Fitness Costs of Minimal Sequence Alterations Causing Protein Instability and Toxicity

Katarzyna Tomala,[1] Elzbieta Pogoda,[1] Agata Jakubowska,[1] and Ryszard Korona*[,1]
[1]Institute of Environmental Sciences, Jagiellonian University, Krakow, Poland
***Corresponding author:** E-mail: ryszard.korona@uj.edu.pl.
Associate editor: Willie Swanson

Abstract

Destabilization of a protein impairs its metabolic efficiency. It is less clear how often destabilization also results in a gain of toxicity. We derived collections of temperature-sensitive, and thus structurally unstable, mutants of the yeast *ADE2* and *LYS2* genes by introducing single or very few amino acids substitutions. Overexpression of these mutant proteins led to a common, although unequal, fitness decrease. Interestingly, although the mutant proteins were functionally redundant, higher expression levels were associated with higher fitness. This result suggests that growth was hampered not by the accumulation of damaged chains but by the activities needed to remove them or by the damage caused before they were removed. Our results support the idea that any protein can become toxic when destabilized by a point mutation.

Key words: **thermal sensitivity, fitness cost, *Saccharomyces cerevisiae*.**

Proteins are often only marginally stable and are therefore readily destabilized in the crowded interior of the cell (Ellis 2001; Hartl et al. 2011). It has been proposed that protein stability is under the constant and fine control of natural selection (DePristo et al. 2005; Tartaglia et al. 2007; Lehner 2011). Mutations that make proteins less stable or that facilitate their destabilization in the case of mistranslation would be purged from genes, especially those coding for abundantly expressed proteins (Drummond et al. 2005; Drummond and Wilke 2008; Yang et al. 2010; Serohijos et al. 2012). Indeed, recent research has shown that genes coding for unstable proteins tend to undergo slow transcription and translation, as this presumably decreases the danger of protein misfolding (Tartaglia et al. 2009; Gsponer and Babu 2012; Zur and Tuller 2012). A number of diseases are known to result from the so-called gain of toxic function by structurally unstable proteins (Winklhofer et al. 2008). These proteins have attracted much attention, but they form a rather restricted class (amyloids, prions) (Hamley 2007). To evaluate general evolutionary hypotheses, the frequency of destabilizing mutations and a number of endangered proteins need to be known. Several studies found it difficult to demonstrate that the destabilization of the protein structure can lead to loss of fitness (Plata et al. 2010; Bershtein et al. 2012; Eames and Kortemme 2012). However, overexpression of a few temperature sensitive GFP variants and one Ura3p variant did show that misfolded and aggregated proteins slow the growth rate in yeast (Geiler-Samerotte et al. 2011). Both the exogenous GFP and especially the endogenous Ura3p variants contained multiple substitutions. Natural selection, however, is more likely to discriminate between alleles differing in only one or very few positions. We sought to study minimal sequence alterations leading to conditional malfunctioning of proteins.

We targeted the yeast *ADE2* and *LYS2* genes. Their products are cytosolic enzymes of moderate abundance (~0.01 and 0.03% [w/w] of total protein present in the yeast cell). Applying standard (not deliberately mutagenic) polymerase chain reaction, we created libraries of plasmids with *ade2* or *lys2* mutated alleles under natural promoters. Strains that lacked chromosomal wild-type alleles but carried mutants on the plasmids were tested on media lacking adenine or lysine, respectively. Mutants able to grow at 25 °C but not at 37 °C were isolated as temperature sensitive (*ts*). This phenotype is a reliable and often-used marker of conditional structural instability as opposed to fatal damage (Pakula and Sauer 1989; Chakshusmathi et al. 2004). We selected seven *ade2(ts)* and six *lys2(ts)* mutants for further study (fig. 1). The *ade2(ts)* mutants contained only single and unique amino acid substitutions. The *lys2(ts)* mutants also contained unique substitutions but more than one (three or more), see supplementary table S1, Supplementary Material online. *LYS2* is a much larger target for mutations than *ADE2*, 4,179 versus 1,716 bp, and thus, multiple mutations were more likely to occur.

We cloned the *ts* alleles into an expression vector, pBG1805 (Gelperin et al. 2005). Within this multicopy plasmid, the gene under study was preceded by an inducible *GAL1* promoter and followed by a tag-coding sequence (His6, HA epitope, protease 3C site, ZZ domain, 19 kDa). Plasmids were inserted into two yeast strains with unique fluorescent protein markers, one coding for yellow fluorescent protein (YFP) and the other for cyan fluorescent protein (CFP). To initiate competition experiments, each mutant was paired with its wild-type progenitors in two configurations of markers (YFP for wild-type and CFP for mutant and vice versa) to eliminate a possible marker effect. (See supplementary methods, Supplementary

Letter

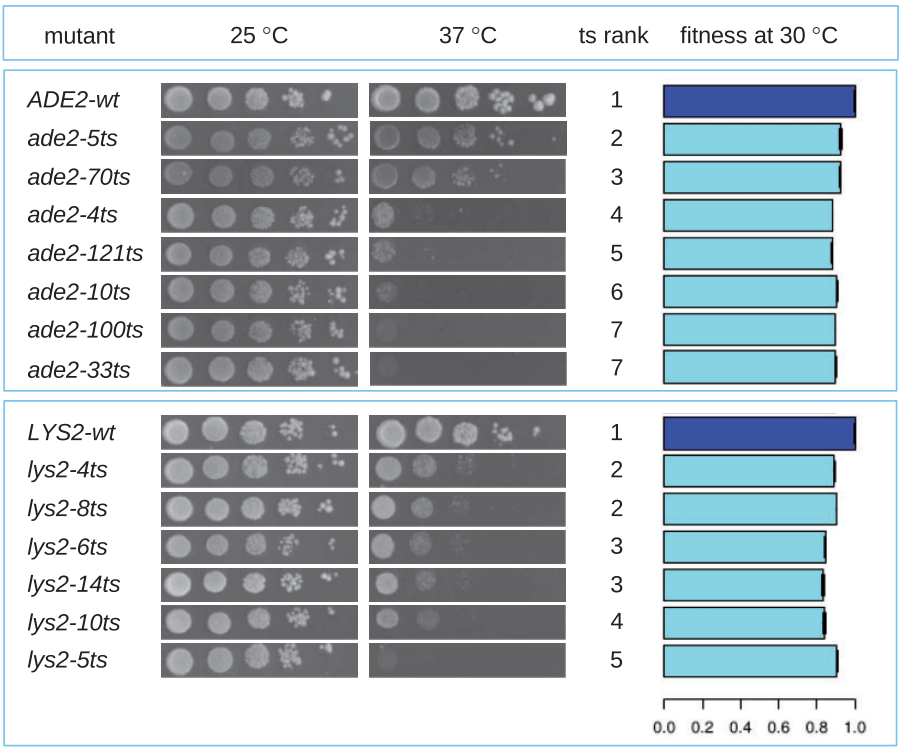

Fig. 1. Temperature sensitive (*ts*) mutants of *ADE2* and *LYS2*. The two left panels show the results of growth tests at ambient and stressful temperatures for strains carrying overexpression plasmids with either wild-type or mutant alleles. In this test, the host strain lacked a functional version of the tested gene, whereas the growth medium lacked adenine or lysine. Therefore, the observed effects mark conditional auxotrophy. The right panel shows the results of competition between strains carrying plasmids with alleles under overexpression (wild type vs. mutant). In this case, the host strains were carrying wild-type alleles on chromosomes, and the media were supplied with both adenine and lysine. Therefore, the observed effects mark the toxicity of overexpressed proteins.

Material online, for more details on the strains, media, and design of competition experiments.) In all competitions, strains overexpressing the wild-type proteins won (fig. 1). The relative fitness of mutants was considerably less than one: 0.901 ± 0.006 (mean ± SE) for the *ade2(ts)* and 0.869 ± 0.020 for the *lys2(ts)* strains.

There are several lines of evidence that differences in fitness were not caused by insufficient metabolic activity of the adenine or lysine synthesis pathways. First, adenine and lysine were added to the medium. More importantly, the strain used as a host in fitness assays was diploid with two (*ADE2*/*ADE2*) or one copy of the tested gene (*lys2Δ*/*LYS2*) being functional. We also carried out two control experiments. In one of them, we asked what factors other than protein destabilization could influence fitness. We transformed the host strain with one of the following plasmids: 1) "empty" centromeric (single-copy) plasmid, 2) empty 2μ (multicopy) plasmid, 3) 2μ with *ADE2* under *GAL1* promoter, or 4) 2μ with *LYS2* under *GAL1* promoter. We found that 2) grew slower than 1), revealing the cost of carrying multiple plasmids (substantial but not overwhelming). Furthermore, 3) and 4) grew slower than 2), demonstrating that overexpression of wild-type proteins was burdensome (not helpful, which is important in further argument). In another control, we compared heterozygotes versus homozygotes, *ADE2*/*ade2Δ* versus *ADE2*/*ADE2*, and *LYS2*/*lys2Δ* versus *LYS2*/*LYS2* and found no significant differences within the pairs. The compared pairs were then transformed with an overexpression plasmid carrying *ts* mutants of either Ade2p or Lys2p gene. The presence of plasmid did not make the growth rates of heterozygotes and homozygotes unequal. This result indicates that abundant destabilized polypeptides did not interfere with rare proficient ones. The results of both control experiments are shown in supplementary figure S1, Supplementary Material online. Finally, we ranked the mutants' phenotypes according to their temperature sensitivity (metabolic inefficiency) and correlated these ranks with fitness scores (fig. 1). There was no interdependence between the two phenotypes for either the *ade2(ts)* (Spearman's rank correlation test $r_s = 0.63$, $n = 8$, $P = 0.09$) or the *lys2(ts)* mutants ($r_s = 0.38$, $n = 7$, $P = 0.40$). In sum, the observed fitness costs could not be interpreted as a metabolic inefficiency of the adenine or lysine pathways.

We sought to relate the observed fitness effects to the amounts of overproduced protein. We carried out a competitive ELISA assay and found that wild-type proteins constituted as much as $3.7 \pm 0.1\%$ (Ade2p) and $8.9 \pm 0.1\%$ (Lys2p) of the total protein mass. Mutant proteins were approximately two to four times less abundant. Figure 2 shows how the level of protein detected by ELISA relates to fitness: the higher the level, the higher the fitness. This result is intriguing because it is implausible that the overexpressed proteins were of use (see the control experiments described earlier). However, our ELISA could detect only those protein molecules that had

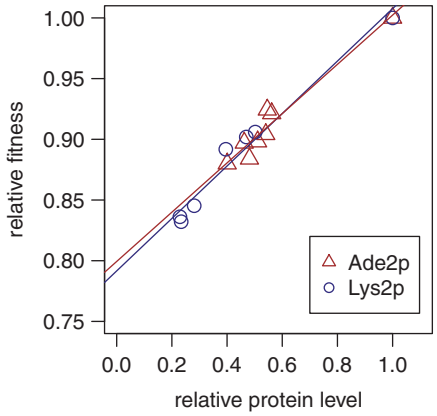

Fig. 2. Folded protein and fitness. Relation between the level of overexpressed and folded protein (detected by ELISA) and the relative fitness (competitive ability) under overexpression. Lines show reduced major axes with normalized data. Squared Pearson's coefficients were high, 0.951 and 0.972 for Ade2 and Lys2, respectively; however, they were strongly dependent on outliers. Statistical significance was confirmed with the nonparametric Spearman's test (Ade2[ts]: $r_s = 0.950$, $n = 8$, $P = 0.001$; Lys2[ts]: $r_s = 0.89$, $n = 7$, $P = 0.012$).

the ZZ domain exposed and were adequately folded, so that they could interact with IgG antibodies. Thus, it is possible that mutants with low levels of soluble proteins had large amounts of aggregates, which remained undetected but were instrumental in decreasing fitness. We tested this hypothesis by extracting total protein, denaturing and dissolving it, and running a Western blot analysis for Ade2p and Lys2p. Figure 3*A* shows that at least in case of Lys2p, the result was negative: Mutants with low levels of soluble protein (ELISA) also had low levels of total protein (Western). (See also supplementary fig. S2, Supplementary Material online, for additional results.) In the case of Ade2p, the total amount of protein was more equal, and therefore, substantial amounts of this protein were potentially present in aggregates. To test for these aggregates, we centrifuged the total cell protein to separate the soluble and insoluble fractions and then probed them for Ade2p (see supplementary methods, Supplementary Material online). Figure 3*B* shows that Ade2p was found in both fractions, including the wild-type strain. (The latter is not surprising because overexpressed proteins tend to aggregate even if their stability is normal.) However, there was no significant correlation between mutant fitness and the fraction of protein located in deposits (Pearson's $r = 0.65$, $n = 8$, $P = 0.082$). (In the case of Lys2p, a substantial portion of protein was partially degraded, making quantitative comparisons between the soluble and insoluble fractions difficult, see supplementary fig. S3, Supplementary Material online.) To recapitulate our protein assays, we found a strong signal from both proteins that fitness was low when the amount of folded tags (marking soluble and at least partly folded protein) was also low (fig. 2). The decline in fitness could not be ascribed to insoluble protein fractions because these were either not enriched, as in case of Lys2p ts mutants (fig. 3*B*), or, when these fractions were enriched (Ade2p ts), their amount did not correlate with fitness (fig. 3*B*). In sum, the diminishing amount of mutated protein (particularly its soluble fraction) was best associated with the observed decline in fitness. We assume that the need to detect, sanitize, and/or remove defective chains could impose considerable metabolic cost. Alternatively, some protein variants may have been so toxic that they poisoned the cell the most, even though they were the most effectively removed.

Considering this idea, we note that there is good evidence that toxicity is caused not by big and largely inactive aggregates but rather by those smaller and more interactive ones (Bhatia et al. 2000; Conway et al. 2000; Chiti and Dobson 2006). Some of those small aggregates can be dismembered and refolded or gathered into stores of polypeptides for future recycling. Other are degraded in cytosol or sanitized as insoluble protein deposits and then removed through autophagy (Kaganovich et al. 2008; Buchberger et al. 2010). It is tempting to suggest the first scenario for our wild-type strains and the least affected mutants while hypothesizing the second scenario for those strains having low fitness and low amounts of mutant protein. We realize, however, that aggregates can show many other activities: sequestering critical proteins (Olzscha et al. 2011), depleting chaperones (Roodveldt et al. 2009), and interfering with membranes (Stefani 2008), translocons (Chakrabarti et al. 2011), or numerous other cellular targets (Winklhofer et al. 2008). The above list of molecular processes involving unstable mutants was compiled over years of intense research; it would be infeasible to verify experimentally how relevant each of these processes is to our mutants. However, the finding that variants of the same polypeptide, often differing by only one amino acid, can adopt apparently different routes of sanitization/removal is revealing and can induce further research on this subject.

To test whether our mutant proteins would be toxic if expressed at normal levels, we focused on a set of *ade2(ts)* alleles, some of which were described here (*ade2-5ts*, *ade2-10ts*) and others of which were described in an earlier study (numbered in that study as 1–3, 5, 7–9) (Tomala and Korona 2008). Strains with *ts* alleles residing in their chromosomal locus against a wild-type allele (*ade2ts/ADE2*) were competed against a control strain (*ade2Δ/ADE2*) (see supplementary methods, Supplementary Material online, for details). Assuming a linear decrease from the cost under overexpression (0.099 of relative fitness for 3.7% of total protein), the cost under normal expression (0.005% of total protein in a hemizygote) would be 0.00013. In our experiments, fitness was estimated from regression of log-transformed ratios of competitors over time. The standard error of these estimates, 0.00034 on average, was higher than the sought effect. Because none of the eight tested mutants was significantly less fit than the control strain, the actual fitness effects were not substantially higher than predicted. In fact, they could have been smaller because the assumption about linear scaling of the fitness effect with the amount of destabilized protein is only hypothetical. However, we consider it unlikely that natural amounts of unstable protein are so much less toxic than expected that selection is unlikely to act on them. This is because the effective population size of the budding yeast is approximately 10^7, and the fitness differential required for selection is close to its inverse (Tsai et al. 2008).

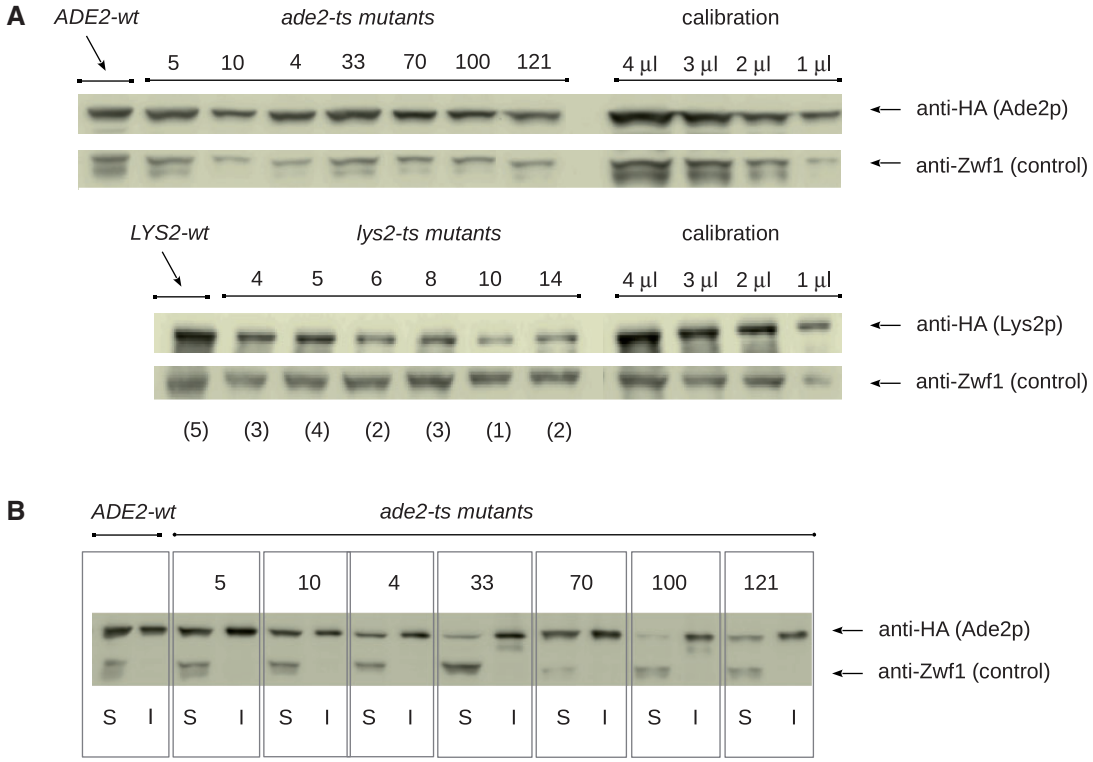

Fig. 3. Total amount of protein. (*A*) The level of total protein (detected by Western blot). The amounts of Ade2p are generally similar to each other; they are difficult to rank reliably. The amounts of Lys2p are highly variable and can be clearly ranked (the ranks are shown as numbers in parentheses). A Spearman's test shows high correlation between these ranks and the scores of the ELISA assay ($r_s = 0.980$, $n = 7$, $P = 0.000083$). (*B*) Relative abundances of the soluble (S) and insoluble (I) protein fractions for Ade2(p). A control protein (here Zwf1) expressed at normal levels is not recovered from aggregates (as expected). One rectangle in (*B*) group proteins extracted from one culture. Assays presented in (*A*) and (*B*) were repeated twice, beginning with independent culturing and protein isolation. The other replicas are presented, in the same way as here, in supplementary figure S2, Supplementary Material online. The soluble and insoluble fractions of Lys2p are not shown in (B). We found that both fractions were present. However, their relative amounts were difficult to quantify. Our attempts are reported in supplementary figure S3, Supplementary Material online.

The per-molecule toxicity would have to drop by some three orders of magnitude between overexpression and normal expression to make the mutations neutral. Admittedly, evidence from direct measurements would surpass extrapolation. Significant refinement of known fitness assays or development of entirely new ones are clearly needed (Blomberg 2011).

In sum, we consider it likely that single/few substitutions can destabilize proteins strongly enough to incite purifying selection. Moreover, they may be ubiquitous because we found a number of them in just two proteins which were experimentally tractable but otherwise arbitrarily chosen.

Supplementary Material

Supplementary table S1, figures S1–S3, and methods are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Daniel Hartl for reading and commenting our manuscript. This work was supported by the Foundation for Polish Science (a "Mistrz" grant to R.K.), a National Science Centre grant no. 2011/01/B/NZ8/00042 to K.T. and DS/MND/WBiNoZ/INoS/21/2012 to E.P.

References

Bershtein S, Mu W, Shakhnovich EI. 2012. Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations. *Proc Natl Acad Sci U S A.* 109:4857–4862.

Bhatia R, Lin H, Lal R. 2000. Fresh and globular amyloid beta protein (1-42) induces rapid cellular degeneration: evidence for AbetaP channel-mediated cellular toxicity. *FASEB J.* 14:1233–1243.

Blomberg A. 2011. Measuring growth rate in high-throughput growth phenotyping. *Curr Opin Biotechnol.* 22:94–102.

Buchberger A, Bukau B, Sommer T. 2010. Protein quality control in the cytosol and the endoplasmic reticulum: brothers in arms. *Mol Cell.* 40:238–252.

Chakrabarti O, Rane NS, Hegde RS. 2011. Cytosolic aggregates perturb the degradation of nontranslocated secretory and membrane proteins. *Mol Biol Cell.* 22:1625–1637.

Chakshusmathi G, Mondal K, Lakshmi GS, Singh G, Roy A, Ch RB, Madhusudhanan S, Varadarajan R. 2004. Design of temperature-sensitive mutants solely from amino acid sequence. *Proc Natl Acad Sci U S A.* 101:7925–7930.

Chiti F, Dobson CM. 2006. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem.* 75:333–366.

Conway KA, Lee SJ, Rochet JC, Ding TT, Williamson RE, Lansbury PT Jr. 2000. Acceleration of oligomerization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset

Parkinson's disease: implications for pathogenesis and therapy. *Proc Natl Acad Sci U S A.* 97:571–576.
DePristo MA, Weinreich DM, Hartl DL. 2005. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat Rev Genet.* 6:678–687.
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, Wilke CO. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134:341–352.
Eames M, Kortemme T. 2012. Cost-benefit tradeoffs in engineered lac operons. *Science* 336:911–915.
Ellis RJ. 2001. Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci.* 26:597–604.
Geiler-Samerotte KA, Dion MF, Budnik BA, Wang SM, Hartl DL, Drummond DA. 2011. Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. *Proc Natl Acad Sci U S A.* 108:680–685.
Gelperin DM, White MA, Wilkinson ML, Kon Y, Kung LA, Wise KJ, Lopez-Hoyo N, Jiang L, Piccirillo S, Yu H, et al. 2005. Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes Dev.* 19:2816–2826.
Gsponer J, Babu MM. 2012. Cellular strategies for regulating functional and nonfunctional protein aggregation. *Cell Rep.* 2:1425–1437.
Hamley IW. 2007. Peptide fibrillization. *Angew Chem Int Ed Engl.* 46: 8128–8147.
Hartl FU, Bracher A, Hayer-Hartl M. 2011. Molecular chaperones in protein folding and proteostasis. *Nature* 475:324–332.
Kaganovich D, Kopito R, Frydman J. 2008. Misfolded proteins partition between two distinct quality control compartments. *Nature* 454: 1088–1095.
Lehner B. 2011. Molecular mechanisms of epistasis within and between genes. *Trends Genet.* 27:323–331.
Olzscha H, Schermann SM, Woerner AC, Pinkert S, Hecht MH, Tartaglia GG, Vendruscolo M, Hayer-Hartl M, Hartl FU, Vabulas RM. 2011. Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell* 144:67–78.
Pakula AA, Sauer RT. 1989. Genetic analysis of protein stability and function. *Annu Rev Genet.* 23:289–310.
Plata G, Gottesman ME, Vitkup D. 2010. The rate of the molecular clock and the cost of gratuitous protein synthesis. *Genome Biol.* 11:R98.
Roodveldt C, Bertoncini CW, Andersson A, van der Goot AT, Hsu ST, Fernandez-Montesinos R, de Jong J, van Ham TJ, Nollen EA, Pozo D, et al. 2009. Chaperone proteostasis in Parkinson's disease: stabilization of the Hsp70/alpha-synuclein complex by Hip. *EMBO J.* 28: 3758–3770.
Serohijos AW, Rimas Z, Shakhnovich EI. 2012. Protein biophysics explains why highly abundant proteins evolve slowly. *Cell Rep.* 2: 249–256.
Stefani M. 2008. Protein folding and misfolding on surfaces. *Int J Mol Sci.* 9:2515–2542.
Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M. 2007. Life on the edge: a link between gene expression levels and aggregation rates of human proteins. *Trends Biochem Sci.* 32: 204–206.
Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M. 2009. A relationship between mRNA expression levels and protein solubility in *E. coli*. *J Mol Biol.* 388:381–389.
Tomala K, Korona R. 2008. Alleviation of deleterious effects of protein mutation through inactivation of molecular chaperones. *Mol Genet Genomics.* 280:409–417.
Tsai IJ, Bensasson D, Burt A, Koufopanou V. 2008. Population genomics of the wild yeast *Saccharomyces paradoxus*: quantifying the life cycle. *Proc Natl Acad Sci U S A.* 105:4957–4962.
Winklhofer KF, Tatzelt J, Haass C. 2008. The two faces of protein misfolding: gain- and loss-of-function in neurodegenerative diseases. *EMBO J.* 27:336–349.
Yang JR, Zhuang SM, Zhang J. 2010. Impact of translational error-induced and error-free misfolding on the rate of protein evolution. *Mol Syst Biol.* 6:421.
Zur H, Tuller T. 2012. Strong association between mRNA folding strength and protein abundance in *S. cerevisiae*. *EMBO Rep.* 13: 272–277.