Functional & Integrative Genomics
https://doi.org/10.1007/s10142-020-00740-1

ORIGINAL ARTICLE

Chaperone client proteins evolve slower than non-client proteins

Manish Prakash Victor[1] · Debarun Acharya[2] · Sandip Chakraborty[1] · Tapash Chandra Ghosh[1]

Received: 21 January 2019 / Revised: 26 March 2020 / Accepted: 17 April 2020

Abstract

Chaperones are important molecular machinery that assists proteins to attain their native three-dimensional structure crucial for function. Earlier studies using experimental evolution showed that chaperones impose a relaxation of sequence constraints on their "client" proteins, which may lead to the fixation of slightly deleterious mutations on the latter. However, we hypothesized that such a phenomenon might be harmful to the organism in a natural physiological condition. In this study, we investigated the evolutionary rates of chaperone client and non-client proteins in five model organisms from both prokaryotic and eukaryotic lineages. Our study reveals a slower evolutionary rate of chaperone client proteins in all five organisms. Additionally, the slower folding rate and lower aggregation propensity of chaperone client proteins reveal that the chaperone may play an essential role in rescuing the slightly disadvantageous effects due to random mutations and subsequent protein misfolding. However, the fixation of such mutations is less likely to be selected in the natural population.

Keywords Chaperone · Chaperone client proteins · Evolutionary rate · Protein aggregation propensity · Protein folding rate

Introduction

Protein synthesis is an expensive cellular process in the cell that costs ~ 60% of ATP in bacteria (Park et al. 2009; Stouthamer 1973) and ~ 90% ATP in the mammalian cell (Schwanhäusser et al. 2011). To maintain the cost efficiency, the synthesized proteins must fold into their stable native structures. In *Escherichia coli*, 10 to 15% of cytoplasmic proteins require special assistance of GroEL (Ewalt et al. 1997) and ~ 20% require DnaK and DnaJ for their folding (Hartl and Hayer-Hartl 2009), which increases such cost efficiency by restoring their native structure(s) (De Maio 1999; Georgopoulos 1992; Wickner et al. 1999). Chaperones are specialized proteins that interact with the unfolded/misfolded proteins (known as "client proteins") and help them to fold into their native structure through different energy-driven mechanisms (Hartl and Hayer-Hartl 2009). This property of chaperone may also help their client proteins to tolerate the deleterious effects of mutations (Aguilar-Rodríguez et al. 2016; Rutherford 2003; Rutherford and Lindquist 1998; Tokuriki and Tawfik 2009). Several studies showed that the disease-causing missense mutations (which cause misfolding, misassembly, and aggregation) were successfully treated by the chemical chaperones (Park et al. 2009; Singh et al. 2007; Suzuki 2014) possibly due to their "stress tolerance" (Estruch 2000). However, the quality control of misfolded/unfolded proteins does not solely depend on chaperone but also involves proteases, where both can recognize the exposed hydrophobic regions in unfolded/misfolded proteins (Wickner et al. 1999).

Unlike the function of chaperones, the proteases eliminate damaged proteins from the cellular system by the energy-dependent process (Goldberg 2003). More importantly, a sizable fraction (~ 30%) of nascent proteins are degraded due to the error in translation (Goldberg and Dice 1974; Schubert et al. 2000). The "molecular triage" (whether a misfolded protein will be destroyed or refolded properly) of non-native protein structures was determined by the kinetics of partitioning between chaperones and proteases (Wickner et al. 1999). Studies also showed that chaperones and co-chaperones could trigger protease activity (McDonough and

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10142-020-00740-1) contains supplementary material, which is available to authorized users.

✉ Sandip Chakraborty
sandip1983in@gmail.com

✉ Tapash Chandra Ghosh
tapash@jcbose.ac.in

[1] Division of Bioinformatics, Bose Institute, Kolkata, West Bengal, India

[2] Department of Microbiology, Bose Institute, Kolkata, West Bengal, India

Published online: 06 May 2020

Patterson 2003) which may degrade mutant proteins (Sherman and Goldberg 1992). Therefore, the misfolded/unfolded proteins due to the mutation(s) on the gene may undergo degradation instead of refolding in natural physiological condition, and become a costly biological waste (Tomala and Korona 2008).

In natural physiological condition, endosymbiotic bacteria accumulate mildly deleterious mutations than the free-living-related lineages, but the altered proteins show native function in the presence of high expression level of the chaperone in those bacteria (Moran 1996). Experiments also suggested that the chaperones buffer the deleterious effect of mutations (Fares et al. 2002; Williams and Fares 2010) and increase mutational robustness in chaperone-rich environment (Aguilar-Rodríguez et al. 2016). However, chaperone clients still evolve slower, maybe due to the other factors that influence the evolutionary rate (Williams and Fares 2010). But interestingly, most of the experiments designed to study the role of chaperone buffering are performed by overexpressing the chaperone(s) (Aguilar-Rodríguez et al. 2016; Tokuriki and Tawfik 2009), which may imbalance the naturally occurring chaperone-protease kinetics. Moreover, the overexpression of a chaperone is energetically expensive (Sabater-Munoz et al. 2015) which may affect the growth rate due to the decline in a cell's energy budget (Wagner 2005). In accordance with this idea, Geiler-Samerotte et al. (2016) also reported that Hsp90 plays a "potentiator" role on genetic variation (which can be considered as mutation) to reduce selection pressure.

In this study, we analyzed the effect of chaperone buffering in controlling the protein evolution of their client in *Escherichia coli*, *Thermus aquaticus*, *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and *Homo sapiens*. We observed chaperone client proteins evolve slower than the non-client proteins in all organisms and the chaperone client type (whether a protein is chaperone client or not client) imposes an independent effect on the evolutionary rates. Moreover, chaperone client proteins are highly expressed and highly connected proteins but their folding rates are slower than the non-client proteins. Thus, they require the assistance of chaperone to avoid aggregation; otherwise, it would be energetically costly for the living cell.

Result and discussion

Analysis of protein evolution

At first, we compared evolutionary rates between chaperone client and non-client proteins in *Escherichia coli* to test the effect of chaperone buffering in protein evolution. We used one-to-one orthologous genes of *Salmonella enterica* to calculate the evolutionary rates of *E. coli.* We retrieved all chaperone proteins (supplementary table S1) from Uniprot (Bateman et al. 2017) and identified chaperone client proteins using the protein-protein interaction network from BioGRID database (version 3.4.152) (Chatr-Aryamontri et al. 2017) (see "Materials and methods" for further details). Interestingly, we observed that chaperone client protein evolve significantly slower than non-client protein (average $dN/dS_{client} = 0.047$, average $dN/dS_{non\text{-}client} = 0.051$, $P = 5.7 \times 10^{-6}$, Mann–Whitney *U* test, Fig. 1a). In general, many of these chaperones play a specific role, like assisting the transport of proteins across biological membranes or are promiscuous binders of proteins destabilized by stressful conditions to prevent their cytotoxic aggregation. Their mechanisms of action are poorly characterized, and they are very different from other more important, complex, and essential chaperones. Thus, it is important to analyze the precise chaperone client dataset to confer significant result. In *E. coli*, DnaK and GroEL play the essential role as protein folding chaperone, and they are well studied. Thus, we retrieved the experimentally validated chaperone clients of DnaK and GroEL from Calloni et al. (2012)) and Kerner et al. (2005) respectively as source of reliable chaperone client data, and observed that same trend persists within client and non-client proteins (average $dN/dS_{DnaK\ client} = 0.045$, average $dN/dS_{non\text{-}client} = 0.052$, $P = 1.6 \times 10^{-15}$, Mann–Whitney *U* test, Fig. 1b; average $dN/dS_{GroEL\ client} = 0.049$, average $dN/dS_{non\text{-}client} = 0.050$, $P = 2.4 \times 10^{-5}$, Mann–Whitney *U* test, Fig. 1c). Moreover, in *E. coli*, the GroEL chaperone clients are specifically characterized as partial, obligate, and non-specific clients according to their chaperone-dependent folding probability (Kerner et al. 2005). The obligate GroEL chaperone client proteins essentially require GroEL for the folding processes. Whereas, partial GroEL chaperone clients require GroEL whenever the proteins step into the misfolding state. There is yet another group of protein category, known as non-specific GroEL client protein. Irrespective of their folding requirements, they bind to GroEL present in the cellular environment. Interestingly, here we also observed that both obligate and partial chaperone clients evolve slower than non-client proteins in *E. coli* (average $dN/dS_{GroEL\ obligate\ client} = 0.041$, average $dN/dS_{non\text{-}client} = 0.052$, $P = 2.3 \times 10^{-3}$, Mann–Whitney *U* test; average $dN/dS_{GroEL\ partial\ client} = 0.045$, average $dN/dS_{non\text{-}client} = 0.052$, $P = 1.0 \times 10^{-3}$, Mann–Whitney *U* test, Fig. 1c), but no significant difference is observed in evolutionary rate between non-specific and non-client proteins (average $dN/dS_{GroEL\ non\text{-}specific\ client} = 0.085$, average $dN/dS_{non\text{-}client} = 0.052$, $P = 6.3 \times 10^{-1}$, Mann–Whitney *U* test, Fig. 1c). These results indicate that chaperone clients evolve slower than non-client proteins in *E. coli* irrespective of dataset.

To get a more general view, we chose four organisms, one prokaryote extremophile (*Thermus aquaticus*), one unicellular eukaryote (*Saccharomyces cerevisiae*), and two multicellular eukaryotes (*Drosophila melanogaster* and *Homo sapiens*) to analyze the evolutionary rate differences between chaperone

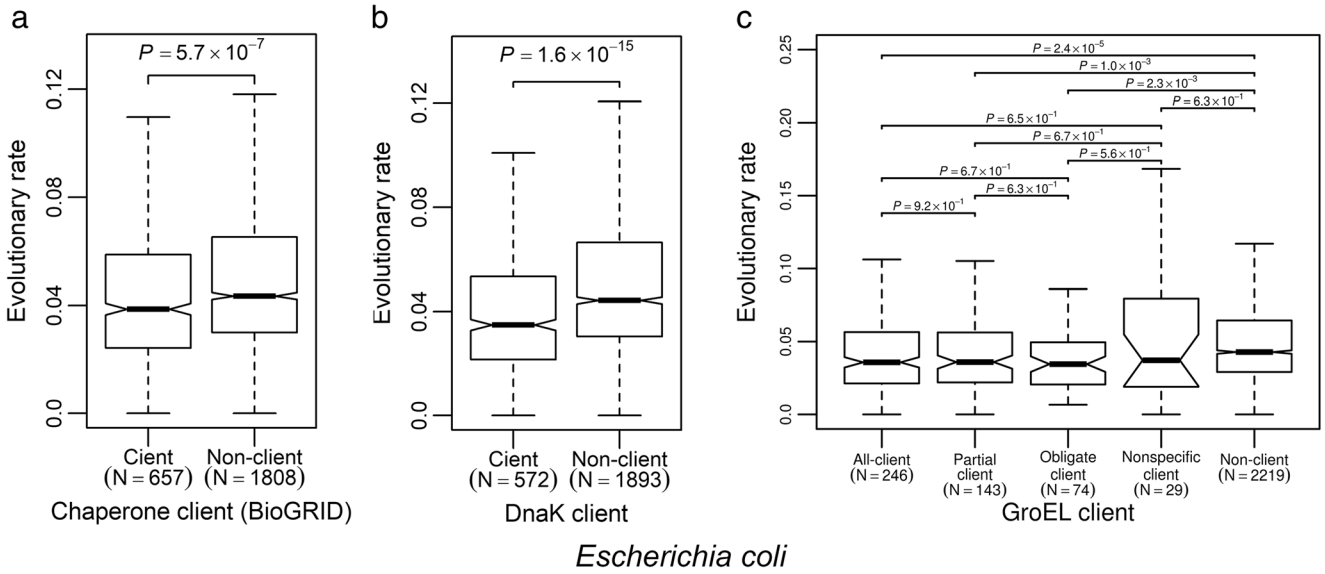

Fig. 1 Evolutionary rates of chaperone client and non-client protein in *E. coli*. **a** Chaperone clients were predicted using BioGRID database (version 3.4.152), **b** DnaK chaperone clients were retrieved from Calloni et al. (2012), **c** GroEL chaperone clients were retrieved from Kerner et al. (2005). Statistical significance is calculated by Mann–Whitney *U* test

client and non-client proteins (details in supplementary material). The similar trend follows in each of the organism (*T. aquaticus:* average $dN/dS_{client} = 0.044$, average $dN/dS_{non\text{-}client} = 0.059$, $P = 2.2 \times 10^{-13}$, Mann–Whitney *U* test; *S. cerevisiae:* average $dN/dS_{client} = 0.057$, average $dN/dS_{non\text{-}client} = 0.070$, $P = 2.7 \times 10^{-19}$, Mann–Whitney *U* test; *D. melanogaster:* average $dN/dS_{client} = 0.113$, average $dN/dS_{non\text{-}client} = 0.152$, $P = 6.2 \times 10^{-6}$, Mann–Whitney *U* test; *H. sapiens:* average $dN/dS_{client} = 0.272$, average $dN/dS_{non\text{-}client} = 0.357$, $P = 1.9 \times 10^{-29}$, Mann–Whitney *U* test, Fig. 2) as observed in *E. coli*. Therefore, it can be inferred that the chaperone client proteins inherently have slower evolutionary rates in prokaryotes as well as in eukaryotes.

Effect of protein abundance and network centrality in controlling the evolutionary rates

Previously, Williams and Fares also observed chaperone client proteins evolve slower than non-client proteins in *E. coli*, but they attributed this trend as an artifact of other factors that influence evolutionary rate (Williams and Fares 2010). Generally, gene/protein expression and network centrality play major role in controlling the evolutionary rates (Alvarez-Ponce et al. 2017; Drummond et al. 2006). Here, we downloaded *E. coli*, *S. cerevisiae*, *D. melanogaster*, and *H. sapiens* protein abundances from PaxDb database (version 4.1) (Wang et al. 2015) and protein-protein interaction

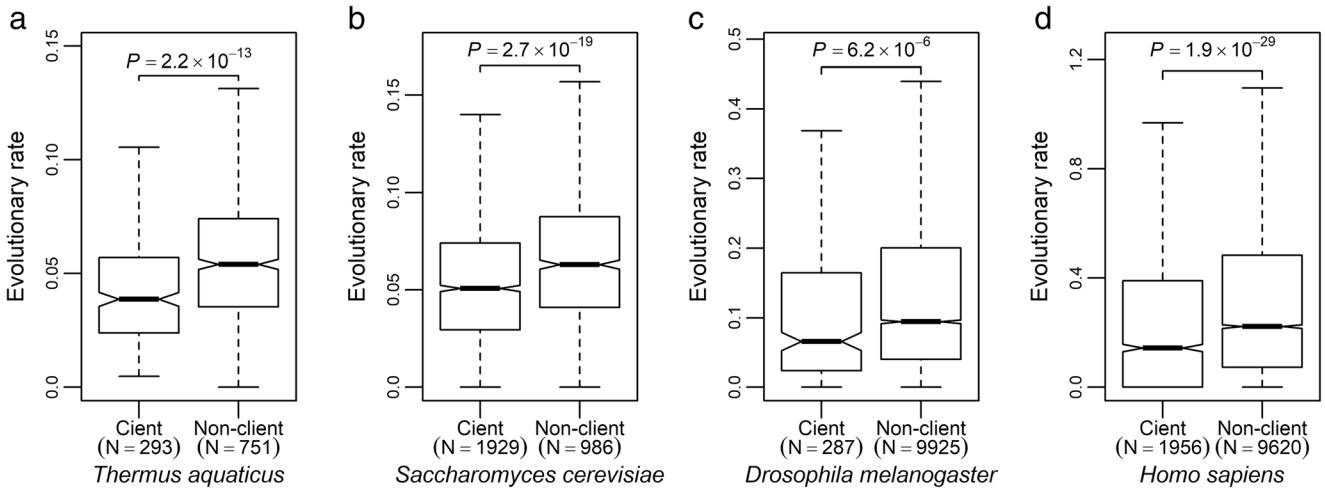

Fig. 2 Evolutionary rates of chaperone client and non-client protein in different organisms. **a** *T. aquaticus*, **b** *S. cerevisiae*, **c** *D. melanogaster*, and **d** *H. Sapiens*. Chaperone clients were predicted using STRING database (version 10.5) for *T. Aquaticus* and BioGRID database (version 3.4.152) for other three organisms. Statistical significance is calculated by Mann–Whitney *U* test

Table 1 Correlation between protein abundance and CAI in different organisms

Organism	Spearman correlation between protein abundance and Codon Adaptation Index
E. coli	$\rho = 0.552$, $N = 2460$, $P = 4.2 \times 10^{-196}$
S. cerevisiae	$\rho = 0.639$, $N = 2914$, $P < 1.0 \times 10^{-196}$
D. melanogaster	$\rho = 0.353$, $N = 4739$, $P = 4.2 \times 10^{-139}$
H. sapiens	$\rho = 0.081$, $N = 9602$, $P = 1.3 \times 10^{-15}$

network (PPIN) from BioGRID database (version 3.4.152) (Chatr-Aryamontri et al. 2017) and observed that both factors show significant correlation with evolutionary rates in *E. coli* (Spearman's $\rho_{\text{protein abundance – evolutionary rates}} = -0.224$, $P = 2.6 \times 10^{-29}$, $N = 2460$; Spearman's $\rho_{\text{PPIN – evolutionary rates}} = -0.108$, $P = 8.3 \times 10^{-8}$, $N = 2459$), *S. cerevisiae* (Spearman's $\rho_{\text{protein abundance – evolutionary rates}} = -0.277$, $P = 1.8 \times 10^{-52}$, $N = 2914$; Spearman's $\rho_{\text{PPIN – evolutionary rates}} = -0.195$, $P = 2.5 \times 10^{-26}$, $N = 2896$), *D. melanogaster* (Spearman's $\rho_{\text{protein abundance – evolutionary rates}} = -0.327$, $P = 8.7 \times 10^{-119}$, $N = 4739$; Spearman's $\rho_{\text{PPIN – evolutionary rates}} = -0.157$, $P = 4.6 \times 10^{-24}$, $N = 4099$), and *H. sapiens* (Spearman's $\rho_{\text{protein abundance – evolutionary rates}} = -0.228$, $P = 2.7 \times 10^{-113}$, $N = 9602$; Spearman's $\rho_{\text{PPIN – evolutionary rates}} = -0.186$, $P = 8.1 \times 10^{-76}$, $N = 9683$). However, the protein abundance data is unavailable in *T. aquaticus*, but Ghaemmaghami et al. (2003) observed Codon Adaptation Index (CAI) (Sharp and Li 1987) is strongly correlated with protein abundance and therefore can be used as an approximation of protein abundance. Here, we also observed significant positive correlations between protein abundance and CAI in all four organisms; the correlations were stronger in unicellular organisms probably due to their less complex cell type (Table 1). Thus, in this study, we used CAI as an approximation of protein abundance for *T. aquaticus* and observed protein-protein interaction (PPI) (calculated from STRING database (version 10.5) (Szklarczyk et al. 2017)) and CAI significantly correlated with evolutionary rates (Spearman's $\rho_{\text{CAI – evolutionary rates}} = -0.411$, $P = 9.8 \times 10^{-44}$, $N = 1044$; Spearman's $\rho_{\text{PPIN – evolutionary rates}} = -0.259$, $P = 1.8 \times 10^{-17}$, $N = 1044$). These results indicate that protein evolutionary rate is significantly controlled by protein-protein interaction and protein abundance. Interestingly, we also observed that chaperone client proteins have higher protein abundance and high protein-protein interactions than non-client proteins in all studied organisms (Table 2). Therefore, it is possible that the slower evolutionary rate of chaperone client proteins may be the artifacts of their higher protein abundance and high interaction as proposed by Williams and Fares (2010). To eliminate the effect of protein abundance, we randomly chose 300 chaperone client proteins from the DnaK client dataset of *E. coli* (since this dataset has a large number of experimentally curated client protein), and then we randomly chose 300 unique non-client proteins which correspond to each client protein with a very similar protein abundance level (i.e., the difference of protein abundance between each client and non-client protein must be less than 5% of the corresponding client protein) so that there is no statistical difference (95% level of confidence) of protein abundance between client and non-client protein. We measured the difference in evolutionary rates between chaperone client and non-client proteins with 95% level of confidence ($P < 0.05$, Mann–Whitney *U* test). Repeating this process for 1000 times, we found in 805 (~80.50%) cases the client proteins evolves slower (Mann–Whitney *U* test, $P < 0.05$) than the non-client proteins despite their similar protein abundance (Mann–Whitney *U* test, $P > 0.05$). Similarly, we controlled the PPI and observed in 994 (~99.40%) cases the client evolves slower (Mann–Whitney *U* test, $P < 0.05$) than the non-client proteins despite the similar protein-protein interaction (Mann–Whitney *U* test, $P > 0.05$). These two results indicate that the slower evolutionary rate of chaperone client protein is not the artifact of protein abundance or protein-protein interaction, but chaperone client imposes an intrinsic constraint on their protein evolution. But, in this method, we can control only one variable at a time between protein abundance and protein-protein interaction. To account for both the factors, we used normalized values of evolutionary rates, protein abundance, and protein-protein interaction in ANCOVA and observed chaperone client independently controlled ($P < 0.05$) the protein evolutionary rates in all organisms (Table 3).

Furthermore, we also performed principal component regression (PCR) analysis (using evolutionary rates as a dependent variable), which is considered a suitable method to establish the relative contributions of factors that influence protein evolution (Drummond et al. 2006). We observed these three parameters (protein abundance, PPI, and chaperone client) explained total (including all three principal components) 6.02% variance of evolutionary rates, in which protein abundance explained 2.55%, chaperone client type explained 1.81%, and PPI explained 1.66% in *E. coli* (DnaK client dataset) (Table 4). The trend is similar in other datasets of *E. coli*. Here, chaperone client type explains the comparable amount of the variation of evolutionary rates as protein abundance and PPI (Table 4). Moreover, in the other four organisms, we also observed consistent trends (Table 4). These results also

Table 2 Protein abundance and protein-protein interactions of chaperone client and non-client protein in different organisms

Organism	Group	Average protein abundance		Average protein-protein interaction	
E. coli	Chaperone client (BioGRID)	0.539 ($N = 657$)	$P = 4.8 \times 10^{-37}$	101.5 ($N = 657$)	$P = 3.8 \times 10^{-114}$
	Non-client	0.445 ($N = 1803$)		43.6 ($N = 1802$)	
	DnaK client	0.589 ($N = 572$)	$P = 1.6 \times 10^{-96}$	84.7 ($N = 572$)	$P = 1.1 \times 10^{-36}$
	Non-client	0.434 ($N = 1888$)		51.3 ($N = 1887$)	
	GroEL client	0.593 ($N = 246$)	$P = 1.9 \times 10^{-39}$	92.2 ($N = 246$)	$P = 1.3 \times 10^{-17}$
	Non-client	0.457 ($N = 2214$)		55.4 ($N = 2213$)	
T. aquaticus	Chaperone client (STRING)	0.765* ($N = 293$)	$P = 8.2 \times 10^{-8}$	105 ($N = 293$)	$P = 1.3 \times 10^{-51}$
	Non-client	0.746* ($N = 751$)		51.4 (N = 751)	
S. cerevisiae	Chaperone client (BioGRID)	0.618 ($N = 1929$)	$P = 8.3 \times 10^{-99}$	35.3 ($N = 1929$)	$P = 7.1 \times 10^{-86}$
	Non-client	0.517 ($N = 985$)		12.8 ($N = 967$)	
D. melanogaster	Chaperone client (BioGRID)	0.614 ($N = 147$)	$P = 7.4 \times 10^{-11}$	35.6 ($N = 203$)	$P = 1.0 \times 10^{-50}$
	Non-client	0.522 ($N = 4592$)		9.3 ($N = 3896$)	
H. sapiens	Chaperone client (BioGRID)	0.514 ($N = 1617$)	$P = 4.1 \times 10^{-124}$	78 ($N = 1996$)	$P < 1.0 \times 10^{-196}$
	Non-client	0.410 ($N = 7985$)		22.3 ($N = 7727$)	

Chaperone clients were predicted using STRING database (version 10.5) for *T. Aquaticus* and BioGRID database (version 3.4.152) for other organisms. Additionally, for *E. coli*, DnaK chaperone clients were retrieved from Calloni et al. (2012) and GroEL chaperone clients were retrieved from Kerner et al. (2005). Statistical significance calculated by Mann–Whitney *U* test

*Codon Adaptation Index is used as an alternative to protein abundance in *T. aquaticus*

indicate that slower evolutionary rate of chaperone client proteins is independent of the effect of other covariates.

Characteristics of chaperone client protein

Generally, the larger proteins (> 100 amino acids) reach their native state via different folding intermediate stages that act as "stepping stones" (Brockwell and Radford 2007), but chaperone client proteins become trapped into those folding intermediates and expose their hydrophobic surfaces that interact with chaperones to rescue their native structures (Hartl et al. 2011; Raineri et al. 2010). Even, the protein folding rate also helps to reach their native structures (Raineri et al. 2010), and may also influence the protein aggregation propensity (van den Berg et al. 1999). We calculated the protein aggregation propensity using TANGO algorithm (Fernandez-Escamilla et al.

Table 3 Results of ANCOVA

Organism	Dataset	Variables	F value
E. coli	Chaperone client (BioGRID)	Chaperone client type	24.73, $P = 7.0 \times 10^{-7}$
	Non-client	Protein abundance	119.07, $P = 4.2 \times 10^{-27}$
		Protein-protein interaction	7.57, $P = 5.6 \times 10^{-3}$
	DnaK client	Chaperone client type	53.19, $P = 4.0 \times 10^{-13}$
	Non-client	Protein abundance	95.00, $P = 4.8 \times 10^{-22}$
		Protein-protein interaction	9.03, $P = 2.7 \times 10^{-3}$
	GroEL different class client	Chaperone client type	6.64, $P = 3.7 \times 10^{-6}$
	Non-client	Protein abundance	120.29, $P = 2.3 \times 10^{-27}$
		Protein-protein interaction	10.35, $P = 1.3 \times 10^{-3}$
T. aquaticus	Chaperone client (STRING)	Chaperone client type	57.83, $P = 7.9 \times 10^{-14}$
	Non-client	Codon Adaptation Index	145.80, $P = 1.6 \times 10^{-31}$
		Protein-protein interaction	15.79, $P = 7.6 \times 10^{-5}$
S. cerevisiae	Chaperone client (BioGRID)	Chaperone client type	60.39, $P = 1.1 \times 10^{-14}$
	Non-client	Protein abundance	136.75, $P = 6.7 \times 10^{-31}$
		Protein-protein interaction	27.15, $P = 2.0 \times 10^{-7}$
D. melanogaster	Chaperone client (BioGRID)	Chaperone client type	6.52, $P = 1.1 \times 10^{-2}$
	Non-client	Protein abundance	425.84, $P = 3.3 \times 10^{-87}$
		Protein-protein interaction	18.38, $P = 1.9 \times 10^{-5}$
H. sapiens	Chaperone client (BioGRID)	Chaperone client type	71.14, $P = 3.9 \times 10^{-17}$
	Non-client	Protein abundance	320.16, $P = 3.0 \times 10^{-70}$
		Protein-protein interaction	71.48, $P = 3.3 \times 10^{-17}$

Table 4 Principal component regression analysis performed using evolutionary rates as a dependent variable

Organism	Group	Variable	*F* statistics	Variation explained (in %)			
				PC1	PC2	PC3	Total
E. coli	Chaperone client (BioGRID)	Chaperone client type	$F = 50.46$ $P = 1.2 \times 10^{-31}$	1.40	0.00	0.00	1.40
	Non-client	Protein abundance		1.10	1.44	0.01	2.55
		Protein-protein interaction		1.64	0.20	0.00	1.84
E. coli	DnaK client	Chaperone client type	$F = 52.41$ $P = 7.9 \times 10^{-33}$	1.77	0.00	0.04	1.81
	Non-client	Protein abundance		2.08	0.00	0.47	2.55
		Protein-protein interaction		1.61	0.00	0.05	1.66
E. coli	GroEL client	Chaperone client type	$F = 51.08$ $P = 5.1 \times 10^{-32}$	1.25	0.01	0.02	1.28
	Non-client	Protein abundance		2.03	0.07	0.57	2.67
		Protein-protein interaction		1.78	0.08	0.06	1.92
T. aquaticus	Chaperone client (STRING)	Chaperone client type	$F = 72.99$ $P = 7.6 \times 10^{-43}$	5.10	0.27	0.00	5.37
	Non-client	Codon Adaptation Index		2.65	3.25	0.01	5.91
		Protein-protein interaction		5.52	0.59	0.00	6.11
S. cerevisiae	Chaperone client (BioGRID)	Chaperone client type	$F = 74.77$ $P = 1.4 \times 10^{-46}$	1.98	0.01	0.00	1.99
	Non-client	Protein abundance		2.27	0.23	0.17	2.67
		Protein-protein interaction		2.22	0.26	0.05	2.53
D. melanogaster	Chaperone client (BioGRID)	Chaperone client type	$F = 153.20$ $P = 8.0 \times 10^{-91}$	2.49	0.16	0.00	2.65
	Non-client	Protein abundance		3.35	3.41	0.82	7.58
		Protein-protein interaction		4.31	1.49	0.07	5.87
H. sapiens	Chaperone client (BioGRID)	Chaperone client type	$F = 154.30$ $P = 3.3 \times 10^{-97}$	1.20	0.00	0.00	1.20
	Non-client	Protein abundance		1.64	0.33	0.00	1.97
		Protein-protein interaction		2.04	0.22	0.00	2.26

Principal component regression is performed with evolutionary rates as dependent variable and chaperone client type, protein abundance, and protein-protein interaction as independent variables. We used principal component regression model pcr(Evolutionary rates ~ chaperone client type + protein abundance (or, Codon Adaptation Index) + protein-protein interaction)

2004), and compared the aggregation propensity between chaperone client and non-client proteins in all five organisms. We observed that the chaperone client proteins have low aggregation propensity than non-client proteins in all organisms (Table 5).

Using the sequence-based protein folding rates prediction by SeqRate (Lin et al. 2010), we calculated the folding rate of the protein. Due to computational limitations, we calculated the folding rate of proteins only in *E. coli* and *S. cerevisiae*. Interestingly, we observed chaperone client proteins also show a slow folding rate than non-client proteins (*E. coli*: average folding $\text{rate}_{\text{client}} = 420.39$ protein/s, average folding $\text{rate}_{\text{non-client}} = 455$ protein/s; Mann–Whitney *U* test, $P = 7.0 \times 10^{-4}$; *S. cerevisiae*: average folding $\text{rate}_{\text{client}} = 151.63$ protein/s, average folding $\text{rate}_{\text{non-client}} = 417.72$ protein/s; Mann–Whitney *U* test,

Table 5 Protein aggregation of chaperone client and non-client protein in different organisms

Organism	Group	Average protein aggregation	
E. coli	Chaperone client (BioGRID)	1934.18 ($N = 657$)	$P = 1.1 \times 10^{-2}$
	Non-client	2902.06 ($N = 1808$)	
	DnaK client	1163.86 ($N = 572$)	$P = 1.3 \times 10^{-18}$
	Non-client	3091.36 ($N = 1893$)	
	GroEL client	977.72 ($N = 246$)	$P = 2.0 \times 10^{-13}$
	Non-client	2828.82 ($N = 2219$)	
T. aquaticus	Chaperone client (STRING)	1798.51 ($N = 293$)	$P = 1.1 \times 10^{-2}$
	Non-client	2270.89 ($N = 751$)	
S. cerevisiae	Chaperone client (BioGRID)	214.13 ($N = 1929$)	$P = 1.5 \times 10^{-10}$
	Non-client	393.94 ($N = 986$)	
D. melanogaster	Chaperone client (BioGRID)	256.73 ($N = 287$)	$P = 4.0 \times 10^{-9}$
	Non-client	498.07 ($N = 9900$)	
H. sapiens	Chaperone client (BioGRID)	2367.94 ($N = 1948$)	$P = 1.8 \times 10^{-12}$
	Non-client	3229.81 ($N = 9559$)	

Chaperone clients were predicted using STRING database (version 10.5) for *T. Aquaticus* and BioGRID database (version 3.4.152) for other organisms. Additionally, for *E. coli*, DnaK chaperone clients were retrieved from Calloni et al. (2012) and GroEL chaperone clients were retrieved from Kerner et al. (2005). Statistical significance is calculated by Mann–Whitney *U* test

$P = 4.5 \times 10^{-4}$). Earlier, Raineri et al. (2010) also expected that proteins with slower folding rate may show low aggregation propensity. Thus, accumulation of folding intermediates with exposed hydrophobic regions may increase the overall aggregation when translated in a chaperone-free system (Niwa et al. 2009).

Conclusion

In general, mutations are expected to occur at the same rate in the same species in wild condition as well as in the laboratory condition, but reduced population size in the laboratory strain may result in reduced efficacy of natural selection (Charlesworth 2009) which may allow the fixation of slightly deleterious mutations (Ohta 1973), especially when there is an added quality control mechanism in the system. Thus, chaperone client proteins might tolerate mutational perturbation in chaperone-rich media in the earlier experiment (Aguilar-Rodríguez et al. 2016). This observation indicates that chaperones may buffer slightly deleterious mutations, and thus these mutations become neutral and can be tolerated (probably with the reduced functionality). However, in natural condition, mutated polypeptides may increase misfolding and that may become a permanent burden for the cellular system in terms of the energetic cost. For this reason, the slightly disadvantageous random mutations are less likely to be selected in the natural population. Therefore, chaperones may promote genetic variation (Moran 1996) but this transient variation stands for shorter duration compared with the evolutionary time scale.

Materials and methods

Evolutionary rates

For each *Escherichia coli* genes, we identified one-to-one orthologous genes from *Salmonella enterica* using reciprocal best hits with cutoff *E*-value $\leq 10^{-5}$, gap < 5, and identity ≥ 80% by the BlastP algorithm (Altschul et al. 1997). The same methods were used to retrieve *Thermus aquaticus–Thermus scotoductus*, *Saccharomyces cerevisiae–Saccharomyces bayanus*, *Drosophila melanogaster–Drosophila simulans*, *and Homo sapiens–Pan troglodytes* orthologous genes. We downloaded all CDS of *Escherichia coli* (GCF_000005845.2_ASM584v2), *Salmonella enteric* (GCF_000195995.1_ASM19599v1), *Thermus aquaticus* (GCF_001399775.1_ASM139977v1), *Thermus scotoductus* (GCF_000187005.1_ASM18700v1), *Saccharomyces cerevisiae* (GCF_000146045.2_R64), *Drosophila melanogaster* (GCF_000001215.4_Release_6_plus_ISO1_MT), *Drosophila simulans* (GCF_000754195.2_ASM75419v2), *Homo sapiens* (GCF_000001405.38_GRCh38.p12), and *Pan troglodytes* (GCF_002880755.1_Clint_PTRv2) from NCBI RefSeq (Pruitt et al. 2007) and ORFs of *Saccharomyces bayanus* were downloaded from *Saccharomyces* Genome Database (https://downloads.yeastgenome.org/sequence/fungi/S_bayanus/archive/MIT/orf_dna/orf_genomic.fasta.gz) (Cherry et al. 2012). Each orthologous protein pairs were aligned by ClustalW (Thompson et al. 2002) and the corresponding CDS pairs were aligned by pal2nal algorithm (Suyama et al. 2006) using

the protein alignment as a template. For each of the resulting alignments, the evolutionary rates (nonsynonymous (dN) to synonymous ratio (dS)) were estimated with CodeML in PAML (version 4.9) (Yang 2007).

Protein-protein interaction data

We downloaded *E. coli*, *S. cerevisiae*, *D. melanogaster*, and *H. sapiens* protein-protein interactions from BioGRID database (version 3.4.152) (Chatr-Aryamontri et al. 2017), and *T. aquaticus* protein-protein interactions from STRING database (version10.5) (Szklarczyk et al. 2017) with confidence score ≥ 40% (Chakraborty and Alvarez-Ponce 2016). We only considered physical interactions and removed the self-interaction from our dataset.

Chaperone client proteins

We collected the chaperones of *E. coli* and *T. aquaticus* from Uniprot (Bateman et al. 2017), *S. cerevisiae* from Gong et al. (2009), *D. melanogaster* from Sorensen et al. (2005) and Tower (2011), and *H. sapiens* from multiple sources (Apweiler et al. 2004; Arakawa et al. 2010; Burdette et al. 2010; Chen et al. 2005; Chung et al. 2002; Dafforn et al. 2001; Hasson et al. 2013; Hietakangas et al. 2006; Kampinga et al. 2009; Lamb et al. 2000; Mymrikov et al. 2017; Myung et al. 2004; Nagaraj et al. 2012; Nisemblat et al. 2015; Qiu et al. 2006; Rabindran et al. 1991; Rauch and Gestwicki 2014; Seo et al. 2010; Sheldon and Kingston 1994; Takayama et al. 1999; Tsao et al. 2006; Vainberg et al. 1998; Wheeler and Jia 2015; Yoshida et al. 2001). We downloaded *E. coli*, *S. cerevisiae*, *D. melanogaster*, and *H. sapiens* protein-protein interactions from BioGRID database (version 3.4.152) (Chatr-Aryamontri et al. 2017), and *T. aquaticus* protein-protein interactions with confidence score ≥ 40% (Chakraborty and Alvarez-Ponce 2016) from STRING database (version 3.4.152) (Szklarczyk et al. 2017). Then we mapped *E. coli*, *T. aquaticus*, *S. cerevisiae*, *D. melanogaster*, and *H. sapiens* chaperones into their corresponding protein-protein interaction partners to retrieve chaperone client proteins in each species. We also downloaded experimentally validated DnaK and GroEL client data form Calloni et al. (2012) and Kerner et al. (2005).

Protein abundance data

We downloaded integrated protein abundance data of *Escherichia coli*, *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and *Homo sapiens* from PaxDb database (Wang et al. 2015). For *Thermus aquaticus*, we calculated the Codon Adaptation Index (CAI) (Sharp and Li 1987) as an alternative representation of protein abundance using in-house PERL script.

Protein folding rate

We used SeqRate (Lin et al. 2010) algorithm from the MULTICOM toolbox (http://sysbio.rnet.missouri.edu/multicom_toolbox/) (Cheng et al. 2012) to calculate protein folding. SeqRate uses protein sequences to predict protein folding rate with support vector machine (Lin et al. 2010).

Protein aggregation propensity

Aggregation propensity of proteins was calculated using TANGO algorithm (Fernandez-Escamilla et al. 2004). TANGO uses protein sequences to predict β-aggregation propensity score.

Statistical analysis

We used Mann–Whitney *U* test to compare the significant difference between the two groups. To calculate correlation, we used Spearman's rank correlation. We performed ANCOVA and principal component regression (PCR) analysis using evolutionary rates as the dependent variable taking chaperone client type, protein abundance (or, Codon Adaptation Index), and protein-protein interaction as the independent variables. Generally, evolutionary rates and protein abundance are exponentially distributed and protein-protein interaction is scale-free distributed. Thus, to perform ANCOVA and PCR, we transformed these three variables to fit the normal distribution. We used R language and environment (https://www.r-project.org/) to perform all statistical analyses.

ANCOVA is performed with evolutionary rates as dependent variable and chaperone client type, protein abundance, and protein-protein interaction as independent variables. We used ANCOVA model lm (evolutionary rates ~ chaperone client type + protein abundance (or Codon Adaptation Index) + protein-protein interaction)

Authors' contributions M.P.V., D.A., and S.C. designed the study. M.A.P. and S.C. performed the analysis. M.P.V., D.A., and S.C. drafted the manuscript. S.C. and T.C.G. completed the final version of the manuscript. All authors read and approved the final manuscript.

Funding information The authors received financial assistance from the Bose Institute and Centre of Excellence in Bioinformatics.

Compliance with ethical standards

Competing interests The authors have declared that they have no competing interests.

References

Aguilar-Rodríguez J, Sabater-Muñoz B, Montagud-Martínez R, Berlanga V, Alvarez-Ponce D, Wagner A, Fares MA (2016) The molecular chaperone DnaK is a source of mutational robustness. Genome Biol Evol 8:2979–2991. https://doi.org/10.1093/gbe/evw176

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402

Alvarez-Ponce D, Feyertag F, Chakraborty S (2017) Position matters: network centrality considerably impacts rates of protein evolution in the human protein-protein interaction network. Genome Biol Evol 9:1742–1756. https://doi.org/10.1093/gbe/evx117

Apweiler R et al (2004) UniProt: the universal protein knowledgebase. Nucleic Acids Res 32:D115–D119. https://doi.org/10.1093/nar/gkh131

Arakawa A et al (2010) The C-terminal BAG domain of BAG5 induces conformational changes of the Hsp70 nucleotide-binding domain for ADP-ATP exchange. Structure 18:309–319. https://doi.org/10.1016/j.str.2010.01.004

Bateman A et al (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45:D158–D169. https://doi.org/10.1093/nar/gkw1099

Brockwell DJ, Radford SE (2007) Intermediates: ubiquitous species on folding energy landscapes? Curr Opin Struct Biol 17:30–37. https://doi.org/10.1016/j.sbi.2007.01.003

Burdette AJ, Churchill PF, Caldwell GA, Caldwell KA (2010) The early-onset torsion dystonia-associated protein, torsinA, displays molecular chaperone activity in vitro. Cell Stress Chaperones 15:605–617. https://doi.org/10.1007/s12192-010-0173-2

Calloni G et al (2012) DnaK functions as a central hub in the E. coli chaperone network. Cell Rep 1:251–264. https://doi.org/10.1016/j.celrep.2011.12.007

Chakraborty S, Alvarez-Ponce D (2016) Positive selection and centrality in the yeast and fly protein-protein interaction networks. Biomed Res Int. https://doi.org/10.1155/2016/4658506

Charlesworth B (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. Nat Rev Genet 10:195–205. https://doi.org/10.1038/nrg2526

Chatr-Aryamontri A et al (2017) The BioGRID interaction database: 2017 update. Nucleic Acids Res 45:D369–D379. https://doi.org/10.1093/nar/gkw1102

Chen B, Piel WH, Gui LM, Bruford E, Monteiro A (2005) The HSP90 family of genes in the human genome: insights into their divergence and evolution. Genomics 86:627–637. https://doi.org/10.1016/j.ygeno.2005.08.012

Cheng JL, Li JL, Wang Z, Eickholt J, Deng X (2012) The MULTICOM toolbox for protein structure prediction. Bmc Bioinform 13. https://doi.org/10.1186/1471-2105-13-65

Cherry JM et al (2012) Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res 40:D700–D705. https://doi.org/10.1093/nar/gkr1029

Chung KT, Shen Y, Hendershot LM (2002) BAP, a mammalian BiP-associated protein, is a nucleotide exchange factor that regulates the ATPase activity of BiP. J Biol Chem 277:47557–47563. https://doi.org/10.1074/jbc.M208377200

Dafforn TR, Della M, Miller AD (2001) The molecular interactions of heat shock protein 47 (Hsp47) and their implications for collagen biosynthesis. J Biol Chem 276:49310–49319. https://doi.org/10.1074/jbc.M108896200

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

Drummond DA, Raval A, Wilke CO (2006) A single determinant dominates the rate of yeast protein evolution. Mol Biol Evol 23:327–337. https://doi.org/10.1093/molbev/msj038

Estruch F (2000) Stress-controlled transcription factors, stress-induced genes and stress tolerance in budding yeast. FEMS Microbiol Rev 24:469–486

Ewalt KL, Hendrick JP, Houry WA, Hartl FU (1997) In vivo observation of polypeptide flux through the bacterial chaperonin system. Cell 90:491–500

Fares MA, Ruiz-González MX, Moya A, Elena SF, Barrio E (2002) Endosymbiotic bacteria: groEL buffers against deleterious mutations. Nature 417:398. https://doi.org/10.1038/417398a

Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol 22:1302–1306. https://doi.org/10.1038/nbt1012

Geiler-Samerotte KA, Zhu YO, Goulet BE, Hall DW, Siegal ML (2016) Selection transforms the landscape of genetic variation interacting with Hsp90. PLoS Biol 14:e2000465. https://doi.org/10.1371/journal.pbio.2000465

Georgopoulos C (1992) The emergence of the chaperone machines. Trends Biochem Sci 17:295–299

Ghaemmaghami S et al (2003) Global analysis of protein expression in yeast. Nature 425:737–741. https://doi.org/10.1038/nature02046

Goldberg AL (2003) Protein degradation and protection against misfolded or damaged proteins. Nature 426:895–899. https://doi.org/10.1038/nature02263

Goldberg AL, Dice JF (1974) Intracellular protein degradation in mammalian and bacterial cells. Annu Rev Biochem 43:835–869. https://doi.org/10.1146/annurev.bi.43.070174.004155

Gong Y, Kakihara 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. https://doi.org/10.1038/msb.2009.26

Hartl FU, Hayer-Hartl M (2009) Converging concepts of protein folding in vitro and in vivo. Nat Struct Mol Biol 16:574–581. https://doi.org/10.1038/nsmb.1591

Hartl FU, Bracher A, Hayer-Hartl M (2011) Molecular chaperones in protein folding and proteostasis. Nature 475:324–332. https://doi.org/10.1038/nature10317

Hasson SA et al (2013) High-content genome-wide RNAi screens identify regulators of parkin upstream of mitophagy. Nature 504:291. https://doi.org/10.1038/nature12748

Hietakangas V, Anckar J, Blomster HA, Fujimoto M, Palvimo JJ, Nakai A, Sistonen L (2006) PDSM, a motif for phosphorylation-dependent SUMO modification. Proc Natl Acad Sci U S A 103:45–50. https://doi.org/10.1073/pnas.0503698102

Kampinga HH et al (2009) Guidelines for the nomenclature of the human heat shock proteins. Cell Stress Chaperones 14:105–111. https://doi.org/10.1007/s12192-008-0068-7

Kerner MJ et al (2005) Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. Cell 122:209–220. https://doi.org/10.1016/j.cell.2005.05.028

Lamb AL, Wernimont AK, Pufahl RA, O'Halloran TV, Rosenzweig AC (2000) Crystal structure of the second domain of the human copper chaperone for superoxide dismutase. Biochemistry 39:1589–1595. https://doi.org/10.1021/bi992822i

Lin GN, Wang Z, Xu D, Cheng J (2010) SeqRate: sequence-based protein folding type classification and rates prediction. BMC Bioinformatics 11 Suppl 3:S1. https://doi.org/10.1186/1471-2105-11-S3-S1

McDonough H, Patterson C (2003) CHIP: a link between the chaperone and proteasome systems. Cell Stress Chaperones 8:303–308

Moran NA (1996) Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc Natl Acad Sci U S A 93:2873–2878

Mymrikov EV, Daake M, Richter B, Haslbeck M, Buchner J (2017) The chaperone activity and substrate spectrum of human small heat shock proteins. J Biol Chem 292:672–684. https://doi.org/10.1074/jbc.M116.760413

Myung JK, Afjehi-Sadat L, Felizardo-Cabatic M, Slavc I, Lubec G (2004) Expressional patterns of chaperones in ten human tumor cell lines. Proteome Sci 2:8. https://doi.org/10.1186/1477-5956-2-8

Nagaraj RH et al (2012) Acetylation of αA-crystallin in the human lens: effects on structure and chaperone function. Biochim Biophys Acta 1822:120–129. https://doi.org/10.1016/j.bbadis.2011.11.011

Nisemblat S, Yaniv O, Parnas A, Frolow F, Azem A (2015) Crystal structure of the human mitochondrial chaperonin symmetrical football complex. Proc Natl Acad Sci U S A 112:6044–6049. https://doi.org/10.1073/pnas.1411718112

Niwa T, Ying BW, Saito K, Jin W, Takada S, Ueda T, Taguchi H (2009) Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of *Escherichia coli* proteins. Proc Natl Acad Sci U S A 106:4201–4206. https://doi.org/10.1073/pnas.0811922106

Ohta T (1973) Slightly deleterious mutant substitutions in evolution. Nature 246:96–98

Park JY, Kim GH, Kim SS, Ko JM, Lee JJ, Yoo HW (2009) Effects of a chemical chaperone on genetic mutations in alpha-galactosidase A in Korean patients with Fabry disease. Exp Mol Med 41:1–7. https://doi.org/10.3858/emm.2009.41.1.001

Pruitt KD, Tatusova T, Maglott DR (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35:D61–D65. https://doi.org/10.1093/nar/gkl842

Qiu XB, Shao YM, Miao S, Wang L (2006) The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. Cell Mol Life Sci 63:2560–2570. https://doi.org/10.1007/s00018-006-6192-6

Rabindran SK, Giorgi G, Clos J, Wu C (1991) Molecular-cloning and expression of a human heat-shock factor, HSF1. Proc Natl Acad Sci U S A 88:6906–6910. https://doi.org/10.1073/pnas.88.16.6906

Raineri E, Ribeca P, Serrano L, Maier T (2010) A more precise characterization of chaperonin substrates. Bioinformatics 26:1685–1689. https://doi.org/10.1093/bioinformatics/btq287

Rauch JN, Gestwicki JE (2014) Binding of human nucleotide exchange factors to heat shock protein 70 (Hsp70) generates functionally distinct complexes in vitro. J Biol Chem 289:1402–1414. https://doi.org/10.1074/jbc.M113.521997

Rutherford SL (2003) Between genotype and phenotype: protein chaperones and evolvability. Nat Rev Genet 4:263–274. https://doi.org/10.1038/nrg1041

Rutherford SL, Lindquist S (1998) Hsp90 as a capacitor for morphological evolution. Nature 396:336–342. https://doi.org/10.1038/24550

Sabater-Munoz B et al (2015) Fitness trade-offs determine the role of the molecular chaperonin groel in buffering mutations. Mol Biol Evol 32:2681–2693. https://doi.org/10.1093/molbev/msv144

Schubert U, Antón LC, Gibbs J, Norbury CC, Yewdell JW, Bennink JR (2000) Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. Nature 404:770–774. https://doi.org/10.1038/35008096

Schwanhäusser B et al (2011) Global quantification of mammalian gene expression control. Nature 473:337–342. https://doi.org/10.1038/nature10098

Seo S, Baye LM, Schulz NP, Beck JS, Zhang QH, Slusarski DC, Sheffield VC (2010) BBS6, BBS10, and BBS12 form a complex with CCT/TRiC family chaperonins and mediate BBSome assembly. Proc Natl Acad Sci U S A 107:1488–1493. https://doi.org/10.1073/pnas.0910268107

Sharp PM, Li WH (1987) The Codon Adaptation Index–a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15:1281–1295

Sheldon LA, Kingston RE (1994) Hydrophobic coiled-coil domains regulate the subcellular-localization of human heat-shock factor-2 VOL 7, PG 1549-1588, 1993. Genes Dev 8:386–386

Sherman MY, Goldberg AL (1992) Involvement of the chaperonin dnaK in the rapid degradation of a mutant protein in *Escherichia coli*. EMBO J 11:71–77

Singh LR, Chen X, Kozich V, Kruger WD (2007) Chemical chaperone rescue of mutant human cystathionine beta-synthase. Mol Genet Metab 91:335–342. https://doi.org/10.1016/j.ymgme.2007.04.011

Sorensen JG, Nielsen MM, Kruhoffer M, Justesen J, Loeschcke V (2005) Full genome gene expression analysis of the heat stress response, in Drosophila melanogaster. Cell Stress Chaperones 10:312–328. https://doi.org/10.1379/csc-128r1.1

Stouthamer AH (1973) A theoretical study on the amount of ATP required for synthesis of microbial cell material. Antonie Van Leeuwenhoek 39:545–565

Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609–W612. https://doi.org/10.1093/nar/gkl315

Suzuki Y (2014) Emerging novel concept of chaperone therapies for protein misfolding diseases. Proc Jpn Acad Ser B Phys Biol Sci 90:145–162. https://doi.org/10.2183/pjab.90.145

Szklarczyk D et al (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 45:D362–D368. https://doi.org/10.1093/nar/gkw937

Takayama S, Xie ZH, Reed JC (1999) An evolutionarily conserved family of Hsp70/Hsc70 molecular chaperone regulators. J Biol Chem 274:781–786. https://doi.org/10.1074/jbc.274.2.781

Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al] Chapter 2

Tokuriki N, Tawfik DS (2009) Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature 459:668–U671. https://doi.org/10.1038/nature08009

Tomala K, Korona R (2008) Molecular chaperones and selection against mutations. Biol Direct 3. https://doi.org/10.1186/1745-6150-3-5

Tower J (2011) Heat shock proteins and Drosophila aging. Exp Gerontol 46:355–362. https://doi.org/10.1016/j.exger.2010.09.002

Tsao ML, Chao CH, Yeh CT (2006) Interaction of hepatitis C virus F protein with prefoldin 2 perturbs tubulin cytoskeleton organization. Biochem Biophys Res Commun 348:271–277. https://doi.org/10.1016/j.bbrc.2006.07.062

Vainberg IE, Lewis SA, Rommelaere H, Ampe C, Vandekerckhove J, Klein HL, Cowan NJ (1998) Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. Cell 93:863–873. https://doi.org/10.1016/s0092-8674(00)81446-4

van den Berg B, Ellis RJ, Dobson CM (1999) Effects of macromolecular crowding on protein folding and aggregation. EMBO J 18:6927–6933. https://doi.org/10.1093/emboj/18.24.6927

Wagner A (2005) Energy constraints on the evolution of gene expression. Mol Biol Evol 22:1365–1374. https://doi.org/10.1093/molbev/msi126

Wang M, Herrmann CJ, Simonovic M, Szklarczyk D, von Mering C (2015) Version 4.0 of PaxDb: protein abundance data, integrated across model organisms, tissues, and cell-lines. Proteomics 15:3163–3168. https://doi.org/10.1002/pmic.201400441

Wheeler B, Jia ZC (2015) Preparation and characterization of human ADCK3, a putative atypical kinase. Protein Expr Purif 108:13–17. https://doi.org/10.1016/j.pep.2014.12.008

Wickner S, Maurizi MR, Gottesman S (1999) Posttranslational quality control: folding, refolding, and degrading proteins. Science 286: 1888–1893

Williams TA, Fares MA (2010) The effect of chaperonin buffering on protein evolution. Genome Biol Evol 2:609–619. https://doi.org/10.1093/gbe/evq045

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24. https://doi.org/10.1093/molbev/msm088

Yoshida H, Okada T, Haze K, Yanagi H, Yura T, Negishi M, Mori K (2001) Endoplasmic reticulum stress-induced formation of transcription factor complex ERSF including NF-Y (CBF) and activating transcription factors 6 alpha and 6 beta that activates the mammalian unfolded protein response. Mol Cell Biol 21:1239–1248. https://doi.org/10.1128/mcb.21.4.1239-1248.2001