BIOINFORMATICS DISCOVERY NOTE Vol. 26 no. 14 2010, pages 1685–1689
doi:10.1093/bioinformatics/btq287

Sequence analysis

Advance Access publication June 2, 2010

A more precise characterization of chaperonin substrates

Emanuele Raineri[1,2,†], Paolo Ribeca[1,2,†], Luis Serrano[3] and Tobias Maier[3,*]

[1]CNAG Centro Nacional de Análisis Genómico, Parc Cientific de Barcelona, Baldiri Reixac 4, E-08028 Barcelona, [2]Center for Genomic Regulation, Bioinformatics and Genomics Unit and [3]Center for Genomic Regulation, Systems Biology Unit, Calle Dr. Aiguader 88, E-08003 Barcelona, Spain

Associate Editor: Burkhard Rost

ABSTRACT

Motivation: Molecular chaperones prevent the aggregation of their substrate proteins and thereby ensure that they reach their functional native state. The bacterial GroEL/ES chaperonin system is understood in great detail on a structural, mechanistic and functional level; its interactors in *Escherichia coli* have been identified and characterized. However, a long-standing question in the field is: What makes a protein a chaperone substrate?

Results: Here we identify, using a bioinformatics-based approach a simple set of quantities, which characterize the GroEL-substrate proteome. We define three novel parameters differentiating GroEL interactors from other cellular proteins: lower rate of evolution, hydrophobicity and aggregation propensity. Combining them with other known features to a simple Bayesian predictor allows us to identify known homologous and heterologous GroEL substrateproteins. We discuss our findings in relation to established mechanisms of protein folding and evolutionary buffering by chaperones.

Contact: tobias.maier@crg.es

Supplementary information: Supplementary data is available at *Bioinformatics* online.

Received on February 28, 2010; revised on May 10, 2010; accepted on May 27, 2010

1 INTRODUCTION

Molecular chaperone systems and their interaction with substrate proteins are a fascinating paradigm in molecular biology. Chaperones assist newly synthesized polypeptide chains to fold and mature to functional proteins: they are able to rescue substrates trapped in energetically unfavorable states (Hartl and Hayer-Hartl, 2002), and are markedly over-expressed under cellular stress to prevent the aggregation of unfolded proteins. In fact, one could say chaperone substrates are proteins on the edge, often carrying out important cellular functions but relying on helper proteins to reach their native state.

Arguably, the best studied chaperone system are the bacterial Hsp60/Hsp10 (GroEL/ES) heat shock proteins (Fenton and Horwich, 1997; Hartl and Hayer-Hartl, 2002; Xu *et al.*, 1997). GroES functions as a lid-like structure, allowing for the full encapsulation of substrate proteins inside the GroEL cavity, thereby providing a secluded folding environment; productive folding can involve various rounds of binding to and release from the chaperonin (Weissman *et al.*, 1994). The *in vivo* substrate proteins of *Escherichia coli* GroEL have recently been identified and characterized experimentally (Chapman *et al.*, 2006; Kerner *et al.*, 2005); between 250 and 320 proteins interact with the chaperone in the cell under regular growth conditions with different interaction propensities.

In spite of these sophisticated experimental studies, a few specific features of chaperonin substrates have been described so far. Due to the limited size of the GroEL cavity, their molecular weights range predominantly between 20 and 60 kDa (Kerner *et al.*, 2005; Sakikawa *et al.*, 1999); they show a specific enrichment in certain fold types, such as the TIM barrel (Kerner *et al.*, 2005); they show lower average folding propensities and higher translation efficiencies (Noivirt-Brik *et al.*, 2005). In addition, an attempt has been made to identify specific chaperone-binding motives in the substrate proteins, after the idea that they would resemble the GroES binding loop (Stan *et al.*, 2006); however, the picture is still far from being complete.

In this article, we establish novel parameters differentiating them from other cellular proteins, such as lower rate of evolution, lower hydrophobicity and aggregation propensity; furthermore, we re-analyze some known factors and examine their relation to our newly determined ones. As a final consistency test, we show that the quantities delineated by this study can be combined into prediction algorithms, allowing at least some degree of quantitative discrimination of GroEL substrates from chaperonin-independent folders.

2 METHODS

We analyzed specific properties of various subsets of proteins of *E.coli* known to interact with the chaperone GroEL *in vivo*, as from the lists published in Kerner *et al.* (2005) and Chapman *et al.* (2006). In Kerner *et al.* (2005), the authors isolated stabilized GroEL–GroES-substrate complexes; a quantitative mass-spectrometry approach allowed the identified 252 GroEL substrates to be sorted in three classes according to their abundance in complex with the chaperone, relative to their native levels in an *E.coli* cell lysate and based on experimental testing of chaperone usage for several example proteins (see the Supplementary Material of Kerner *et al.*, 2005, for a full list). GroEL substrates in *E.coli* have more recently been identified by a second independent study (Chapman *et al.*, 2006) from inclusion bodies after depletion of cellular GroEL levels, yielding 317 proteins; 136 proteins are common to both datasets (Noivirt-Brik *et al.*, 2005). Our analysis is based both on various combinations of GroEL substrate proteins belonging to Classes I, II and III of Kerner *et al.* (2005), and on the full set of identified GroEL substrates of Chapman *et al.* (2006).

Statistical setup: to assess the statistical significance of the differences between the GroEL substrates and the entire proteome, we consistently adopted the same procedure already used in Noivirt-Brik *et al.* (2005), which

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

is independent of the specific feature one is looking at, and hinges on the fact that the substrates are nothing but a particular subset of the proteome of cardinality n: (i) we extracted from the proteome N random subsets of n elements each (ii) for each of the subsets i, we computed its mean $m_i, i=1,\ldots,N$. Note that for very large N, we expect the m_is to be Gaussianly distributed, with the same average as the population's. We empirically took $N=5000$, which guarantees that the deviations from Gaussianity are small enough to legitimate the use of Z-scores as statistical indicators for significance (iii) for each feature under examination, we compared the mean of the GroEL substrates m_{GroEL} to the mean m and the SD σ of the m_is. It was then possible to give a Z-score equal to $(m-m_{\mathrm{GroEL}})/\sigma$.

Rate of evolution: for each transcribed gene of *E.coli*, we computed the corresponding closest gene in eight other organisms (Table 2), whose sequences we obtained from the KEGG database (Kanehisa and Goto, 2000). Closeness is defined by sequence identity calculated according to the Needleman-Wunsch algorithm as implemented in NEEDLE (Rice *et al.*, 2000). In principle, phylogenetic proximity of the considered organisms suggests that gene pairs selected on the basis of high sequence similarity correspond to proteins with orthologous function; for this study, only gene pairs with at least 40% sequence identity were considered. The assigned gene pairs were verified by confirming that they share the same classification in terms of KEGG orthology classes (Kanehisa and Goto, 2000), where such a classification was available. In this manuscript, we use the ratio dN/dS of the number of non-synonymous and synonymous amino acid substitutions between two given proteins with high sequence identity as a measure for their rate of evolution. To determine it for each pair of genes assessed, we ran the software `yn00` from the package PAML (Yang, 2007) taking the distance described in Yang and Nielsen (2000). This in turn implied first aligning the amino acid sequence of each gene product [for which we used MUSCLE (Edgar, 2004)] and then using this result to align the nucleotide sequences via TRANALIGN (Rice *et al.*, 2000).

Hydrophobicity and aggregation propensity: the average hydrophobicity of proteins was assessed with PEPWINDOW, which gives scores based on the Kyte-Doolitle index (Kyte and Doolittle, 1982); aggregation propensity values were computed with TANGO (Fernandez-Escamilla *et al.*, 2004), an algorithm considering stretches of amino acids rather than entire amino acid sequences, which seems more relevant for aggregation during protein folding. To avoid a bias, membrane proteins were excluded for the analysis of both protein hydrophobicity and aggregation propensity. Likewise, to exclude a possible bias induced by mass spectrometry hydrophobicity computations were repeated considering only experimentally identified *E.coli* proteins (Corbin *et al.*, 2003; Gevaert *et al.*, 2002; Ishihama *et al.*, 2008), with entirely similar results. We also tested the FoldIndex algorithm (Prilusky *et al.*, 2005) and found it to correlate well with both hydrophobicity and aggregation propensity.

Fold type analysis: as of SCOP database version 1.73, 217 of the 252 GroEL substrate proteins (and 2644 of the 3928 annotated *E.coli* proteins in total) had SCOP classes assigned. The substrates fall into the following classes: a:11, b:17, c:126, d:52, e:6, f:4 and g:1. To assess the contribution of structural properties to the rate of evolution of GroEL substrates, we randomly extracted 100 samples from the *E.coli* proteome. To build each sample, we chose randomly 11 proteins of Class a, 17 of Class b, etc. as to reflect the structural properties of the GroEL substrates. We then applied the statistical procedure described above to the sample distribution. Essentiality, expression level, molecular weight, GroES-type binding motif and amino acid chain length were additional parameters tested. Results are shown in the Supplementary Material.

3 RESULTS AND DISCUSSION

In this section, we illustrate our main results (see Table 1): we find that GroEL substrates show lower dN/dS than other proteins; that they are less hydrophobic and have lower aggregation propensity than other proteins; Our results are statistically highly significant (Table 1). Finally structural properties (like the SCOP fold class) are partly responsible for all the observed effects. To the best of our knowledge, the first result is entirely new in the literature, and the second one essentially new, although Noivirt-Brik *et al.* (2005) proved a similar behavior for folding propensity, which is related to hydrophobicity. As for the third result, it was known that GroEL substrates are enriched in some SCOP classes (Kerner *et al.*, 2005), but the effects of the enrichment had not been quantified precisely. Interestingly, our results hold irrespectively of the dataset considered [either Kerner *et al.* (2005) or (Chapman *et al.*, 2006)]; in particular, as for Kerner *et al.* (2005) taking into account only 84 stringent GroEL substrates belonging to Class III, or omitting Class I substrates, does not significantly change our results (Supplementary Fig. S1).

Table 1. GroEL substrates show lower rate of evolution dN/dS, hydrophobicity and aggregation propensity

Dataset	$\frac{dN}{dS}$	Z	Hydroph.	Z	Aggreg.	Z
Proteome	0.073	0	−0.081	0	5.6	0
Substrates from Kerner *et al.* (2005)	0.050	−4.6	−0.25	−3.5	3.1	−5.9
Substrates from Chapman *et al.* (2006)	0.050	−5.1	−0.22	−2.4	3.9	−4.4

Table 2. Rates of evolution w.r.t. *E.coli* for different gamma-proteobacteria. $dN/dS|_s$ is computed on substrates only

Organism	Genes	Matched pairs	Subs.	$\frac{dN}{dS}$	$\frac{dN}{dS}\Big\vert_s$	Z	Sequence identity (%)
Buchnera aphidicola	574	495	66	0.091	0.078	−2.7	94.5
Haemophilus influenzae	1657	1069	113	0.118	0.095	−3.6	94.2
Pasteurella multocida	2015	1255	135	0.129	0.099	−4.9	63.1
Photorhabdus luminescens	4683	1973	176	0.122	0.090	−5.7	95.4
Salmonella typhimurium	4527	3187	219	0.073	0.050	−4.6	99.6
Shigella flexneri	4445	3216	207	0.141	0.085	−3.4	75.2
Vibrio cholerae	3835	1495	171	0.175	0.147	−4.2	92.5
Yersinia pestis	4066	2310	190	0.121	0.093	−4.9	96.4
Escherichia coli	4132	4132	204	0	0	0	100.0

3.1 Parameters describing GroEL substrates

We compared the rate of evolution (dN/dS) between pairs of genes with high sequence identity coding for proteins of *E.coli* and the closely related organism *Salmonella typhimurium*. Several reasons suggested this pair of organisms as prime example: (i) both organisms have comparable genome sizes—*E.coli*: 4132 genes, *S.typhimurium*: 4527 genes, see Table 2—(ii) setting the sequence identity threshold to 90% still allowed us to confidently map 3316 gene pairs (iii) the two GroEL genes are 100% identical in their sequence, suggesting that—although not a prerequisite for this study—proteins of *S.typhimurium* highly similar to *E.coli* GroEL substrates are also likely to interact with their own chaperonin. We found that GroEL substrate proteins diverge less from their best matching partners (i.e. the ones with highest sequence identity) than all mapped gene pairs between *E.coli* and *S.typhimurium* (Fig. 1A, and Supplementary Fig. S1a). The measured dN/dS for the matched

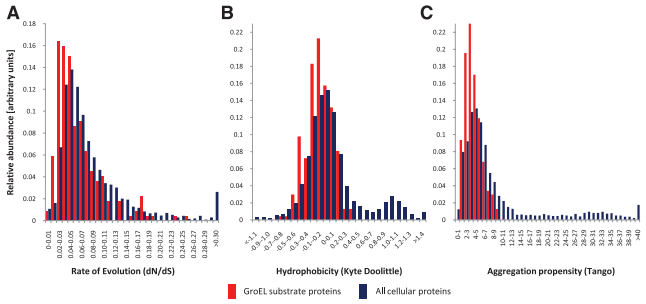

Fig. 1. Rate of evolution, hydrophobicity and aggregation propensity are three parameters differentiating GroEL substrates from other cellular proteins [substrate set from Kerner *et al.* (2005) was used for the plots]. Red histograms: GroEL substrate proteins. Blue histograms: all *E.coli* proteins excluding membrane proteins. (**A**) Rate of evolution dN/dS; (**B**) hydrophobicity; and (**C**) aggregation propensity. The peaks of the distributions for GroEL substrate proteins are in all cases shifted to lower values when compared with other cellular proteins. Additionally, a considerable fraction of cellular *E.coli* proteins is observed with high hydrophobicity and aggregation propensity values. It is important to note that membrane proteins were excluded from the analysis.

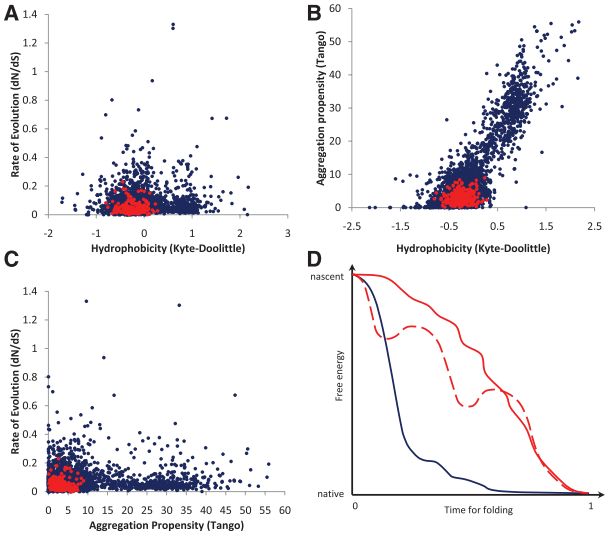

Fig. 2. In the first three panels red dots: GroEL substrates; blue dots: other *E.coli* proteins. (**A**) and (**B**) Rate of evolution does only weakly correlate with hydrophobicity and aggregation propensity. (**C**) Hydrophobicity and aggregation propensity correlate well, but for the subset of GroEL substrates. In all three panels, the supports of the distributions of the hydrophobicity and aggregation propensity values are markedly different between GroEL substrates and proteins folding independently of GroEL. A list of outliers can be found in Supplementary Table S1. (**D**) Proposed model depicting differences in the folding pathways of GroEL substrates (red curves) and other cellular proteins (blue curve). GroEL substrates might have a slower rate of folding (continuous red curve) or form kinetically trapped folding intermediates (dashed red curve); other cellular proteins undergo fast hydrophobic collapse, effectively removing them from the pool of chaperonin interacting proteins.

GroEL substrates was 0.050; the average rate of evolution of all mapped pairs 0.073. We then extended the analysis to other gamma-proteobacteria (*Buchnera aphidicola*, *Haemophilus influenzae*, *Pasteurella multocida*, *Photorhabdus luminescens*, *Shigella flexneri*, *Vibrio cholerae* and *Yersinia pestis*). In all eight pairs of organisms, genes coding for GroEL substrate proteins diverge less than their respective control sets taken as the group of all mapped gene pairs (Table 2).

We assessed the hydrophobicity and the aggregation propensity of proteins using two different, well-established algorithms, Kyte-Doolittle (Kyte and Doolittle, 1982) and TANGO (Fernandez-Escamilla *et al.*, 2004). Both independently showed a clear indication towards GroEL substrates being less hydrophobic and less aggregation-prone than chaperonin-independent folders. More in detail, our study shows that the hydrophobicity averages for substrates and non-substrates are -0.25 and -0.08, respectively; analogous results can be obtained for aggregation propensity (Table 1, Fig. 1C and Supplementary Fig. S1). An interesting fact is that although Kyte-Doolittle and TANGO values for all non-GroEL substrates correlate well ($r = 0.69$, Fig. 2C), aggregation propensity and hydrophobicity only show a very weak correlation when considering GroEL substrate proteins alone ($r = 0.02$, Fig. 2C). We also observed a very weak correlation ($r = 0.06$) between the calculated rates of evolution for matched gene pairs and the respective hydrophobicity values of the corresponding *E.coli* proteins (Fig. 2A). This establishes both low rate of evolution and low hydrophobicity as criteria describing the GroEL substrate proteome independently. Aggregation propensity also describes the GroEL-substrate proteome independently of the rate of evolution (Fig. 2B). We observe that in general our results, and what has already been found in Noivirt-Brik *et al.* (2005) about a lower average folding propensity in GroEL substrates, corroborate each other, since the considered quantities are related.

A detailed structural analysis of GroEL substrate proteins revealed among proteins stringently depending on GroEL a bias towards certain SCOP fold classes, such as the TIM barrel fold (Kerner *et al.*, 2005). We tested the hypothesis that the SCOP class bias can also account for the observed differences: to this end, we created random sets reflecting the SCOP class distribution of the GroEL substrates, and measured their rate of evolution, hydrophobicity and aggregation propensity. While the dN/dS of the random subsets shows a mean of 0.073, the SCOP class correction leads to a mean of 0.062, which is considerably closer to the value for the substrates; similar effects hold for all other quantities (Supplementary Fig. S2). Therefore, we reason that specific structural properties of the GroEL substrate set are at least partly responsible for the observed lower value of our quantities. We additionally tested if a bias in the expression level (Supplementary Fig. S4) or the essentiality of GroEL substrate proteins could account for the observed lower rate of evolution, but found no significant effect. We further analyzed various other parameters; the results are reported in the Supplementary Material.

Finally, as a proof of the consistency and applicability of our findings, we implemented a Bayesian predictor based on the above features, which we also exercise on known heterologous substrates (see Table S2) as well as on a second set of experimentally identified GroEL substrate proteins (Chapman *et al.*, 2006). All results related to our predictor are shown in the Supplementary Material.

3.2 Folding kinetics and evolutionary buffering

We found three novel parameters differentiating GroEL substrates from other cellular proteins, namely rate of evolution, hydrophobicity and aggregation propensity. Why are the identified factors suitable criteria to describe the GroEL substrate proteome? A common answer could lie in the folding pathway of proteins considering robustness to mutations and differences in kinetic stability of GroEL substrates and proteins folding without the help of chaperones. The protein folding process has been depicted as a descent through a funnel-shaped energy landscape, where the native structure with the least free energy sits at the bottom (Onuchic and Wolynes, 2004). Proteins with a robust folding pathway reach the native state fast and seamlessly without the aid of chaperones. Additionally, these proteins can accommodate for certain mutations without severe effects on their thermodynamic stability (Bloom *et al*., 2006; Tokuriki *et al*., 2007). On the other hand, chaperone substrates routinely form kinetically trapped folding intermediates, leading to the accumulation of non-native isoforms; interaction with chaperones smoothes their energy landscape and rescues folding (Altschuler and Willison, 2008). Mutations in chaperonin substrates, although potentially beneficial for function, are hence more likely to lead to a further reduction in folding fidelity and protein stability, explaining the observed lower average rate of evolution for the GroEL substrate proteome. The rate of folding is another crucial parameter to successfully reach the native state in the cytosol. Upon emergence from the ribosome, nascent amino acid chains of globular proteins need to bury hydrophobic residues in the core of the protein, thereby stabilizing their structure and preventing unproductive inter-protein interactions. Proteins with a relatively lower hydrophobicity and low aggregation propensity (as observed for the bulk of the GroEL substrates) might have slower and only partial initial collapsing steps, leading to the accumulation of folding intermediates with exposed hydrophobic residues functioning as recognition sequences for GroEL (Fig. 2D) and which likely cause increased aggregation when translated in a chaperone-free *in vitro* system (Niwa *et al.*, 2009).

Several groups have shown that chaperones function as evolutionary buffers or capacitors, shielding chaperone substrate proteins from deleterious effects of mutations both at the phenotypic (Fares *et al.*, 2002; Sangster *et al.*, 2004) and the genetic level (Tokuriki and Tawfik, 2009). Our results indicate that GroEL substrates have on average a lower rate of evolution than other cellular proteins; this fact seems to question the applicability of the so-called 'buffering hypothesis' under *in vivo* conditions. We speculate that even though GroEL promotes genetic and phenotypic variation, such an effect might be transient and represent only a comparatively short phase on evolutionary time scales, getting rapidly exhausted after the introduction of the chaperone into the organism. Today, at equilibrium, constraints in the folding pathways of genuine chaperone substrate proteins under physiological GroEL expression levels might keep them from reaching their potentially higher rate of evolution *in vivo*: because GroEL substrates fold on the verge of what is energetically possible, they might just not be able to accommodate more mutations, which are likely to be detrimental and dramatically impair the folding process. It is also important to appreciate that other mechanisms already proposed in the literature might contribute as well to keep evolutionary buffering effects unobservable *in vivo*: among them the role of chaperones play as examiners (Tomala and Korona, 2008) and the recently proposed intriguing finding (Specchia *et al.*, 2010) that chaperones might interfere with genome-wide mutagenesis processes other than simple random mutations.

Funding: Spanish Ministerio de Educación y Ciencia (Consolider grant CSD2007-00050 to P.R.); European Molecular Biology Organisation long-term Fellowship (to T.M.).

Conflict of Interest: none declared.

REFERENCES

Altschuler,G. and Willison,K. (2008) Development of free-energy-based models for chaperonin containing TCP-1 mediated folding of actin. *J. R. Soc. Interface*, **5**, 1391–1408.

Bloom,J. *et al*. (2006) Structural determinants of the rate of protein evolution in yeast. *Mol. Biol. Evol.*, **23**, 1751–1761.

Chapman,E. *et al*. (2006) Global aggregation of newly translated proteins in an Escherichia coli strain deficient of the chaperonin GroEL. *Proc. Natl Acad. Sci. USA*, **103**, 15800–15805.

Corbin,R. *et al*. (2003) Toward a protein profile of Escherichia coli: comparison to its transcription profile. *Proc. Natl Acad. Sci. USA*, **100**, 9232–9237.

Edgar,R. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.

Fares,M. *et al*. (2002) Endosymbiotic bacteria: GroEL buffers against deleterious mutations. *Nature*, **417**, 398.

Fenton,W. and Horwich,A. (1997) GroEL-mediated protein folding. *Protein Sci*., **6**, 743–760.

Fernandez-Escamilla,A. *et al*. (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.*, **22**, 1302–1306.

Gevaert,K. *et al*. (2002) Chromatographic isolation of methionine-containing peptides for gel-free proteome analysis: identification of more than 800 Escherichia coli proteins. *Mol. Cell. Proteomics*, **1**, 896–903.

Hartl,F. and Hayer-Hartl,M. (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science*, **295**, 1852–1858.

Ishihama,Y. *et al*. (2008) Protein abundance profiling of the Escherichia coli cytos. *BMC Genomics*, **9**, 102.

Kanehisa,M. and Goto,S. (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.

Kerner,M. *et al*. (2005) Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. *Cell*, **122**, 209–220.

Kyte,J. and Doolittle,R. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, **157**, 105–132.

Niwa,T. *et al*. (2009) Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins. *Proc. Natl Acad. Sci. USA*, **106**, 4201–4206.

Noivirt-Brik,O. *et al*. (2005) Low folding propensity and high translation efficiency distinguish in vivo substrates of GroEL from other Escherichia coli proteins. *Bioinformatics*, **23**, 3276–3279.

Onuchic,J. and Wolynes,P. (2004) Theory of protein folding. *Curr. Opin. Struct. Biol.*, **14**, 70–75.

Prilusky,J. *et al*. (2005) Foldindex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*, **21**, 3435–3438.

Rice,P. *et al*. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.

Sakikawa,C. *et al*. (1999) On the maximum size of proteins to stay and fold in the cavity of GroEL underneathGroES. *J. Biol. Chem*., **274**, 21251–21256.

Sangster,T. *et al*. (2004) Causes, effects and implications of Hsp90-mediated genetic capacitance. *Bioessays*, **26**, 348–362.

Specchia,V. *et al*. (2010) Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. *Nature*, **463**, 662–665.

Stan,G. *et al*. (2006) Residues in substrate proteins that interact with GroEL in the capture process are buried in the native state. *Proc. Natl Acad. Sci. USA*, **103**, 4433–4438.

Tokuriki,N. and Tawfik,D. (2009) Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature*, **459**, 668–673.

Tokuriki,N. *et al*. (2007) The stability effects of protein mutations appear to be universally distributed. *J. Mol. Biol.*, **269**, 1318–1332.

Tomala,K. and Korona,R. (2008) Molecular chaperones and selection against mutations. *Biol. Direct*, **26**, 3–5.

Weissman,J. *et al*. (1994) GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms. *Cell*, **78**, 693–702.

Xu,Z. *et al*. (1997) The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature*, **388**, 741–750.

Yang,Z. and Nielsen,R. (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.*, **17**, 32–43.

Yang,Z. (2007) PAML4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.*, **24**, 1586–1591.