Chaperonin-Dependent Accelerated Substitution Rates in Prokaryotes

David Bogumil, and Tal Dagan*

Institute of Botany III, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany

*Corresponding author: E-mail: tal.dagan@uni-duesseldorf.de.

Accepted: 16 July 2010

Abstract

Many proteins require the assistance of molecular chaperones in order to fold efficiently. Chaperones are known to mask the effects of mutations that induce misfolding because they can compensate for the deficiency in spontaneous folding. One of the best studied chaperones is the eubacterial GroEL/GroES system. In *Escherichia coli*, three classes of proteins have been distinguished based on their degree of dependency on GroEL for folding: 1) those that do not require GroEL, 2) those that require GroEL in a temperature-dependent manner, and 3) those that obligately require GroEL for proper folding. The buffering effects of GroEL have so far been observed in experimental regimens, but their effect on genomes during evolution has not been examined. Using 446 sequenced proteobacterial genomes, we have compared the frequency of amino acid replacements among orthologs of 236 proteins corresponding to the three categories of GroEL dependency determined for *E. coli*. Evolutionary rates are significantly correlated with GroEL dependency upon folding with GroEL dependency class accounting for up to 84% of the variation in amino acid substitution rates. Greater GroEL dependency entails increased evolutionary rates with GroEL obligatory proteins (Class III) evolving on average up to 15% faster than GroEL partially dependent proteins (Class II) and 35% faster than GroEL-independent proteins (Class I). Moreover, GroEL dependency class correlations are strictly conserved throughout all proteobacteria surveyed, as is a significant correlation between folding class and codon bias. The results suggest that during evolution, GroEL-dependent folding increases evolutionary rate by buffering the deleterious effects of misfolding-related mutations.

Key words: genome evolution, misfolding, GroEL, codon usage.

Introduction

Chaperones (Ellis 1987), also called heat-shock proteins (HSPs), are essential in both prokaryotes and eukaryotes as they assist protein folding, prevent protein aggregation, and play a crucial role in survival under stress conditions (Young et al. 2004). Moreover, chaperones have been shown to buffer mutational effects both in eukaryotes and in prokaryotes (Rutherford 2003). In *Arabidopsis thaliana*, the reduction of Hsp90 expression level exposes genotype-independent phenotypic variation (Queitsch et al. 2002). In prokaryotes, Hsp60 (GroEL) is essential to organismal fitness under high mutational loads in *Escherichia coli* (Fares et al. 2002; Maisnier-Patin et al. 2005) and in *Buchnera aphidicola* (Moran 1996). Hence in individual organisms, chaperones exert a buffering effect on slightly deleterious mutations, presumably by compensating for decreased folding stability of mutated proteins (Moran 1996; Todd et al. 1996; Fares et al. 2002; Queitsch et al. 2002; Maisnier-Patin et al. 2005; Tokuriki and Tawfik 2009). Is this property widespread in nature and does it affect prokaryote genome evolution?

The chaperone pathway in eubacteria includes a ribosome-bound trigger factor that meets polypeptides as they emerge from the ribosome. The DnaK (Hsp70) and its co-chaperone DnaJ may bind alternatively to nascent polypeptides. Subsequently, the GroEL/GroES (Hsp60) chaperonine system operates on a subset of the proteins whose folding requires further energy investment (Young et al. 2004). In *E. coli*, GroEL/GroES is found to interact with about 10% of all soluble proteins (Kerner et al. 2005) and is the only chaperone essential to the bacterium under all tested conditions (Horwich et al. 1993). The GroEL/GroES chaperones are found in all eubacteria except a few highly reduced endosymbionts (Lund et al. 2003). Proteins found in interaction with GroEL in *E. coli* can be classified into three dependency classes (Kerner et al. 2005): GroEL-independent proteins (Class I) fold spontaneously in standard conditions (37 °C)

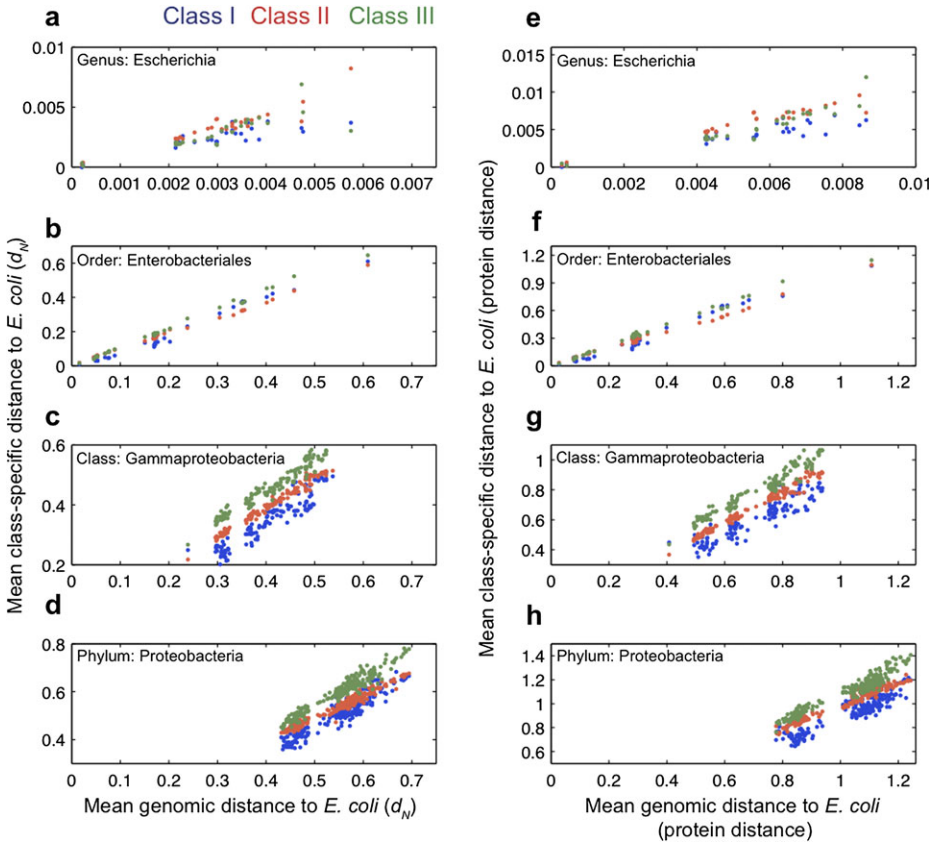

Fig. 1.—Evolutionary rates of proteins in the three GroEL dependency classes within 445 Proteobacteria compared with their *Escherichia coli* strain O157H7 EDL933 ortholog. Each dot in the figure represents the mean distance of all proteins in the same class within the same species from their ortholog in *E. coli* O157H7 EDL933.

and attain on average 55% of their activity independent of chaperones, GroEL, or otherwise. GroEL partially dependent proteins (Class II) require GroEL/GroES assistance, in addition to other chaperons, at 37 °C but do not require GroES at 25 °C, where spontaneous folding is observed. GroEL obligatory proteins (Class III) fail to fold spontaneously at 37 °C and have an obligate requirement for GroEL/GroES in order to attain activity (Kerner et al. 2005). GroEL is known to be a capacitor for slightly deleterious mutations in vitro (Fares et al. 2002; Queitsch et al. 2002; Maisnier-Patin et al. 2005; Tokuriki and Tawfik 2009). If this is also true in nature, Class III proteins should exhibit increased numbers of nonsynonymous substitutions in comparison to Classes I and II.

Materials and Methods

GroEL dependency classes were obtained from Kerner et al. (2005). The Kerner et al. (2005) list contains 249 SWISSPROT accession numbers from various *E. coli* strains. Four proteins that are classified into more than one class were removed. Completely sequenced genomes of 446 Proteobacteria were downloaded from NCBI (http://www.ncbi.nlm.nih.gov/; July 2009 version). Non-proteobacterial taxa were not included in the analysis because we cannot assume that protein interaction with GroEL is conserved in all prokaryotes. In order to use a single reference genome in our analysis, the Kerner et al. (2005) proteins were Blasted (Altschul et al. 1990) on *E. coli* O157H7 EDL933. Proteins that had hits below 98% identical amino acids were curated manually and nine proteins were removed. The remaining proteins distribute as follows: 37 Class I, 120 Class II, and 79 Class III proteins.

Orthologs to *E. coli* strain O157H7 EDL933 proteins in all completely sequenced Proteobacteria were inferred using a reciprocal best Blast hit procedure (Tatusov et al. 1997) with an *e* value $<1 \times 10^{-10}$ cutoff. All orthologous protein pairs were aligned using ClustalW (Thompson et al. 1994). Pairwise alignment reliability was tested using HoT (Landan and Graur 2007), and alignments having column score <90% were excluded. Protein alignments were translated into nucleotide alignments using PAL2NAL (Suyama et al. 2006). Rates of nonsynonymous nucleotide substitutions were calculated by an approximation to maximum likelihood method using yn00 (Yang 2007). Protein distances were calculated by PROTDIST (Felsenstein 2005) using Jones, Taylor, and Thorton (JTT) substitution matrix (Jones et al. 1992). Preferred codons

Table 1
Statistical Tests for Homogeneity of Medians among the GroEL Dependency Classes

Variable	Taxonomic Group	Homogeneity of Medians (*P* value)[a]	Post hoc Comparisons[b]
d_N	Genus: Escherichia	$7.5 \times 10^{-15*}$	I < II, III and II = III[c]
	Order: Enterobacteriales		
	Class: Gammaproteobacteria	$<2.2 \times 10^{-16*}$	I < II < III
	Phylum: Proteobacteria		
Protein distance	Genus: Escherichia	$1.1 \times 10^{-16*}$	I < II, III and II = III
	Order: Enterobacteriales		
	Class: Gammaproteobacteria		
	Phylum: Proteobacteria	$<2.2 \times 10^{-16*}$	I < II < III
CAI	Genus: Escherichia	$<2.2 \times 10^{-16*}$	I > II, III and II = III
	Order: Enterobacteriales		
	Class: Gammaproteobacteria		I > II > III
	Phylum: Proteobacteria		I > II, III and II = III

[a] Using Friedman test.
[b] $\alpha = 0.05$, using Tukey's test.
[c] Roman numbers denote the classes. The notation I < II means that the values of the tested variable are significantly smaller in Class I proteins than in Class II proteins.
*P value << 0.01.

for each genome and codon adaptation index (CAI) (Sharp and Li 1987) for all genes were calculated using the EMBOSS package (Rice et al. 2000). Amino acid usage and GC content were calculated using an in-house PERL script. Statistical analysis was performed using MatLab statistical toolbox.

To test our hypothesis in different phylogenetic ,we grouped the species in the genome sample into four groups according to their relatedness with *E. coli* strain O157H7 EDL933: 1) Genus: Escherichia, 2) Order: Enterobacterialles, 3) Class: Gammaproteobacteria, and 4) Phylum: Proteobacteria. In order to keep the groups independent, each genome is included in a single group. The genomes are sorted into the groups by their phylogenetic relations with *E. coli*.

Results

To compare nonsynonymous substitution rates among orthologs of the *E. coli* GroEL Class I (37 members), Class II (120 members), and Class III (79 members) proteins, we identified and aligned (Thompson et al. 1994) their orthologs from 446 sequenced proteobacterial genomes. Numbers of nonsynonymous nucleotide substitutions (d_N) (Nei and Gojobori 1986) and amino acid replacements were calculated in pairwise genome comparisons (Yang 2007). For a given genome comparison, the three class-specific mean d_N values were plotted against the mean of all comparisons for the genome pair; this compensates for genome- and lineage-specific differences in substitution rate and nucleotide bias.

Plotting these values at different phylogenetic depths revealed strong and distinct differences in evolutionary rate for the three protein classes, differences which become increasingly apparent with increasing sequence divergence (fig. 1). For intraspecific comparisons within *E. coli* (fig. 1*a*), the differences among the three GroEL dependency classes are not readily visible because of stochastic variation for small d_N values, but they are significant ($P = 7.55 \times 10^{-15}$, using the Friedman test; Zar 1999), with Class I proteins having

Table 2
Explained Variability and Mean Ratios of Class-Specific Values for All Tested Samples

	Genus: Escherichia	Order: Enterobacteriales	Class: Gammaproteobacteria	Phylum: Proteobacteria
d_N				
Explained variability[a]	0.36	0.4	0.87	0.8
Class III/II	0.92	1.06	1.14	1.1
Class III/I	1.1[b]	1.4	1.31	1.18
Protein distance				
Explained variability	0.6	0.3	0.84	0.76
Class III/II	0.87	1.06	1.15	1.1
Class III/I	1.17[b]	1.36	1.35	1.2
CAI				
Explained variability	0.96	0.57	0.48	0.53
Class III/II	0.99	1	0.99	1
Class III/I	0.95	0.98	0.97	0.97

[a] Explained variability was calculated by partial $\eta^2 = \eta^2 = \frac{SS_{treatment}}{SS_{treatment}+SS_{error}}$ with Friedman test.
[b] *Escherichia coli* K12 MG1655 and *E. coli* O157H7 comparisons resulted in zero distance for Class I proteins and were omitted from the calculation.

significantly lower rates than Class II and Class III proteins ($\alpha = 0.05$, using Tukey's post hoc test; Zar 1999). The same test on a larger and ~100-fold more divergent orthologs set from 60 enterics (but excluding *E. coli*) shows a more significant difference in d_N among the GroEL dependency classes ($P < 2.2 \times 10^{-16}$, using Friedman test; fig. 1*b*), with Class I proteins having significantly lower d_N than Class II proteins, and the latter having significantly lower d_N than Class III proteins ($\alpha = 0.05$, using Tukey's post hoc test).

Comparisons within the Gammaproteobacteria (135 genomes; excluding enterics) yielded even more significant correlations (table 1) and furthermore a striking distinction of the three classes (fig. 1*c*). Differences between the GroEL dependency classes account for 87% of the variation between class-specific mean d_N values (table 2). Extending the sample to include 227 Proteobacteria (excluding Gammaproteobacteria) entailed comparisons of greater divergence, with most d_N values exceeding 0.5 substitutions per site (fig. 1*d*), but the significance and the trends remained (table 1), with GroEL dependency class accounting for 80% of the observed differences in class-specific mean d_N (table 2). These correlations held up for GroEL dependency class in amino acid sequence comparisons for the same phylogenetic samples (fig. 1*e*–*h*). At the level of amino acid replacements estimated by JTT (Jones et al. 1992) protein distances for Gammaproteobacteria, Class III proteins evolve on average 15% faster than Class II and 35% faster than Class I proteins (table 2). GroEL folding dependency thus appears to be a major and hitherto undetected determinant of sequence divergence in prokaryotes.

But is the correlation causal? Protein conservation and expression level are known to be correlated (Krylov et al. 2003 ; Drummond et al. 2006; Pál et al. 2006). If chaperon dependency is related to expression level, then it is possible that expression level is the determinant of evolutionary rate differences among the GroEL dependency classes (Warnecke and Hurst 2010). A comparison of protein expression levels measured for *E. coli* strain K12 MG1655 (Lu et al. 2007) shows that these are not equal among the three classes ($P = 2.1 \times 10^{-5}$, using Kruskal–Wallis) with Class I proteins having significantly higher expression levels than Classes II and III proteins, whereas Classes II and III do not differ significantly from each other in their expression levels ($\alpha = 0.05$, using Tukey's post hoc test; fig. 2). To test if protein expression level has any effect on our results, we compared the evolutionary rates among the three GroEL dependency classes while adjusting for the variability in protein expression levels using analysis of covariance (ANCOVA). For the comparison within the genus level and order level, we found significant differences between the three GroEL dependency classes also when protein expression level is considered as the covariate variable (table 3). The ANCOVA was not applicable for the class and phylum levels because the underlying assumptions for that test were not met.

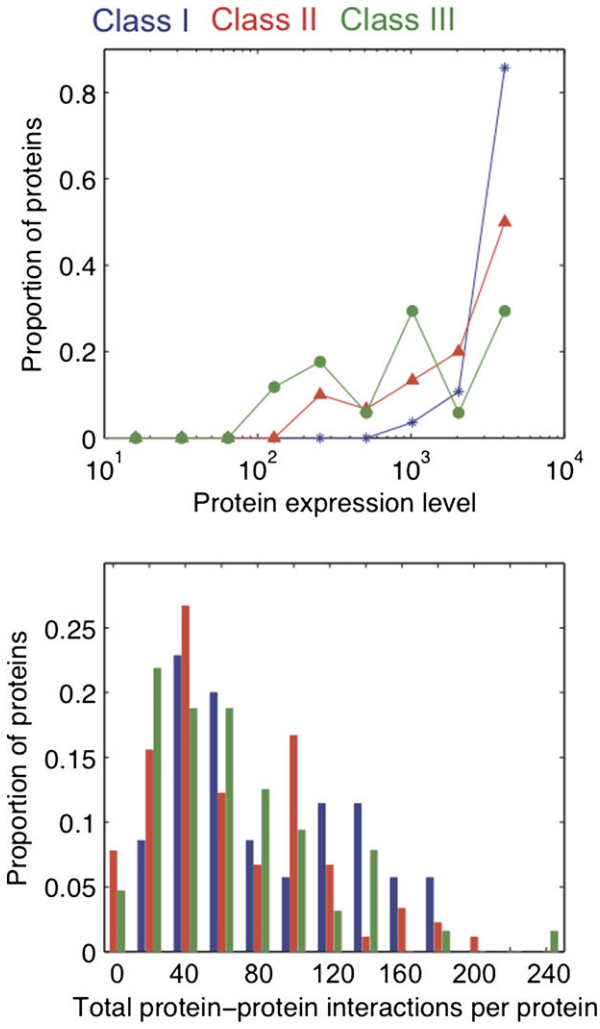

Fig. 2.—Distribution of protein expression levels (Lu et al. 2007) (top) and number of protein-protein interactions (Hu et al. 2009) (bottom) in the three GroEL dependency classes.

Protein expression level has been shown to be positively correlated with the connectivity of a protein within the cellular protein–protein interaction (PPI) network in yeast (von Mering et al. 2002). However, the correlation strength is highly dependent upon the method used to detect interacting proteins (von Mering et al. 2002). Here we tested for difference in PPI frequency among the three dependency classes by using PPI from Hu et al. (2009). We find that the three dependency classes are statistically different in their PPI frequency ($P = 0.049$, using Kruskal–Wallis test) with Class I proteins having a slightly higher frequency of PPIs (median PPI per protein—Class I: 64, Class II: 50; Class III: 52; fig. 2).

We also compared the CAI (Sharp and Li 1987), which is positively correlated, and strongly so, with expression level (Sharp and Li 1987), among orthologs in the three dependency classes at different phylogenetic depths. Class I

Table 3
Statistical Tests for Differences in Evolutionary Rates among the Three GroEL Dependency Classes with a Covariate

Response Variable (y)	Covariate (x)	Taxonomic Group	Pooled Regression[a]	Homogeneity of Slopes among Groups[b]	Homogeneity of Intercepts among Groups[c]
d_N	Protein expression level	Genus: Escherichia	0.026^{*}	0.074	0.0049^{*}
		Order: Enterobacteriales	$6.5 \times 10^{-6**}$	0.52	$<2.2 \times 10^{-16**}$
		Class: Gammaproteobacteria	$<2.2 \times 10^{-16**}$	$<2.2 \times 10^{-16**}$	n.a.
		Phylum: Proteobacteria	$<2.2 \times 10^{-16**}$	$<2.2 \times 10^{-16**}$	n.a.
Protein distance	Protein expression level	Genus: Escherichia	0.0044^{*}	0.15	6.5×10^{-4}
		Order: Enterobacteriales	$1.6 \times 10^{-4**}$	0.49	$<2.2 \times 10^{-16}$
		Class: Gammaproteobacteria	$<2.2 \times 10^{-16**}$	$1.1 \times 10^{-16**}$	n.a.
		Phylum: Proteobacteria	$<2.2 \times 10^{-16**}$	$<2.2 \times 10^{-16**}$	n.a.
d_N	CAI	Genus: Escherichia	$1.3 \times 10^{-9**}$	$5.5 \times 10^{-4**}$	n.a.
		Order: Enterobacteriales	$<2.2 \times 10^{-16**}$	$<2.2 \times 10^{-16**}$	n.a.
		Class: Gammaproteobacteria	$<2.2 \times 10^{-16**}$	$6.1 \times 10^{-6**}$	n.a.
		Phylum: Proteobacteria	$<2.2 \times 10^{-16**}$	0.74	$<2.2 \times 10^{-16**}$
Protein distance	CAI	Genus: Escherichia	$7.7 \times 10^{-13**}$	$<2.2 \times 10^{-16**}$	n.a.
		Order: Enterobacteriales	$<2.2 \times 10^{-16**}$	$5.1 \times 10^{-9**}$	n.a.
		Class: Gammaproteobacteria	$<2.2 \times 10^{-16**}$	$1.9 \times 10^{-13**}$	n.a.
		Phylum: Proteobacteria	$<2.2 \times 10^{-16**}$	0.42	$<2.2 \times 10^{-16**}$

NOTE.—Results of the ANCOVA test and its underlying assumptions (Sokal and Rohlf 1995) are presented. To adjust for overall differences among species, the response variable was divided by the genomic average.

[a] Using F-test for linear relation between the response and covariate $y = ax + b$ testing the null hypothesis H_0: $a = 0$.

[b] Using F-test for equality of slopes among the groups. Each group is fitted with a linear regression $y_{class} = a_{class}x_{class} + b_{class}$ followed by testing the null hypothesis H_0: $a_{class\ I} = a_{class\ II} = a_{class\ III}$.

[c] Using F-test for equality of intercepts among the groups. This is equivalent to a test for equality of means with the null hypothesis H_0: $\mu_{class\ I} = \mu_{class\ II} = \mu_{class\ III}$.

*P value < 0.05.

**P value $<< 0.01$.

proteins have significantly higher CAI than Classes II and III proteins, whereas CAI values of Class II proteins are either similar (in the order and phylum sets) or slightly increased in comparison to Class III proteins (table 1 and fig. 3). This trend is true not only for *E. coli* (Warnecke and Hurst 2010) but throughout the proteobacteria. Thus, although high expression levels can explain the decreased evolutionary rates for Class I proteins, it cannot explain the increased evolutionary rates in Class III proteins in comparison to Class II proteins. Hence, the difference in evolutionary rates among the three GroEL dependency groups does indeed appear to be attributable to GroEL buffering effects.

Proteins in the three dependency classes are highly dissimilar in their amino acid composition. A comparison of *E. coli* O157H7 EDL933 proteins shows that Class II and Class III proteins comprise significantly more positively charged amino acids (Fujiwara et al. 2010) and less negatively charged amino acids than Class I proteins. No significant difference is found in hydrophobic amino acids or polar uncharged amino acids composition (supplementary table S1, Supplementary Material online). Cysteine and proline usage is significantly higher in Class II and Class III proteins in comparison to Class I proteins. No significant difference in glycine usage among the classes was found (supplementary table S1 and supplementary fig. S1, Supplementary Material online). Genes encoding for Class III proteins are significantly GC richer than Class I proteins (supplementary table S1, Supplementary Material online). This result is attributable to the amino acid usage of Class III proteins, most of them are encoded by GC-rich codons. Repeating this analysis for all orthologs in all phylogenetic depths reveals that the same trends in amino acid usage are general for all tested proteobacteria (supplementary table S2 and supplementary figs. S2–S5, Supplementary Material online). No correlation was found between any of the amino acid usage measures and evolutionary rates (supplementary table S1, Supplementary Material online); hence, the difference in amino acid usage among the GroEL dependency classes may be attributed to the interaction with GroEL (Fujiwara et al. 2010).

Discussion

GroEL can buffer slightly deleterious mutations in experimental setups. In nature this same capacity leads to increased evolutionary rates for GroEL-dependent proteins. It has recently been suggested that protein misfolding has a key role in determining evolutionary rates (Drummond et al. 2005; Drummond and Wilke 2008; Lobkovsky et al. 2010; Warnecke and Hurst 2010). Our results indicate that GroEL-dependent folding is a biological mechanism that can manifest such effects. However, the correlation of GroEL dependency classes with evolutionary rates, protein expression levels, and CAI implies that the promiscuous amino acid substitution regime allowed by the GroEL buffering might not be uniformly distributed within the cellular protein network. The Class I proteins comprise a group of highly

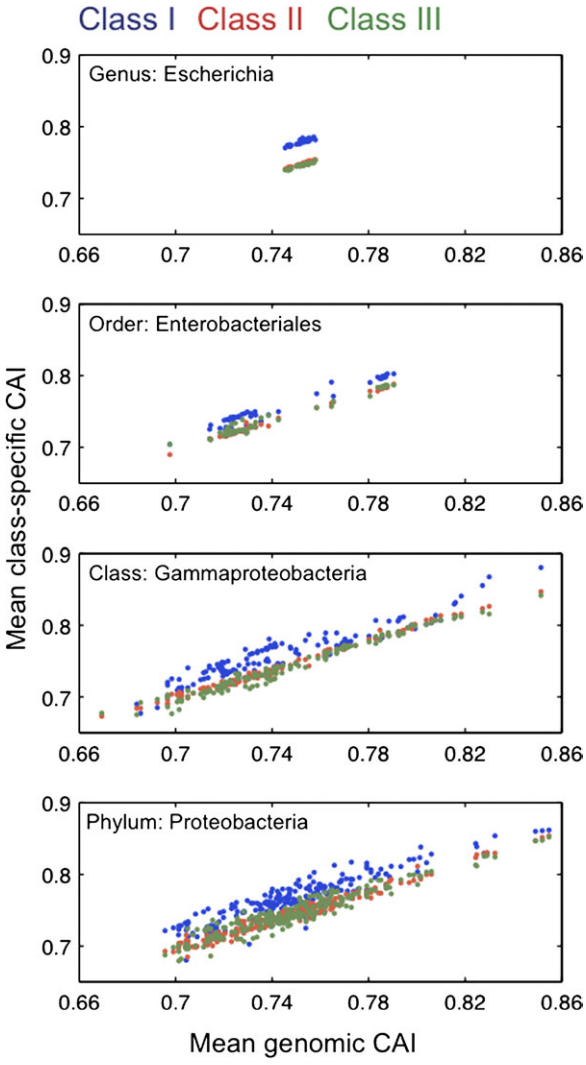

Fig. 3.—CAI of proteins in the three GroEL dependency classes.

conserved, highly expressed proteins having higher CAIs. In contrast, the Class III proteins evolve with an increased evolutionary rate (fig. 1), are expressed at lower levels (fig. 2), and are encoded by less preferred codons (Warnecke and Hurst 2010) (fig. 3). Protein expression level is positively correlated with the number of protein interactions and negatively correlated with dispensability (Pál et al. 2006), whereas CAI is correlated with translation accuracy and efficiency (Drummond and Wilke 2008; Tuller et al. 2010). Hence, proteins that are essential to the cell and that are highly connected in the *E. coli* protein network are not only more conserved but also translated with higher accuracy and tend to fold spontaneously. Conversely, proteins that have a more peripheral role within the cell are more tolerant to increased evolutionary rates and are protected from slightly deleterious mutations by the buffering effect of the GroEL/GroES chaperone.

Supplementary Material

Supplementary figures S1–S6 and tables S1 and S2 are available at *Genome Biology and Evolution* online (http://www.oxfordjournals.org/our_journals/gbe/).

Acknowledgments

We thank the German Federal Ministry of Education and Research (BMBF) for financial support. We are thankful to Giddy Landan for his help in statistical analysis and to Bill Martin, Rotem Sorek, and Martin Lercher for their help in refining the manuscript.

Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005. Why highly expressed proteins evolve slowly. Proc Natl Acad Sci U S A. 102:14338–14343.

Drummond DA, Raval A, Wilke CO. 2006. A single determinant dominates the rate of yeast protein evolution. Mol Biol Evol. 23:327–337.

Drummond DA, Wilke CO. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell. 134:341–352.

Ellis RJ. 1987. Proteins as molecular chaperones. Nature. 328:378–379.

Fares MA, Ruiz-González MX, Moya A, Elena SF, Barrio E. 2002. Endosymbiotic bacteria: groEL buffers against deleterious mutations. Nature. 417:398.

Felsenstein J. 2005. PHYLIP (phylogeny inference package). Version 3.6. Seattle (WA): Department of Genome Sciences, University of Washington.

Fujiwara K, Ishihama Y, Nakahigashi K, Soga T, Taguchi H. 2010. A systematic survey of in vivo obligate chaperonin-dependent substrates. EMBO J. 29:1552–1564.

Horwich AL, Low KB, Fenton WA, Hirshfield IN, Furtak K. 1993. Folding in vivo of bacterial cytoplasmic proteins: role of GroEL. Cell. 74:909–917.

Hu P, et al. 2009. Global functional atlas of Escherichia coli encompassing previously uncharacterized proteins. PLoS Biol. 7:e96.

Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation rate matrices from protein sequences. Comput Appl Biosci. 8:275–282.

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

Krylov DM, Wolf YI, Rogozin IB, Koonin EV. 2003. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. Genome Res. 13:2229–2235.

Landan G, Graur D. 2007. Heads or tails: a simple reliability check for multiple sequence alignments. Mol Biol Evol. 24:1380–1383.

Lobkovsky AE, Wolf YI, Koonin EV. 2010. Universal distribution of protein evolution rates as a consequence of protein folding physics. Proc Natl Acad Sci U S A. 107:2983–2988.

Lu P, Vogel C, Wang R, Yao X, Marcotte EM. 2007. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol. 25:117–124.

Lund PA, Large AT, Kapatai G. 2003. The chaperonins: perspectives from the archaea. Biochem Soc Trans. 31:681–685.

Maisnier-Patin S, et al. 2005. Genomic buffering mitigates the effects of deleterious mutations in bacteria. Nat Genet. 37:1376–1379.

Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc Natl Acad Sci U S A. 93:2873–2878.
Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 3:418–426.
Pál C, Papp B, Lercher MJ. 2006. An integrated view of protein evolution. Nat Rev Genet. 7:337–348.
Queitsch C, Sangster TA, Lindquist S. 2002. Hsp90 as a capacitor of phenotypic variation. Nature. 417:618–624.
Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16:276–277.
Rutherford SL. 2003. Between genotype and phenotype: protein chaperones and evolvability. Nat Rev Genet. 4:263–274.
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.
Sokal RR, Rohlf FJ. 1995. Biometry. 3rd ed. San Francisco (CA): Freeman.
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.
Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. Science. 278:631–637.
Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.
Todd MJ, Lorimer GH, Thirumalai D. 1996. Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism. Proc Natl Acad Sci U S A. 93:4030–4035.
Tokuriki N, Tawfik DS. 2009. Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature. 459:668–673.
Tuller T, Waldman YY, Kupiec M, Ruppin E. 2010. Translation efficiency is determined by both codon bias and folding energy. Proc Natl Acad Sci U S A. 107:3645–3650.
von Mering C, et al. 2002. Comparative assessment of large-scale data sets of protein–protein interactions. Nature. 417:399–403.
Warnecke T, Hurst LD. 2010. GroEL dependency affects codon usage—support for a critical role of misfolding in gene evolution. Mol Syst Biol. 6:340.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.
Young JC, Vishwas RA, Siegers K, Hartl FU. 2004. Pathways of chaperone-mediated protein folding in the cytosol. Nat Rev Mol Cell Biol. 5:781–791.
Zar JH. 1999. Biostatistical analysis. Upper Saddle River (NJ): Prentice Hall.

Associate editor: Takashi Gojobori