and centrality in protein networks. *Nature* 411:41–42
- Joy MP, Brock A, Ingber DE, Huang S (2005) High-betweenness proteins in the yeast protein interaction network. *J Biomed Biotechnol* 2005: 96–103
- Juhas M, Eberl L, Glass JI (2011) Essence of life: essential genes of minimal genomes. *Trends Cell Biol* 21:562–568
- Kampinga HH, Craig EA (2010) The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nat Rev Mol Cell Biol* 11(8):579–592
- Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Harrys Kishore CJ, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadrans S, Chaerkady R, Pandey A (2009) Human protein reference database—2009 update. *Nucleic Acids Res* 37(Database issue):D767–D772
- Khuri S, Wuchty S (2015) Essentiality and centrality in protein interaction networks revisited. *BMC Bioinformatics* 16:109
- Kohl M, Wiese S, Warscheid B (2011) Cytoscape: software for visualization and analysis of biological networks. *Methods Mol Biol (Clift NJ)* 696:291–303
- Kovács IA, Palotai R, Szalay MS, Csermely P (2010) Community landscapes: an integrative approach to determine overlapping network module hierarchy, identify key nodes and predict network dynamics. *PLoS One* 5(9)
- Kühner S, van Noort V, Betts MJ, Leo-Macias A, Batisse C, Rode M, Yamada T, Maier T, Bader S, Beltran-Alvarez P, Castaño-Diez D, Chen WH, Devos D, Güell M, Norambuena T, Racke I, Rybin V, Schmidt A, Yus E, Aebersold R, Herrmann R, Böttcher B, Frangakis AS, Russell RB, Serrano L, Bork P, Gavin AC (2009) Proteome organization in a genome-reduced bacterium. *Science* 326(5957): 1235–1240
- Laksanalamai P, Whitehead TA, Robb FT (2004) Minimal protein-folding systems in hyperthermophilic archaea. *Nat Rev Microbiol* 2(4):315–324
- Macario AJ, Conway de Macario E (1999) The archaeal molecular chaperone machine: peculiarities and paradoxes. *Genetics* 152(4):1277–1283
- Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21:3448–3449
- Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D (1999) Detecting protein function and protein-protein interactions from genome sequences. *Science* 285(5428):751–753
- Mihalik Á, Csermely P (2011) Heat shock partially dissociates the overlapping modules of the yeast protein-protein interaction network: a systems level model of adaptation. *PLoS Comput Biol* 7(10): e1002187
- Newman MEJ (2006) Modularity and community structure in networks. *Proc Natl Acad Sci U S A* 103:8577–8582
- Orchard S, Kerrien S, Abbani S, Aranda B, Bhate J, Bidwell S, Hermjakob H (2012) Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat Methods* 9(4):345–350
- Pavithra SR, Kumar R, Tatu U (2007) Systems analysis of chaperone networks in the malarial parasite *Plasmodium falciparum*. *PLoS Comput Biol* 3(9):1701–1715
- Pazos F, Valencia A (2008) Protein co-evolution, co-adaptation and interactions. *EMBO J* 27(20):2648–2655
- Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO (1999) Assigning protein functions by comparative genome analysis: protein phylogenomic profiles. *Proc Natl Acad Sci U S A* 96: 4285–4288
- Przulj N, Wigle DA, Jurisica I (2004) Functional topology in a network of protein interactions. *Bioinformatics* 20(3):340–348
- Raman K (2010) Construction and analysis of protein-protein interaction networks. *Autom Exp* 2(1):2
- Ran J, Li H, Fu J, Liu L, Xing Y, Li X, Shen H, Chen Y, Jiang X, Li Y, Li H (2013) Construction and analysis of the protein-protein interaction network related to essential hypertension. *BMC Syst Biol* 7:32
- Rani S, Srivastava A, Kumar M, Goel M (2016) CrAgDb—a database of annotated chaperone repertoire in archaeal genomes. *FEMS Microbiol Lett* 363(6)
- Rao VS, Srinivas K, Sujini GN, Sunand Kumar GN (2014) Protein-protein interaction detection: methods and analysis. *Int J Proteomics* 2014:147648–147660

- Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL (2002) Hierarchical organization of modularity in metabolic networks. *Science* 297(5586):1551–1555
- Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, Pico AR, Bader GD, Ideker T (2012) A travel guide to Cytoscape plugins. *Nat Methods* 9(11):1069–1076
- Scardoni G, Petterlini M, Laudanna C (2009) Analyzing biological network parameters with CentiScape. *Bioinformatics* 25:2857–2859
- Schaefer CF, Anthony K, Krupa S, Buchhoff J, Day M, Hannay T, Buetow KH (2009) PID: the pathway interaction database. *Nucleic Acids Res* 37:D674–D679
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498–2504
- Sharma A, Gautam V, Costantini S, Paladino A, Colonna G (2012) Interactomic and pharmacological insights on human sirt-1. *Front Pharmacol* 23:3–40
- Sharma A, Costantini S, Colonna G (2013) The protein-protein interaction network of the human Sirtuin family. *Biochim Biophys Acta* 1834(10):1998–2009
- Snijders APL, Walther J, Peter S, Kinnman I, de Vos MGJ, van de Werken HJG, Brouns SJJ, van der Oost J, Wright PC (2006) Reconstruction of central carbon metabolism in *Sulfolobus solfatriticus* using a two-dimensional gel electrophoresis map, stable isotope labeling and DNA microarray analysis. *Proteomics* 6(15):1518–1529
- Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* 34:D535–D539
- Szalay-Beko M, Palotai R, Szappanos B, Kovacs IA, Papp B, Csermely P (2011) ModuLand plug-in for Cytoscape: extensively overlapping modules, community centrality and their use in biological networks. *Bioinformatics* 28(16):2202–2204
- Szalay-Beko M, Palotai R, Szappanos B, Kovács IA, Papp B, Csermely P (2012) ModuLand plug-in for Cytoscape: determination of hierarchical layers of overlapping network modules and community centrality. *Bioinformatics* 28(16):2202–2204
- Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 45(D1):D362–D368
- Takagi M, Tamaki H, Miyamoto Y, Leonardi R, Hanada S, Jackowski S, Chohnan S (2010) Pantothenate kinase from the thermoacidophilic archaeon *Picrophilus torridus*. *J Bacteriol* 192(1):233–241
- Thürmer A, Voigt B, Angelov A, Albrecht D, Hecker M, Liebl W (2011) Proteomic analysis of the extremely thermoacidophilic archaeon *Picrophilus torridus* at pH and temperature values close to its growth limit. *Proteomics* 11(23):4559–4568
- Vallabhajosyula RR, Chakravarti D, Lufteali S, Ray A, Raval A (2009) Identifying hubs in protein interaction networks. *PLoS One* 4:e5344
- von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B (2003) STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* 31(1):258–261
- Vorderwülbecke S, Kramer G, Merz F, Kurz TA, Rauch T, Zachmann-Brand B, Bukau B, Deuerling E (2004) Low temperature or GroEL/ES overproduction permits growth of *Escherichia coli* cells lacking trigger factor and DnaK. *FEBS Lett* 559(1–3):181–187
- Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393(6684):440–442
- Wuchty S, Almaas E (2005) Evolutionary cores of domain co-occurrence networks. *BMC Evol Biol* 5:24
- Yoon J, Blumer A, Lee K (2006) An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality. *Bioinformatics* 22:3106–3108
- Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M (2007) The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput Biol* 3(4):e59
- Zak DE, Tam VC, Aderem A (2014) Systems-level analysis of innate immunity. *Annu Rev Immunol* 32:547–577
- Zhan T, Boutros M (2016) Towards a compendium of essential genes—from model organisms to synthetic lethality in cancer cells. *Crit Rev Biochem Mol Biol* 51(2):74–85
- Zhang W, Culley DE, Nie L, Brockman FJ (2006) DNA microarray analysis of anaerobic *Methanosarcina barkeri* reveals responses to heat shock and air exposure. *J Ind Microbiol Biotechnol* 33(9):784–790
- Zhu R, Ribeiro AS, Salahub D, Kauffman SA (2007) Studying genetic regulatory networks at the molecular level: delayed reaction stochastic models. *J Theor Biol* 246(4):725–745