æ

18. INTRACELLULAR ERRORS

3 November 2017

As noted repeatedly in prior chapters, few (if any) cellular processes have been pushed to the limits of perfection dictated by the laws of physics. A likely reason for this is the barrier to natural selection imposed by random genetic drift, combined with the recurrent introduction of deleterious mutations. As a consequence of imperfections in the cellular machinery, cells make errors, which if not removed are expected to lead to progressive damage, resulting in elongated cell-division times and/or shortened lifespans. The challenges are often quite multifaceted. For example, the production of properly constructed proteins requires the avoidance of problems potentially arising via dozens of cellular processes (Figure 18.1).

Cells have evolved multiple processes that seemingly mitigate the physiological consequences of error production, . For example, the incidence of errors arising during the replication of new DNA strands (heritable mutations) is reduced by a proof-reading domain in the primary DNA polymerases (Chapter 8), and problems arising in nonreplicating DNA are dealt with by a variety of repair mechanisms. Mechanisms also exist for the detection and elimination of some types of erroneous transcripts, and some of the stages leading to translation involve proofreading mechanisms. A wide array of other cellular processes, including interactions of enzymes with inappropriate substrates (Chapter 17) and faulty assembly of proteins and their higher-order structures (Chapter 12), are subject to error.

This chapter will focus entirely on the rates at which errors arise at the levels of transcription and translation, the mechanisms by which these are mitigated, the energetic burden of error surveillance, and the magnitude of selection operating to increase the fidelity of transcription and translation. Transcription and translation errors arise at rates orders of magnitude higher than those incurred during replication, and this is likely an evolutionary consequence of the transient nature of such errors. Unlike replication errors, which create cumulative damage in regions linked to mutator alleles, alleles associated with phenotypic errors will generally become decoupled from their source in no more than a single generation, reducing the strength of purifying selection against mechanisms of error production. The effects of such errors may also be diminished by complementation from the frequently large pool of error-free molecules.

This scenario of diminished selection intensity raises the question as to how cells have evolved an array of mechanisms for error surveillance at the transcript and translational levels, and why error rates are so high despite the presence of various proof-reading mechanisms. A key point made below is that although multiple layers of surveillance lead to the impression of a highly refined system, and to the common

assertion that cells are robust to perturbations, the overall level of performance is likely no greater than that possible with a much simpler system. Such a conclusion is entirely consistent with the drift barrier on the total performance of a system being distributed over multiple traits.

Transcript Fidelity

The first step in the successful development of a gene product is the generation of an appropriate RNA transcript from the underlying genomic sequence. The RNA polymerases responsible for transcription are typically comprised of several protein subunits. In eukaryotes, one of these complexes (Pol II) is reserved for the production of messenger RNAs and micro RNAs, another (Pol I) for the synthesis of the ribosomal RNA subunits, and a third (Pol III) for transfer RNA production (Werner and Grohmann 2011). Land plants deploy two additional RNA polymerases to generate small RNAs used in transcriptional silencing (Wierzbicki et al. 2009; Haag and Pikaard 2011); these seem to be derived from Pol II, but are highly divergent in sites that are otherwise conserved in Pols I-III, suggesting the possibility of reduced fidelity (Luo and Hall 2007; Landick 2009). In contrast, prokaryotes use just a single RNA polymerase (designated below as RNAP) to service all genes.

Despite their shared functions, the complexity of these enzymes is quite variable, with eubacterial and archaeal RNAPs consisting of 5 and 12 subunits respectively, eukaryotic Pols I and III containing 14 and 17 subunits respectively (Carter and Drouin 2010), and Pols II, IV, and V all comprised of 12 (Haag and Pickaard 2011). Yet, as will be noted below, despite the fact that eukaryotic RNA polymerases contain more than twice the number of components as those from eubacteria, there is no evidence that the former carry out their tasks more efficiently.

Like replication, transcription involves phases of initiation, elongation, and termination, but several additional processing steps are involved as well (e.g., $5'$-capping, intron removal, and addition of poly-A to $3'$ ends). Problems may arise at each of these stages: 1) transcription initiation at an inappropriate location, which can be particularly disastrous if it occurs downstream from the translation-initiation site; 2) inaccurate removal of introns (a problem largely confined to eukaryotes, and especially significant in multicellular lineages, which typically average five or more introns per protein-coding gene); 3) premature transcriptional termination (prior to the translation-termination codon being particularly detrimental); and 4) base-substitution and insertion/deletion errors.

Errors of the first three types are thought to be common (e.g., Suzuki et al. 2000; Frith et al. 2006), with Struhl (2007) suggesting that 90% of transcription initiation by Pol II in yeast is nothing more than transcriptional noise. However, it is often difficult to know whether all aberrant transcripts are actually errors as opposed to being indirectly involved in various kinds of regulatory pathways. The remainder of this section will be focused simply on base-substitution errors.

Although the mechanisms underlying transcription fidelity are not fully understood, they differ from those involved in replication (Chapter 8) in that RNA polymerases have no proof-reading domains. Instead, transcriptional accuracy relies largely on correct nucleotide recruitment, with incorrect base incorporation simply

resulting in polymerase pausing, which provides time for the recruitment of factors for removing the dangling base (Zenkin et al. 2006; Sydow and Cramer 2009; Sydow et al. 2009; Kaplan 2010; Yusenkova et al. 2010). In contrast to the situation with replication, there is no known mechanism for correcting errors after chain elongation is complete (e.g., by recognizing mismatches in the DNA-RNA hybrid).

As the life spans of individual transcripts are generally substantially shorter than the longevities of cells, transcription must progress at fairly high rates to meet cellular demands. The few estimates of average speeds of progression exhibit only a small range of phylogenetic variation: 46 bp/sec in *E. coli* (Golding and Cox 2004; Proskin et al. 2010); 20 to 60 in yeast (Mason and Struhl 2005; Larson et al. 2012; Lisica et al. 2016; Ucuncuoglu et al. 2016); 21 in *Drosophila* (Ardehali and Lis 2009); 23 in rat (Ardehali and Lis 2009); and 56 in human (Ardehali and Lis 2009). In contrast, replication rates are typically in the range of 100 to 1000 bp/sec in prokaryotes (Hiriyanna and Ramakrishnan 1986; Stillman 1996; Myllykallio et al. 2000), but just 10 to 50 bp/sec in yeast, flies, and mammals (summarized in Lynch 2007). Thus, transcription is much slower than replication in prokaryotes, whereas both processes proceed at comparable rates in eukaryotes.

Erroneous transcripts may typically arise by simple copying errors, although these may sometimes be exacerbated by base damage within the gene itself, as transcription often proceeds across damaged sites by simply substituting an A (Brégeon et al. 2003; Clauson et al. 2010). Post-transcriptional errors cannot be ruled out, which is the reason the term transcript- rather than transcriptional-error is adhered to here. One might think that errors in transcripts could easily be enumerated by simply comparing the sequences of cDNAs to their genomic sources. However, the sequencing-error rate is generally substantially greater than the transcription-error rate, so this approach will not work. Thus, until recently almost all information on transcript-error rates has been derived via indirect *in vitro* methods, generally by measuring the relative incorporation rates of two competing nucleotides across a specified template.

The average of three *E. coli* estimates obtained with this kind of approach (which themselves exhibit a 40-fold range of variation) is 1.4×10^{-4} (Springate and Loeb 1975; Blank et al. 1986; Goldsmith and Tawfik 2009), whereas the thermophilic bacterium *Thermus aquaticus* has an estimated error rate of 6.5×10^{-4} per nucleotide site (Yuzenkova et al. 2010). The sole reporter-construct estimate for budding yeast *S. cerevisiae* is 2.0×10^{-6} (Kireeva et al. 2008; Walmacq et al. 2009), whereas a single estimate for wheat is 2.4×10^{-4} (de Mercoyrol et al. 1992). These early estimates, most likely quite subject to experimental biases, suggest that transcription-error rates (in terms of sequence fidelity) fall in the broad range of 10^{-5} to 10^{-3} per nucleotide site.

More recently, it has become possible to estimate the genome-wide *in vivo* error rate by directly sequencing individual mRNA molecules multiple times (Gout et al. 2013; Traverse and Ochman 2016; Li et al., in prep.) These rates are generally lower than the indirect estimates, all falling in the range of 10^{-6} and 10^{-4} (Table 18.1), and revealing no obvious phylogenetic pattern (McCandlish and Plotkin 2013). The highest direct estimates are all from one study (Traverse and Ochman 2016), and methodological differences may underlie much of the variation, as our own estimate of the error rate for *E. coli,* 8×10^{-6}, is ten times lower than the estimate for the

same species by Traverse and Ochman (2016). If their estimates are anomalous, the range in known transcription-error rates shrinks to 10^{-6} to 10^{-5}. Thus, the earlier *in vitro* estimates are anomalously high.

To put these rates into perspective, our estimate of the transcript-error rate in *E. coli* exceeds the known genomic mutation rate per nucleotide site in this species (Lee et al. 2012) by a factor of 40,000. For all other species for which both rates are known, the transcript-error rate is inflated by factors of 2,000 to 34,000. Thus, there is little question that transcription (and/or the downstream accumulation of damage in circulating transcripts) is substantially more error-prone than DNA replication, as previously suggested by Ninio (1991a,b). The probability of a base-substitution error in a small mRNA of ~ 1000 bp in length is expected to be on the order of 0.1 to 1.0%.

Table 18.1. Estimated error rates associated with transcription and translation. Standard errors (in parentheses), where calculable, are derived from results of independent studies. Transcript-error rates and total translation-error rates refer to single nucleotide and amino-acid substitutions, respectively. From Lynch (in prep.)

Species	Transcription ($\times 10^{-4}$)	Synthetase Loading	Total Translation	Translation Readthrough
Prokaryotes:				
Agrobacterium tumefaciens	0.099			
Bacillus subtilis	0.077			
Buchnera aphidicola	0.823			
Carsonella ruddii	0.509			
Escherichia coli	0.452	0.0011 (0.0004)	0.0052 (0.0027)	0.0050 (0.0017)
Mesoplasma florum	0.154			
Mycobacterium smegmatis			0.0168 (0.0147)	
Salmonella typhimurium				0.0034 (0.0014)
Five bacteria		0.0017 (0.0009)		
Eukaryotes:				
Saccharomyces cerevisiae	0.040	0.0055 (0.0018)	0.0162 (0.0124)	0.0111 (0.0047)
Arabidopsis thaliana	0.115			
Lupinus luteus		0.0009 (0.0005)		
Caenorhabditis elegans	0.040			
Mus musculus	0.150		0.0152 (0.0065)	0.0042
Homo sapiens		0.0004 (0.0001)	0.0147 (0.0146)	

Translational Fidelity

Even when an mRNA emerges error-free, several additional challenges must be met if a faithful protein product is to be synthesized (Parker 1989; Zaher and Green 2009). First, specific amino-acyl synthetase proteins (AARSs) must initially harvest their cognate amino acids. Second, charged AARSs must then pass their cargo on to the appropriate transfer RNA (tRNA). There is considerable room for error in both of these steps because the structural differences between some amino acids

are quite minimal (e.g., valine and isoleucine differ by the presence of just a single methyl group). AARSs are endowed with proof-reading mechanisms to minimize misloading errors (Hussain et al. 2010; Reynolds et al. 2010), although some species of *Mycoplasma* have lost the capacity for proof-reading in multiple synthetases (Li et al. 2011; Yadavalli and Ibba 2013). Third, at the ribosome, each codon in an mRNA must be recognized by its cognate tRNA via codon:anticodon recognition. Proof-reading appears to occur twice after initial tRNA loading, involving processes that require GTP hydrolysis (Ieong et al. 2016).

Infidelities at any one of these steps can lead to a diversity of errors in translated products. For example, misreads of sense codons can lead to alterations in protein structure/function, with misreading as termination codons leading to prematurely truncated chains. Misreading of termination codons as sense codons leads to termination read-through.

Despite these immediate functional problems, error rates at the level of translation are even higher than transcript-error rates. Most attempts to estimate the rate at which AARSs are mischarged have involved *in vitro* competition experiments between cognate and noncognate amino acids. These measures are only rough estimates of likely *in vivo* error rates for two reasons. First, such evaluations almost always involve simple binary experiments, leaving questions as to the total error rate expected in a more natural setting where all twenty amino acids are present simultaneously. Second, most binary tests have focused on the loading of erroneous amino acids with physical features most similar to the cognate substrate of the focal AARS, raising the additional caveat that pairwise estimates of misloading rates may be upwardly biased.

Using this approach, the average pairwise misloading rate for a variety of species ranges from 0.0004 to 0.0055 (Table 18.1). As these estimates have been obtained with different methods, different AARSs, and different pairs of cognate and noncognate amino acids, no conclusions can be drawn with respect to phylogenetic differences in AARS loading fidelity. Data on the rate of mischarging of tRNAs (by inappropriate AARSs) are scant, but the few available estimates are of order 10^{-3} per tRNA (Yadavalli and Ibba 2013; Shepherd and Ibba 2014). Thus, given the sum of known AARS and tRNA misloading rates, it appears that the potential for translation error even before a transcript meets a ribosome is far higher than the transcript-error rate. Additional errors in translation will arise at the level of codon reading, although the only detailed estimates are for *E. coli,*, which fall in the range of 10^{-7} to 10^{-4} (Zhang et al. 2016). Thus, 10^{-3} would appear to be the lower limit to the total error rate per codon for the species that have been examined.

Several attempts have been made to estimate the total rate of amino-acid misincorporation into protein (which summarizes the net consequence of errors in all preceding steps, including transcription). As it is not easy to sequence single amino-acid chains, such studies have often been performed with target genes that do not encode a particular amino acid, and then searching for the incidence of that amino acid in synthesized proteins. In other cases, genes have been engineered to produce defective products unless a particular codon is misloaded by a specific amino acid, with the degree of rescue providing insight into the specific error rate at that one codon. As both of these methods are limited with respect to the amino-acid misloadings that can be detected, to obtain the total translation-error rate, correc-

tion needs to be made to account for the likely errors involving all amino acids not monitored (Foundations 18.1). There are potential biases associated with such correction, but with these caveats in mind, average *in vivo* translation-error estimates (per codon) fall in the range of 0.005 to 0.017 (Table 18.1), substantially greater than the AARS-misloading rates.

A rough check on these numbers can be acquired from observations on another type of translational error – misreading of a termination codon as a sense codon (Parker 1989). Typically, such studies monitor the expression of reporter constructs containing premature termination codons that completely abrogate gene function unless experiencing read-through. The read-through error rate is then estimated as the fraction of gene expression relative to that for an intact gene copy. Although there can be substantial variation in the read-through rate depending on the local context of the nonsense codon, most studies average over several such sites, with reported rates of nonsense-codon misreading ranging from 0.003 to 0.011 (Table 18.1).

Thus, despite the uncertainties involved, the translation-error rate per codon appears to be on the order of $10,000\times$ greater than the transcript-error rate per nucleotide site. This means that only a small fraction of errors at the protein level are associated with transcription. Assuming an average translation-error rate of 0.01 per codon, a newly synthesized protein of moderate size of 300 amino acids would contain an average of three erroneous amino acids, and assuming a Poisson distribution of errors, only a fraction $e^{-3} = 0.05$ of proteins of this size would be error free. For large complexes, such as the ribosome involving $\sim 10,000$ amino acids summed over all subunits, essentially every composite structure would be expected to contain errors.

The implications here are that within a population of genetically uniform cells, each cell will harbor a statistically and transiently unique distribution of variants for most proteins. Recalling Equation 2.2a, which predicts the numbers of protein molecules within cells, and again assuming an average of 300 amino acids per protein, a bacterial-sized cell of $\sim 1\ \mu\text{m}^3$ is expected to contain $\sim 5 \times 10^6$ protein-sequence errors. Average protein lengths are more on the order of 500 amino acids in eukaryotes, so a yeast-sized cell of $\sim 100\ \mu\text{m}^3$ can be expected to harbor $\sim 6 \times 10^8$ errors, and a moderate sized eukaryotic cell of $\sim 10^5\ \mu\text{m}^3$ would harbor $\sim 4 \times 10^{10}$ errors.

Biophysical Limits to Substrate Discrimination and the Cost of Proofreading

As in any cellular process, there are two potential limits to which a level of molecular perfection can be approached. The biophysical barrier represents the ultimate limits that could be achieved by a supreme biochemist, capable of constructing limited only by diffusion limitations and energetic features of various substrate-binding mechanisms. The evolutionary barrier is the boundary set by the degree to which random genetic drift reduces the efficiency of natural selection. The issues associated with each of these barriers are considered in the following two sections.

Errors arise during transcription and translation as statistical consequences of

random diffusion and attachment of alternative substrates to catalytic sites, with the frequencies of usage depending on the relative concentrations and binding affinities involved. As the binding affinities of two alternative substrates become arbitrarily close, both substrates would be equally likely to bind to the enzyme provided the substrate concentrations are the same, although with unequal concentrations competitive binding trend towards favoring the more abundant substrate. As the difference in binding energies increases, the relative rate of an enzyme engaging with an inappropriate substrate declines exponentially. These issues can be addressed more formally as follows.

Under a simple competitive binding situation, the limit to accuracy can be formally evaluated for two competing Michaelis-Menten reactions by considering the ratio of rates of engagement with incorrect (W) and correct (R) substrate molecules (Foundations 18.1),

$$E = \frac{[\mathrm{W}]}{[\mathrm{R}]} \cdot \frac{k_{\mathrm{d,R}}}{k_{\mathrm{d,W}}}, \tag{18.1a}$$

where the first term represents the ratio of substrate concentrations, and the second term is the ratio of dissociation constants. The latter are functions of the binding energies between substrates and enzyme, and their ratio can be represented in statistical-mechanic terms relative to the background energy of the system (see Foundations 2.3; and pp. 1011-1016 in Phillips et al. 2013),

$$E = \frac{[\mathrm{W}]}{[\mathrm{R}]} \cdot e^{-\Delta E/(k_B T)}, \tag{18.1b}$$

where ΔE is the difference of binding strengths involving correct and incorrect substrate molecules, k_B is the Boltzmann constant (Chapter 2), and T is the temperature in degrees Kelvin. The actual error rate (the fraction of incorrect reactions, ϵ) is equivalent to $E/(1+E)$, which is essentially the same as E provided $E \ll 1$. Assuming that this condition is met and that concentrations of alternative substrates are equal,

$$\epsilon \simeq e^{-\Delta E/(k_B T)}. \tag{18.2}$$

To gain an appreciation for the biological limits to accuracy under simple competitive binding, note that most enzymes bind their specific substrates with energies in the range of 12 to 24 $k_B T$ (Kuntz et al. 1999). The strength of a single hydrogen bond, 5 to 15 $k_B T$ depending on the context (Fersht 1999), puts this in perspective, as a G:C pairing involves three hydrogen bonds, whereas an A:T pairing involves two. It then follows that binding-strength differentials between preferred and nonpreferred substrates (ΔE) are generally smaller than $\sim 10\ k_B T$, rendering most biological processes to be error-prone. For example, taking 2 or 5 $k_B T$ to be binding-energy differences between two substrates yields error rates of $\epsilon \simeq 0.13$ and 0.007, respectively, whereas extreme differences of 10 and 15 $k_B T$ still yield $\epsilon \simeq 5 \times 10^{-5}$ and 3×10^{-7}, respectively. Even the latter is substantially higher than known DNA replication error rates (Chapter 8), showing that replication fidelity must involve processes beyond simple single-step competitive binding of alternative nucleotides to single-stranded DNA.

Hopfield (1974) and Ninio (1975) realized how a simple deviation from linear Michaelis-Menten enzyme kinetics can lead to a dramatic amplification in enzyme

fidelity (Foundations 18.1). The key point is that if an enzyme can pause long enough for substrate molecules to dissociate before completing a reaction, this opens the opportunity for the repeated interrogation of a population of alternative substrates. Molecules that are less likely to dissociate before conversion to the final product will then be utilized more frequently. However, this path to increased accuracy comes at a cost, as all known proofreading processes consume energy in the form of ATP or GTP hydrolyses. For example, an ATP hydrolysis is required each time an amino acid is attached to an AARS, and this must be repeated whenever a substrate molecule is rejected prior to tRNA attachment.

This energetic cost to proofreading was first shown directly without any detailed knowledge of the underlying mechanism. For example, Hopfield et al. (1976) considered an *in vitro* system involving the transfer of either isoleucine or valine to the isoleucine tRNA via isoleucyl-tRNA synthetase. When isoleucine was the sole substrate, 1.6 ATP hydrolyses occurred per charged isoleucyl-tRNA (implying that a correct substrate molecule is examined 1.6 times prior to permanent attachment), whereas 270 ATPs were consumed per charged tRNA when only valine was presented. These results suggest an error rate of $\sim 1.6/270 = 0.006$ resulting from the differential rejection of the two residue types. In additional work, Yamane and Hopfield (1977) found that 25 to 40 ATPs are consumed when properly charged AARSs are forced to deliver an amino acid to a noncognate tRNA, implicating energy-consuming proofreading at the tRNA stage. Finally, when properly charged tRNAs encounter inappropriate codons, GTP is hydrolyzed, implicating additional proofreading at the codon-anticodon recognition step on the ribosome (Thompson and Stone 1977; Yates 1979).

In a more general analysis of the energetic cost of proofreading, Andersson et al. (1986) found that hyper-accurate ribosomes in *E. coli* require about twice the number of GTPs to produce a peptide bond as in wild-type cells, presumably as a consequence of the increased number of rejection cycles per accepted amino acid. Likewise, Muzyczka et al. (1972) found that bacteriophage DNA polymerases with mutations in their proofreading domains hydrolyzed more or less nucleotides relative to wild-type when they were antimutators vs. mutators.

One can view the energetic cost of proofreading at two levels: the baseline cost of multiple interrogations of correct substrate molecules; and the additional cost incurred by engaging with inappropriate substrates. As an example of the first cost, consider the prior example in which 1.6 ATP hydrolyses are experienced per correct substrate molecule. This implies an intrinsic cost of proofreading of 0.6 ATPs per residue incorporated, as just 1 ATP would be consumed if the preferred substrate was never rejected.

The total energetic cost of erroneous substrate removal cannot be inferred without a detailed knowledge of the relative concentrations of alternative substrates and their kinetic coefficients. However, a rough idea can be gained by noting that the rate of amino-acid misloading is approximately equal to the reciprocal of the average number of cycles through the proofreading system per loading event, whereas the cost (in ATP hydrolyses) per misloading is equal to 1 ATP per cycle. This suggests that the energetic cost of proofreading may be largely independent of the relative binding energies of alternative substrates, as noncognate substrate molecules that bind with affinities closer to those of the cognate substrate are advanced more

rapidly through the system, and consume less ATP in doing so.

As an explicit example following from the preceding paragraph, assuming equal concentrations of isoleucine and valine molecules within the cytosol implies an additional consumption of $0.006 \cdot (270 - 1.6) = 1.6$ ATP molecules per fixed valine, i.e., a doubling in ATP consumption relative to the situation in which valine is the only substrate. The total cost of proofreading by this particular AARS is therefore approximately three-fold (~ 3.2 ATP hydrolyses per amino-acid incorporation, as opposed to 1.0). These observations imply that the evolution of proofreading mechanisms, which arose very early in cellular evolution, imposed an energetic cost equivalent to at least a doubling or tripling in the consumption of ATP molecules for each proofreading step involving DNA/RNA transactions.

From Chapter 3, the cost of building a cell scales nearly linearly with cell volume, averaging $\sim 27V$ billions of hydrolyzed ATPs (where V is cell volume in units of μm^3), whereas from Chapter 2, the number of protein molecules (in millions) per cell is $\simeq 1.6V$. Assuming 400 amino-acids per protein (an approximate average over prokaryotes and eukaryotes), two ATPs consumed by proofreading per amino-acid incorporation, and ignoring protein turnover, the fraction of a cell's total growth budget allocated to proofreading is $\sim 5\%$. Using a rather different approach, and earlier, less certain numbers, Savageau and Freter (1979) obtained an estimate of 2%. These are rough estimates, but even if somehow overestimated by a factor of ten, it is clear that surveillance at the level of translation consumes a substantial fraction of a cell's energy budget. In principle, there are no limits to the level of accuracy that can be achieved by kinetic-proofreading mechanisms (Foundations 18.1), but each increment in accuracy will involve additional energetic costs on the cell, hence reducing cellular rates of reproduction.

Foundations 18.1. Kinetic proofreading. Numerous cell biological processes, including DNA replication and mRNA translation, involve proofreading steps. As first suggested by Hopfield (1974) and Ninio (1975), such processes exploit weak binding energies and the even smaller differences between correct and incorrect substrates to repeatedly interrogate bound substrates until passing them on to the next biochemical stage. in principle, there are no limits to the level of accuracy that can be achieved by such mechanisms, but any increase in accuracy comes at the cost of increased reaction times and energy consumption.

To gain an appreciation for the mechanisms by which proofreading can lead to a quantum leap in accuracy, we first consider the null situation in which two alternative substrates are engaged in the same Michaelis-Menten reaction (Chapter 17), albeit at different rates. The correct and incorrect substrates will be designated R (right) and W (wrong) or as X when nonspecified, with both being processed by the same enzyme E. Recall that under the Michaelis-Menten model, before the final product is arrived at, an enzyme-substrate complex EX is formed (Figure 18.2), which then either returns to the prior state by dissociation or proceeds to product formation, at rates $k_{\text{d,X}}$ and $k_{\text{cat,X}}$, respectively.

The rate of production of the correct intermediate is $k_{\text{a,R}}[\text{R}]$, where the brackets denote concentration, and once formed the intermediate is converted to final product with probability $\lambda_{\text{R}} = k_{\text{cat,R}}/(k_{\text{d,R}} + k_{\text{cat,R}})$, with similar expressions following for the incorrect substrate. Provided the error rate is $\ll 1$, it will be very closely approximated

by the ratio of the forward rates for the incorrect and correct products (see text),

$$\epsilon = \frac{k_{\mathrm{a,W}} \cdot \lambda_{\mathrm{W}} \cdot [\mathrm{W}]}{k_{\mathrm{a,R}} \cdot \lambda_{\mathrm{R}} \cdot [\mathrm{R}]}, \tag{18.2.1a}$$

which can be seen to depend on ratios involving both substrate concentrations and kinetic coefficients. Removal of the extrinsic effect of concentration yields the error rate under the assumption of equal concentrations of the two substrates, and further supposing the association rates of the two substrates to be the same, which is reasonable if encounters are based on diffusion of two similar sized particles,

$$\epsilon = \frac{k_{\mathrm{cat,W}} \cdot (k_{\mathrm{d,R}} + k_{\mathrm{cat,R}})}{k_{\mathrm{cat,R}} \cdot (k_{\mathrm{d,W}} + k_{\mathrm{cat,W}})}. \tag{18.2.1b}$$

If the catalytic rates greatly exceed the dissociation rates, the error rate may be high enough to violate the assumption of the preceding derivation, but will approach

$$\epsilon \simeq \frac{k_{\mathrm{cat,W}}}{k_{\mathrm{cat,R}} + k_{\mathrm{cat,W}}}. \tag{18.2.1d}$$

which is equal to one half when the two catalytic rates are equivalent. On the other hand, if the dissociation rates are large relative to the catalytic rates, and again assuming the latter are equivalent for both substrates, Equation 18.2.1b reduces to the ratio of dissociation rates,

$$\epsilon \simeq \frac{k_{\mathrm{d,R}}}{k_{\mathrm{d,W}}}. \tag{18.2.1e}$$

This is the limit to what can be achieved by an enzyme that discriminates solely on the basis of the sticking times (the inverse of the dissociation rates) of the substrates.

Now consider the situation in which a proofreading step is inserted into the previous scheme, designating this as the creation of a secondary complex EX^* from EX at rate m_X. Under such conditions, the secondary complex is rejected and returned to state EX with decay rate $k_{\mathrm{d}^*,\mathrm{X}}$ (Figure 18.3). This has the effect of creating a recurrent loop in the progression of a substrate molecule to the final-product step, with the number of excursions back to EX depending on the magnitude of the dissociation constant. The combination of repetitive interrogation and enhanced rejection of loosely bound complexes leads to an amplification of the level of fidelity to the appropriate substrate. A full exposition of the model can be found in Hopfield (1974) and Ninio (1975), but to focus on the central point, the simplest case will be examined here – the situation in which all rate coefficients are equal for both substrates except for the dissociation constants.

The rate of production of product $\mathrm{P_X}$ can be determined by first partitioning the series of intermediate events into net rates / probabilities associated with four subcategories,

$$\lambda_{1,\mathrm{X}} = k_{\mathrm{on}}[\mathrm{X}] \qquad \text{for } \mathrm{E} + \mathrm{X} \rightarrow \mathrm{EX} \tag{18.2.2a}$$

$$\lambda_{2,\mathrm{X}} = \frac{m}{m + k_{\mathrm{d,X}}} \qquad \text{for } \mathrm{EX} \rightarrow \mathrm{EX}^* \tag{18.2.2b}$$

$$\lambda_{3,\mathrm{X}} = \frac{k_{\mathrm{d}^*,\mathrm{X}}}{k_{\mathrm{cat}} + k_{\mathrm{d}^*,\mathrm{X}}} \qquad \text{for } \mathrm{EX}^* \rightarrow \mathrm{EX} \tag{18.2.2c}$$

$$\lambda_{4,\mathrm{X}} = \frac{k_{\mathrm{cat}}}{k_{\mathrm{cat}} + k_{\mathrm{d}^*,\mathrm{X}}} \qquad \text{for } \mathrm{EX}^* \rightarrow \mathrm{E} + \mathrm{P_X} \tag{18.2.2d}$$

The overall forward rate of production of P_X can be summarized as the series

$$\Lambda_{1,X} = \lambda_{1,X}\lambda_{2,X}\lambda_{4,X}[1 + (\lambda_{2,X}\lambda_{3,X}) + (\lambda_{2,X}\lambda_{3,X})^2 + \cdots] \tag{18.2.3a}$$

$$= \frac{\lambda_{1,X}\lambda_{2,X}\lambda_{4,X}}{1 - \lambda_{2,X}\lambda_{3,X}} \tag{18.2.3b}$$

Note that from the standpoint of the intrinsic error rate, $\lambda_{1,X}$ is irrelevant as k_{on} is the same for both substrates, so the ratio $\lambda_{1,W}/\lambda_{1,R} = [W]/[R]$. Again, factoring out the concentration effect, the error rate is

$$\epsilon = \frac{\lambda_{2,W}\lambda_{4,W}(1 - \lambda_{2,R}\lambda_{3,R})}{\lambda_{2,R}\lambda_{4,R}(1 - \lambda_{2,W}\lambda_{3,W})}, \tag{18.2.4}$$

and substituting from above yields

$$\epsilon = \frac{mk_{cat} + k_{d,R}k_{cat} + k_{d,R}k_{d^*,R}}{mk_{cat} + k_{d,W}k_{cat} + k_{d,W}k_{d^*,W}} \tag{18.2.5a}$$

Comparison to Equation 18.2.1c shows that the key difference between the error-rate function with an intermediate error-checking step is the presence of products of terms, most notably of the dissociation constants associated with each substrate. If the dissociation coefficients are large relative to m and k_{cat}, and the ratios of coefficients at both steps are the same,

$$\epsilon \simeq \left(\frac{k_{d,R}}{k_{d,W}}\right)^2. \tag{18.2.5b}$$

Thus, in the limiting case, a proofreading step can reduce the error rate down to the square of that expected in the absence of proofreading.

Finally, we can inquire as to the average number of intermediate steps that are cycled through before reaching the final product (n_c). Expanding from the logic underlying Equation 18.2.3a,

$$n_c = \sum_{i=1}^{\infty} i \cdot (\lambda_{2,X}\lambda_{3,X})^{i-1}$$

$$= \frac{1}{[1 - \lambda_{2,X}\lambda_{3,X}]^2} \tag{18.2.6a}$$

For the limiting case in which the dissociation coefficients are large relative to the other rates,

$$n_c \simeq \frac{1}{[1 - (m/k_{d^*,X})]^2}. \tag{18.2.6b}$$

The specific results noted above, which make various assumptions about the equality of some pairs of rates for R and W substrates and relative magnitudes of different classes of coefficients, lead to a maximum level of error reduction. More general formulae can be found in Hopfield (1974), Ninio (1975), and Murugan et al. (2012). By increasing the number of steps from substrate to final product, all other things being equal, proofreading will in general increase the reaction time, although it is unclear whether this is a limiting factor with respect to cell-division time, which depends on many other processes operating simultaneously. Moreover, it can be shown that an increase in proofreading rates (which in part influence the overall reaction time) can lead to an increase in both reaction speed and accuracy (by promoting correct substrates more rapidly to the final product before dissociation) (Banerjee et

al. 2017), so it is not inevitable that there is a tradeoff between speed and accuracy, contrary to common view (Johansson et al. 2008; Wohlgemuth et al. 2011). Rather, the directionality of this relationship depends on the full set of parameters in Figure 18.3, some of which are jointly favorable for both traits and others of which are not.

The Limits to Selection on Error Rates

Although an increase in phenotypic errors can have clear negative fitness consequences, natural selection has not driven error rates down to their biophysical limits. The substantial room that exists for improvement in translational fidelity is amply revealed by the fact that hyperaccurate ribosomes are readily obtained in microbial systems (Gorini 1971; Piepersberg et al. 1979; Bouadloun et al. 1983; Andersson et al. 1986; Zaher and Green 2010). Mikkola and Kurland (1992) found that natural isolates of *E. coli* have a ten-fold range of translation-error rates, bracketing the values found for long-established lab populations. Although they found no correlation between the growth rates and translation accuracies of different strains, this is perhaps not surprising given the difficulties of measuring growth rates to a high degree of resolution. Wild-type *E. coli* grow more rapidly than those with hyperaccurate ribosomes at high-nutrient levels, but growth rates are approximately the same under low-nutrient conditions that may more closely reflect the natural state (Andersson et al. 1986).

Why has natural selection not been able to reduce transcript and translation error rates to the levels observed for replication? One obvious difference here is that genomic errors remain associated with mutator alleles until separated by recombination, whereas transcription and translation errors are transient. In bacteria, the half lives of individual transcripts are typically 5 to 10 minutes, well below cell-division times (Bernstein et al. 2002; Hambraeus et al. 2003; Taniguchi et al. 2010; Dressaire et al. 2013). In *S. cerevisiae,* mRNA half lives are on the order of 10 to 20 minutes (Wang et al. 2002; Neymotin et al. 2014), and they can be a few hours in mouse cells (Schwanhäusser et al. 2011), again shorter than the time necessary for cell division.

It need not follow, however, that the damage from transcript errors is quickly erased. For one thing, proteins have considerably longer half lives than mRNAs. More importantly, the rate of molecular turnover need not have much of a bearing on the cellular load of errors at all. Rather, a roughly steady-state density of erroneous transcripts can be expected within the cellular environment, reflecting a balance between the decay of old mRNAs and the transcriptional production of new ones, and the same will be true of proteins. Thus, regardless of the transience of individual errors, one can expect a relatively constant number of total errors per cell (at both the transcript and translational levels). For each expressed gene, the total expected number of errors per cell at steady state will simply equal the product of the error rate per codon, the number of codons per mRNA, and the total number of mRNAs (for transcript errors) or protein molecules (for translation errors).

On the other hand, although the total number of errors in a cell must increase

with the number of molecules, the fitness effect of any single error in a particular protein may be diluted out with increasing numbers of copies of the protein free of the specific error. The setting for phenotypic mutations is fundamentally different than that for genetic mutations, which are either fully expressed (in haploids or homozygous diploids), or 50% expressed in diploid heterozygotes (assuming additivity). As a consequence, for a gene with a steady-state number n of transcripts or proteins, the phenotypic manifestation of error expression will be a function of the product of the number of transcripts (proteins) and a dilution factor, $\phi(n)$ (Foundations 13.2). In the case of additivity, the dilution factor is simply $1/n$, and the net effect of errors is independent of the number of molecules per cell.

To obtain the total burden of errors on fitness, we require the total number of erroneous amino acids within proteins, each discounted by the dilution effect. Let p be the number of expressed protein-coding genes, L be the mean number of codons per protein, and u be the number of errors per codon (which is the translation error rate in the case of proteins, and roughly twice the transcript-error rate for mRNAs owing to the redundancy of the genetic code). The expected number of erroneous amino acids incorporated into proteins by a particular route is then $up\overline{Ln}$. Further letting $\overline{\delta}$ be the average reduction in fitness if a mutation were to be fully expressed (as in an encoded genomic error in a haploid organism), the total reduction in fitness associated with a particular type of error is $up\overline{Ln}\phi(n)\cdot\overline{\delta}$. This further implies that the absolute value of the selection coefficient associated with a modifier of the transcript- or translational-error rate scales with $p\overline{Ln}\phi(n)\cdot\overline{\delta}$ or just $p\overline{L\delta}$ if the dilution factor is $1/n$.

As in the case of replication fidelity, theory also suggests that selection on transcript and translational fidelity should scale approximately inversely with the effective size of a population (N_e), as $1/N_e$ dictates the power of genetic drift (and hence the efficiency of selection) (Foundations 13.2). If this general theoretical framework is correct, we then expect u to scale inversely with $N_e \cdot p\overline{Ln}\phi(n)\cdot\overline{\delta}$, or more simply as $N_e \cdot p\overline{Ln}\phi(n)$ provided $\overline{\delta}$ is fairly constant across species. Because all of the underlying cellular factors can covary with each other, as well as with N_e, no simple scaling with single parameters is likely to emerge, and as can be seen in Table 18.1, there is no immediately obvious phylogenetic patterning associated with the transcript-error rate.

However, a composite analysis is consistent with the expected scaling. From Chapter 2, we know how the number of molecules n scales with cell volume, and also have information on the coefficient of variation of n, which enters $\phi(n)$ (Foundations 13.3). Estimates of p and $\overline{L}$ are generally available from genome sequencing data, and N_e from population-level polymorphism data (Chapter 8). For the five species for which data are available for these parameters along with measures of the transcript-error rate, the latter is seen to scale negatively with the composite parameter $N_e \cdot p\overline{Ln}\phi(n)$ (Figure 18.4). With this limited number of taxa, a simple model in which the dilution factor is $1/n$ cannot be rejected, but the slope of the overall regression is much shallower than the expectation of -1, either simply because of inadequate taxon sampling or because some aspect of the model is adequate.

The central point is that even with relentless selection to improve transcriptional and translation fidelity, error rates are expected to become stalled at high levels owing to their small phenotypic effects and the reduced efficiency of selection

operating at the drift barrier. This is not to say that an erroneous transcript or protein molecule cannot occasionally (and stochastically) yield a transient benefit in an extreme environment, but it does argue that such errors are little more than by-products of the limits to natural selection.

Finally, another suggested contribution to the evolution of high transcription and translation error rates is the cost of kinetic proofreading (Ehrenberg and Kurland 1984a). The idea here is that whereas increasingly accurate transcription and translation should improve cellular health, these advantages might become offset by the bioenergetic costs of proofreading, which generally consumes ATP. Under this view, either too high an error rate or too high a level of accuracy would lead to reduced fitness, motivating the idea that the fidelity of transcription / translation is kept at some intermediate optimal state by natural selection (Ehrenberg and Kurland 1984a,b; Kurland and Ehrenberg 1987). However, such optimization would only be possible if the supposed optimum error rate were higher than the drift barrier, else the latter would take precedence (Figure 18.5). Another implicit assumption in this argument is that increases in accuracy can only be achieved via proofreading improvement rather than through modifications of the basic efficiency of the pre-proofreading steps in transcription and translation. Finally, this cost argument does not explain the phylogenetic range in error rates. The evolutionary consequences of proofreading will be viewed in an entirely different light in the following section.

Foundations 18.2. The evolutionary bounds on the transcription-error rate. Given that natural selection relentlessly promotes the sequences of protein-coding genes toward their optimal state, it is reasonable to assume that most errors in transcripts are deleterious. Here, we consider the expected selective advantage of a genomic variant that improves transcriptional fidelity (or conversely the disadvantage of a variant that exacerbates the transcript-error rate). With slight modifications, the same approach can be used to ascertain the magnitude of selection operating on a variant that influences the translation-error rate. To achieve such an understanding, several factors must be considered: 1) the expected number of errors per molecule produced (transcript or protein) manifest at the amino-acid sequence level; 2) the total number of molecules associated with each gene; and 3) the fitness effects of such errors.

Letting u be the rate of amino-acid misincorporation per codon, and L_i be the number of amino acids in a protein of type i, the numbers (j) of erroneous amino acids in individual protein molecules of this type will be Poisson distributed with expectation uL_i,

$$P(j|uL_i) = \frac{e^{-uL_i}(uL_i)^j}{j!}. \tag{18.2.1}$$

From the standpoint of translation errors, u is defined to be the error-rate per codon, whereas if μ were to be the transcription-error rate per nucleotide site, because there are 3 nucleotide sites per codon, and $\sim$ 3/4s of nucleotide substitutions cause an amino-acid substitution, $u \simeq 3\mu \cdot (3/4) = 9\mu/4$.

Although all products of an error-containing gene will themselves contain the error, transcription and translation errors are singular events, and individual variant proteins will generally be just a fraction of the total pool of molecules for specific genetic loci. This raises the question of the degree to which the fitness effects of

single molecules are manifest at the cellular level. As with variant alleles at a locus, transcription and translation errors might behave in an additive, recessive, or dominant fashion, with the magnitude of the latter two conditions depending on the number of molecules per cell. Letting n_i be the number of molecules per cell for protein i, a flexible function that allows for alternative modes of dilution of effects is

$$f(n_i) = \frac{1}{n_i^x}, \tag{18.2.2}$$

which equals 1.0 when $n_i = 1$ (effects are fully felt), and converges to 0.0 (effects are completely masked) as $n_i \to \infty$ at a rate that depends on the exponent x. When $x = 1$, $f(n_i) = 1/n_i$, and the number of copies of a protein has no effect, as the number of error-containing proteins, which is proportional to n_i, is compensated by the dilution effect, i.e., $n_i \cdot f(n_i) = 1$. Values of $0 < x < 1$ result in a relatively slow decline in the dilution effect with increasing n_i (with $x = 0$ implying complete dominance of errors), whereas $x > 1$ results in a relatively rapid decline in $f(n_i)$ (i.e., relatively recessive effects of errors).

From Chapter 8, the average reduction in fitness associated with fully expressed deleterious mutations is generally < 0.1, and based on known transcription- and translation-error rates and typical gene lengths, the number of errors per protein will generally be $\ll 10$. Thus, letting δ be the fitness loss per single error in a single molecule if fully revealed, unless there are very strong epistatic effects, the total reduction in fitness in a protein containing j errors can be closely approximated as $j\delta \simeq 1 - e^{-j\delta}$.

For each locus i, the expected fractional reduction in fitness associated with the error burden (s_i) will then depend on the number of proteins per cell over which the errors are distributed, n_i, the degree of expression of individual errors, $f(n_i)$, and the distribution of the numbers of errors per protein $P(j|uL_i)$,

$$s_i = n_i \cdot f(n_i) \cdot \sum_{j=1}^{L_i} P(j|uL_i) \cdot (1 - e^{-j\delta}), \tag{18.2.3a}$$

$$\simeq n_i \cdot f(n_i) \cdot \left[1 - \exp\left(-\frac{uL_i\delta}{1+\delta}\right)\right]. \tag{18.2.3b}$$

Further simplification is possible by assuming independent n_i and L_i, ignoring variation in protein length, and noting that δ and $uL_i\delta \ll 1$,

$$s_i \simeq n_i \cdot f(n_i) \cdot (up\overline{L}) \cdot \delta, \tag{18.2.3c}$$

where p denotes the number of protein-coding loci in the proteome.

Equation 18.2.3c shows that the fitness consequences of mistranslation are a simple function of three quantities: 1) the total mistranslation rate per expressed proteome ($up\overline{L}$); 2) the net effect of the steady-state numbers of proteins per gene (the middle term, which is the product of copy number and dilution effect); and 3) the average effect of an amino-acid substitution if fully expressed ($\overline{\delta}$).

Finally, we require an expression to map the entire burden of errors throughout the proteome to cell fitness. As the average fitness effects of single errors in individual proteins will generally be $\ll 0.1$, we will assume the effects of errors arising at each genetic locus to act independently, such that total mean fitness can be denoted as

$$W(u) = \prod_{i=1}^{p}(1 - s_i) \simeq \exp\left(-\sum_{i=1}^{p} s_i\right), \tag{18.2.4}$$

where the approximation follows from $s_i \ll 1$. The quantity $[1 - W(u)]$ defines the fractional selective disadvantage of error rate u relative to the situation in which $u = 0$. The difference in this quantity for two alleles resulting in error rates u_1 and u_2, $[W(u_1) - W(u_2)]$, provides a measure of the selective advantage of the first allele over the second.

As noted in Chapter 8, the magnitude of a selection coefficient dictates the capacity for natural selection to improve a trait. For a haploid population with genetic effective size N_e, as a first-order approximation, the absolute value of $[W(u_1) - W(u_2)]$ must exceed $1/N_e$ for natural selection to discriminate between alternative alleles. For $|W(u_1) - W(u_2)| < 1/N_e$, drift overwhelms the power of selection, and hence this point is referred to as the drift barrier. Thus, letting Δ denote the fractional decline in the translation-error rate between consecutive states in a hierarchy of mutationally connected alleles with effects on the error rate, the drift barrier is the error rate u^* that satisfies

$$N_e = \frac{1}{W[(1-\Delta)u^*] - W[u^*]}. \tag{18.2.5}$$

For $u > u^*$, natural selection is capable of driving the error rate to a lower value, whereas a situation in which $u < u^*$ is expected to be transient as selection is unable to maintain such a state.

Rearranging, and substituting from above, the lower-bound to the error rate achievable by selection is

$$u^* \simeq \frac{1}{\Delta \cdot N_e \cdot p\overline{L} \cdot \overline{n}\phi(\overline{n}) \cdot \delta}. \tag{18.2.6}$$

where $\phi(n) = \{1 - [x(1-x)C_n^2/2]\}/\overline{n}^x$ is the average dilution factor, with C_n denoting the mean and coefficient of variation in expression level (obtained by Taylor-series expansion of Equation 18.2.2). This shows that, all other things being equal, there is an expected inverse relationship between N_e and the drift barrier. However, there is also an inverse scaling with the total number of codons in the proteome $p\overline{L}$, the copy-number effect $n\phi(n)$, and the average fitness effect of mutations δ. The granularity of mutational changes in alleles influencing the error rate (Δ) operates as a simple scaling factor, but does not change the form of the relationship – the higher the value of Δ, the greater the difference of allelic effects, and hence the greater the efficiency of selection for a lower error rate.

The Evolutionary Consequences of Proofreading

As noted above, the accuracy-demanding processes of replication, transcription, and translation all involve layers of mechanisms that improve fidelity. For example, genome replication involves highly selective DNA polymerases, with the small fraction of resultant base misincorporations being subject to correction by subsequent proofreading, and the still smaller fraction of errors that escape proofreading being subject to mismatch repair.

The subset of erroneous transcripts either missing the translation start or termination site or containing premature termination codons can be removed by mRNA surveillance mechanisms that occur after the initiation of translation. These include: the nonsense-mediated decay (NMD) pathway, which eliminates a fraction

of inappropriate mRNAs carrying premature termination codons (PTCs); no-go decay, which degrades mRNAs associated with stalled ribosomes; and non-stop decay, which removes mRNAs lacking a stop codon (Graille and Séraphin 2012; Kervestin and Jacobson 2012).

The central point is that some fraction of erroneous mRNAs is removed from the cell in the very first round of translation, and this may help explain why *in vivo* transcript error-rate estimates are lower than those obtained by *in vitro* methods, which exclude translation-associated processes. Notably, many of these surveillance mechanisms are present only in eukaryotes, or are substantially elaborated in the eukaryotic lineage. Eukaryotes also have a quality-control pathways for removing nonfunctional ribosomal RNAs and transfer RNAs, which are apparently absent from prokaryotes (LaRiviere et al. 2006; Kramer and Hopper 2013).

The eukaryotic NMD process is associated with spliceosomal introns, which are unique to eukaryotes (often exceeding an average of five per protein-coding gene) and must be spliced out of precursor mRNAs to produce productive transcripts prior to translation. Failure to remove an intron, which may be the most common mode of production of erroneous transcripts, will lead to a downstream frameshift two-thirds of the time, which will in turn usually lead to the appearance of PTCs. If not removed from the cytoplasm, such aberrant transcripts will yield truncated proteins, which will generally be harmful to cell health. In general, the NMD process removes such transcripts during their first round of translation, distinguishing erroneous from true termination codons by use of information on the distance from the length of the 3′ end of the transcript, including information laid down in the form of proteins marking intron-exon junctions (Hentze and Kulozik 1999; Gonzalez et al. 2001; Lykke-Andersen 2001; Mango 2001; Maquat and Carmichael 2001; Wilusz et al. 2001; Maquat 2004, 2006).

Not all PTC-containing mRNAs are detectable by the NMD process, but the importance of NMD is illustrated by experiments in which the pathway has been silenced, which show substantial increases in PTC-containing mRNAs, with greater efficiency of removal of transcripts with PTCs near the 5′ end (Mendell et al. 2004; Mitrovich and Anderson 2005; Gout et al. 2017). Knockouts of the NMD pathway have small phenotypic effects in the yeasts *S. cerevisiae* and *S. pombe* (Leeds et al. 1992; Dahlseid et al. 1998; Mendell et al. 2000), moderate fitness effects in the nematode *C. elegans* (Hodgkin et al. 1989), and lethal effects in mice (Medghalchi et al. 2001). The enhanced sensitivity of multicellular species to NMD inactivation may simply be a consequence of greater rates of production of erroneous transcripts in complex genomes with more opportunities for splicing errors, although this conclusion is clouded by evidence that some of the proteins in the NMD pathway have additional cellular functions (Maquat 2006).

The accuracy of translation also depends on a series of surveillance mechanisms for proper loading of tRNA synthetases by their cognate amino acids, proper recognition of tRNAs by their cognate synthetases, and proper codon recognition by tRNAs, with all three steps incorporating proofreading mechanisms. It is common to view such layered security systems as reflections of the extraordinarily creative power of mutation combined with natural selection – once selection has brought a particular mechanism to perfection, a second layer can emerge, yielding a quantum leap in accuracy, with still other layers being subsequently added. Under this view,

copying fidelity can be continuously pushed to higher and higher levels. Left unexplained, however, is the fact that eukaryotes, which have more elaborate surveillance mechanisms for errors in DNA, RNAs, and proteins, nonetheless exhibit high net error rates in all cases.

This view of ever-improving fidelity is likely incorrect. Suppose that prior to the addition of a secondary line of defense, the primary mechanism is not constrained by biophysical limits, but rather by the drift barrier. In that case, the fortuitous addition of a second layer of defense (with a large enough immediate effect) might lead to a larger boost in accuracy than possible under incremental changes made to the primary mechanism, thereby vaulting over the prior limits to natural selection. However, the initial boost in accuracy need not be permanent, as incremental reductions in the efficiency of both layers are likely to result in decay back to the drift barrier, rendering the overall system no more accurate than the previous single-layered system (Figure 18.6). The end result is a more complex and bioenergetically expensive system, which superficially looks robust, but is in fact no more accurate than the simpler ancestral state (Gros and Tenaillon 2009; Lynch 2012). Frank (2007) called this phenomenon the "paradox of robustness."

In effect, the combination of multiple lines of defense results in the relaxation of selection on the efficiency of individual layers, and hence the eventual degeneration of earlier established mechanisms. This is because natural selection operates on the output of an entire system, leaving multiple degrees of freedom for change in individual components so long as the summed level of efficiency remains at the drift barrier. With two layers, the bivariate drift barrier will have a line of equivalence with pairs of phenotypes on the line being equivalent with respect to overall accuracy (Figure 18.6), and with three layers, there will be a trivariate drift barrier. Such systems will be vulnerable to eventual loss of one component, provided such loss can be compensated by improved performance in the other(s).

Observations on the mutational properties of microbes support for this view. For example, although the elimination of mismatch-repair from *E. coli, S. cerevisiae* and other organisms generally results in an $\sim$ 100-fold increase in the mutation rate, *Mycobacterium smegmatis,* a bacterium lacking the mismatch-repair pathway, and *Deinococcus radiodurans,* a bacterium in which mismatch repair only improves accuracy by a few fold, both have mutation rates similar to that in other microbes (Long et al. 2015; Kucukyildirim et al. 2016). The fact that bacterial populations founded with a mutator genotype frequently evolve lower mutation rates on relatively short timescales through compensatory molecular changes at genomic sites not involved in the initial mutator construct provides further evidence that individual fidelity mechanisms are not limited by biophysical constraints (McDonald et al. 2012; Turrientes et al. 2013; Wielgoss et al. 2013; Williams et al. 2013). Finally, different yeast species with very similar mutation rates have substantial differences in their mutation spectra, implying differences in the underlying mechanisms of mutation and repair (Long et al. 2016).

From the standpoint of translational fidelity, one can also point to cases involving the fidelity mechanisms involving AARSs, the enzymes responsible for sequestering cognate amino acids prior to passing them on to their appropriate tRNAs. Many AARSs are capable of post-transfer editing of mischarged tRNAs. However, the PheRS in *Mycoplasma mobile* has lost the capacity to edit, and instead sim-

ply relies on pre-transfer kinetic proofreading for discriminating against noncognate amino acids (Yadavalli and Ibba 2013). Although this AARS is not sufficient to support *E. coli* growth, presumably owing to problems associated with mistranslation, the introduction of just two amino-acid changes into the evolutionarily deactivated editing domain removes this deficiency, increasing the level of accuracy by several fold. Yeast (*S. cerevisiae*) cytoplasmic PheRS is capable of editing a tRNA mischarged with an erroneously transferred tyrosine, whereas the mitochondrial PheRS is incapable of editing but nonetheless has a comparable error rate (Reynolds et al. 2010). Thus, the accuracy of the latter is solely dependent on a high level of initial specificity, which has apparently been displaced by post-transfer editing in the cytoplasmic version. Many other examples are known in which tRNA-charging accuracy depends on a mixture of pre- and post-transfer, with a switch to strong reliance on just a single mechanism being conferred by no more than two amino-acid substitutions (Martinis and Boniecki 2010).

Adaptive Significance of Errors

Given that the vast majority of amino-acid altering mutations have negative fitness effects (Chapter 10), it is reasonable to expect the same to be true of translation errors. Indeed, high translation-error rates are known to lead to malfunctioning cells (Lee et al. 2006; Nangle et al. 2006; Bacher and Schimmel 2007; Schimmel and Guo 2009), and as noted above, the removal of the surveillance capacity for aberrant mRNAs also causes fitness loss. By evaluating the growth rates of *Salmonella* cells containing various mutations influencing translation fidelity, Hughes (1991; Hughes and Tubulekas 1993) found that a ten-fold increase in the error rate generates a two-fold reduction in growth rate. This growth-rate reduction was not a consequence of reduced processivity of translation, but was shown by an enzyme reporter assay to be associated with the production of erroneous proteins – a 3× increase in the translation-error rate yielded a 40% reduction in enzyme activity and a 15% reduction in protein stability, with no associated change in protein abundance. Similarly, in comparisons of the catalytic activities of two enzymes in wild-type *E. coli* with those in a mutant line with enhanced translational fidelity, Musa et al. (2016) found in both cases an $\sim 30\%$ increase in enzyme activity in the latter case.

Despite these observations, following the grand tradition of assuming that everything biological must be optimized to maximize organismal fitness, a number of authors have argued that translational errors have been wrongly viewed as deleterious (Peltz et al. 1999; Santos et al. 1999; Pezo et al. 2004; Bacher et al. 2007; Pan 2013; Ribas de Pouplana et al. 2014; Fan et al. 2015; Wang and Pan 2016). This contrarian argument asserts that mistranslation is a regulated phenomenon, with organisms "deliberately" making errors in order to expand the chemical diversity of the cell. The view here is that there is an optimal intermediate level of mistranslation, fine-tuned by natural selection to yield populations of variant molecules, some of which will have fitness-enhancing functions. What drives this point of view?

First, if one engineers an *E. coli* cell line to be auxotrophic (unable to synthesize a particular nutrient) by introducing a missense mutation in a gene required for the synthesis of the nutrient, an editing defective tRNA synthetase can rescue

the line, presumably by promoting the production of a small fraction of proteins with translation-error produced reversions to phenotypic function (Min et al. 2003). However, such an extreme starting point provides little (if any) evidence for the adaptive significance of error production, as auxotrophic mutants are expected to be rapidly purged from populations by natural selection except in cases where the nutrient is already freely available in the environment.

Second, it has commonly been observed that cells under stress have increased translation-error rates. For example, in bacteria stressed with antibiotics and/or oxidative damage, error rates can increase by 10 to 100× (Bacher and Schimmel 2007; Kramer and Farabaugh 2007; Javid et al. 2014; Fan et al. 2015; Leng et al. 2015). Methionine misacylation (i.e., the mischarging of non-methionine tRNAs with methionine) is particularly common in stressful situations, with up to 10% of incorporated methionine being erroneous in some conditions for mammalian, yeast, and bacterial cells (Jones et al. 2011; Wiltrout et al. 2012; Schwartz and Pan 2017). In the hyperthermophilic archeon *Aeropyrum pernix,* growth at low temperatures is accompanied by global mischarging of leucine tRNAs by methionine AARS. It has been argued that such conservation of a particular type of "misfunction" must imply maintenance by natural selection as a mechanism for the adaptive exploitation of errors (Netzer et al. 2009; Pan 2013; Schwartz et al. 2016).

In principle, widespread methionine misincorporation might yield advantages beyond the direct effects on protein functions. There is, for example, biochemical evidence that methionine residues can serve as scavengers for reactive oxygen species via conversion to methionine sulfoxide (Levine et al. 1999; Stadtman and Levine 2003; Wang and Pan 2016), which would in turn protect the proteins containing them. Moreover, a common enzyme (methionine sulfoxide reductase) confers the ability to revert the modified base back to methionine, implying that such residues can be recycled as antioxidants.

Nonetheless, these observations leave many questions unanswered. Although methionine misincorporation can alter the properties of individual proteins in potentially beneficial ways (Schwartz and Pan 2016; Wang and Pan 2015), this is not a demonstration of an overall induced selective advantage, as the majority of such variants at other loci are likely deleterious. If methionine serves such a useful function, then why is the cytoplasm not populated with higher free amounts of this amino acid, and why are more methionines not directly encoded into the proteome? There are no known disadvantages of methionine in nonstressful conditions.

Third, when cells are extremely starved for one particular amino acid, mischarging of the cognate tRNA synthetase can increase misincorporation rates by up to an order of magnitude (Feeney et al. 2013), in some cases providing rescue from an otherwise lethal situation. In *Acinetobacter baylyi,* for example, a mutation that allows the isoleucine-AARS to mischarge with valine can increase the growth rate when isoleucine is strongly limiting (Bacher et al. 2007). Should this be surprising, however, given the alternative outcome of no translation? In nature, a more common situation would likely be generic shortage of all amino acids, which under this view would call for global mischarging of all AARSs, an intracellular free-for-all. Moreover, as discussed above (e.g., Equation 18.1a), utilization of inappropriate substrates is expected to naturally increase whenever preferred substrates are relatively rare.

Fourth, some organisms, such as members of the genus *Mycoplasma,* have one or more AARSs with editing defective domains (Li et al. 2011). The fungal pathogen *Candida albicans* has experienced a reassignment of one particular leucine codon (CUG) to serine (an alteration of the genetic code), but still incorporates 1 to 6% leucines at such codons (Rocha et al. 2011; Bezerra et al. 2013). This codon ambiguity reduces fitness in normal environments, and induces the expression of stress-response proteins, which can create a competitive advantage in stressful environments (Santos et al. 1999). Again, however, although rare accidents can occasionally be useful, that does not mean that the proclivity to incur accidents is promoted by natural selection. Notably, the organisms with intrinsically defective translation efficiency are pathogens, which may be highly vulnerable to random genetic drift and loss of nonessential functions.

In summary, although there is clear evidence that translation inaccuracy can increase during times of stress, on occasion even stochastically creating protein variants capable of improving a precarious situation, there is a lack of direct evidence that error-prone translation has been promoted by selection as a strategy for adaptively generating variant protein pools. When cells are stressed, cellular functions go wrong, and there is no obvious reason why translation should be an exception. Notably, the examples promoted as poster children of adaptive translation inaccuracy are highly idiosyncratic in that they involve different amino acids – leucine and serine in the case of *Candida,* phenylalanine and leucine in the case of *Mycoplasma,* asparagine and aspartate in the case of *Mycobacteria,* and methionine in the case of *E. coli,* yeast, and mammals. There is no obvious reason why such lineage-specific variation would be driven by specific phylogenetic adaptive requirements.

Finally, in one experiment imposing a selective challenge to evolve antibiotic resistance, the rate of improvement was not enhanced by elevated translation error, although there was some improvement in the evolved strains when later exposed to some antibiotics that were not the source of the actual selection (Bratulic et al. 2017). In a rather novel twist, these authors argue that mistranslation improves long-term fitness not by providing stochastic short-term solutions to selective challenges but by imposing epistatic effects that somehow magnify the effects of preexisting deleterious mutations, hence enhancing the ability of natural selection to purge such mutations from the population. Whether this is generally true remains to be seen. But either way, it should again be realized that the existence of such an effect does not provide evidence for the origin and maintenance of translational inaccuracy by natural selection.

Literature Cited

Andersson, D. I., H. W. van Verseveld, A. H. Stouthamer, and C. G. Kurland. 1986. Suboptimal growth with hyper-accurate ribosomes. Arch. Microbiol. 144: 96-101.

Ardehali, M. B., and J. T. Lis. 2009. Tracking rates of transcription and splicing *in vivo*. Nat. Struct. Mol. Biol. 16: 1123-1124.

Bacher, J. M., and P. Schimmel. 2007. An editing-defective aminoacyl-tRNA synthetase is mutagenic in aging bacteria via the SOS response. Proc. Natl. Acad. Sci. USA 104: 1907-1912.

Banerjee, K., A. B. Kolomeisky, and O. A. Igoshin. 2017. Elucidating interplay of speed and accuracy in biological error correction. Proc. Natl. Acad. Sci. USA 114: 5183-5188.

Bernstein, J. A., A. B. Khodursky, P. H. Lin, S. Lin-Chao, S. N. Cohen. 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. Proc. Natl. Acad. Sci. USA 99: 9697-9702.

Bezerra, A. R., et al. 2013. Reversion of a fungal genetic code alteration links proteome instability with genomic and phenotypic diversification. Proc. Natl. Acad. Sci. USA 110: 11079-11084.

Blank, A., J. A. Gallant, R. R. Burgess, and L. A. Loeb. 1986. An RNA polymerase mutant with reduced accuracy of chain elongation. Biochemistry 25: 5920-5928.

Bouadloun, F., D. Donner, and C. G. Kurland. 1983. Codon-specific missense errors *in vivo*. EMBO J. 2: 1351-1356.

Bratulic, S., M. Toll-Riera, and A. Wagner. 2017. Mistranslation can enhance fitness through purging of deleterious mutations. Nat. Commun. 8: 15410.

Brégeon, D., Z. A. Doddridge, H. J. You, B. Weiss, and P. W. Doetsch. 2003. Transcriptional mutagenesis induced by uracil and 8-oxoguanine in *Escherichia coli*. Mol. Cell 12: 959-970.

Carter, R., and G. Drouin. 2010. The increase in the number of subunits in eukaryotic RNA polymerase III relative to RNA polymerase II is due to the permanent recruitment of general transcription factors. Mol. Biol. Evol. 27: 1035-1043.

Clauson, C. L., K. J. Oestreich, J. W. Austin, and P. W. Doetsch. 2010. Abasic sites and strand breaks in DNA cause transcriptional mutagenesis in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 107: 3657-3662.

Dahlseid, J. N., J. Puziss, R. L. Shirley, A. L. Atkin, P. Hieter, and M. R. Culbertson. 1998. Accumulation of mRNA coding for the ctf13p kinetochore subunit of *Saccharomyces cerevisiae* depends on the same factors that promote rapid decay of nonsense mRNAs. Genetics 150: 1019-1035.

Dressaire, C., F. Picard, E. Redon, P. Loubiére, I. Queinnec, L. Girbal, and M. Cocaign-Bousquet. 2013. Role of mRNA stability during bacterial adaptation. PLoS One 8: e59059.

de Mercoyrol, L., Y. Corda, C. Job, and D. Job. 1992, Accuracy of wheat-germ RNA polymerase II. General enzymatic properties and effect of template conformational transition from right-handed B-DNA to left-handed Z-DNA. Eur. J. Biochem. 206: 49-58.

Ehrenberg, M., and C. G. Kurland. 1984. Costs of accuracy determined by a maximal growth rate constraint. Q. Rev. Biophys. 17: 45-82.

Fan, Y., J. Wu, M. H. Ung, N. De Lay, C. Cheng, and J. Ling. 2015. Protein mistranslation

protects bacteria against oxidative stress. Nucleic Acids Res. 43: 1740-1748.

Feeney, L., V. Carvalhal, X. C. Yu, B. Chan, D. A. Michels, Y. J. Wang, A. Shen, J. Ressl, B. Dusel, and M. W. Laird. 2013. Eliminating tyrosine sequence variants in CHO cell lines producing recombinant monoclonal antibodies. Biotechnol. Bioeng. 110: 1087-1097.

Fersht, A. 1999. Structure and Mechanism in Protein Science. W. H. Freeman & Co., New York, NY.

Frank, S. A. 2007. Maladaptation and the paradox of robustness in evolution. PLoS ONE 2: e1021.

Frith, M. C., J. Ponjavic, D. Fredman, C. Kai, J. Kawai, P. Carninci, Y. Hayashizaki, and A. Sandelin. 2006. Evolutionary turnover of mammalian transcription start sites. Genome Res. 16: 713-722.

Golding, I., and E. C. Cox. 2004. RNA dynamics in live *Escherichia coli* cells. Proc. Natl. Acad. Sci. USA 101: 11310-11305.

Goldsmith, M., and D. S. Tawfik. 2009. Potential role of phenotypic mutations in the evolution of protein expression and stability. Proc. Natl. Acad. Sci. USA 106: 6197-6202.

Gonzalez, C. I., A. Bhattacharya, W. Wang, and S. W. Peltz. 2001. Nonsense-mediated mRNA decay in *Saccharomyces cerevisiae.* Gene 274: 15-25.

Gorini, L. 1971. Ribosomal discrimination of tRNAs. Nat. New Biol. 234: 261-264.

Gout, J.-F., W. Li, C. Fritsch, A. Li, S. Haroon, L. Singh, D. Hua, H. Fazelinia, S. Seeholzer, M. Lynch, and M. Vermulst. 2017. The landscape of transcription errors in eukaryotic cells. Science Advances (in press).

Gout, J. F., W. K. Thomas, Z. Smith, K. Okamoto, and M. Lynch. 2013. Large-scale detection of *in vivo* transcription errors. Proc. Natl. Acad. Sci. USA 110: 18584-18589.

Graille, M., and B. Séraphin. 2012. Surveillance pathways rescuing eukaryotic ribosomes lost in translation. Nat. Rev. Mol. Cell Biol. 13: 727-735.

Gros, P. A., and O. Tenaillon. 2009. Selection for chaperone-like mediated genetic robustness at low mutation rate: impact of drift, epistasis and complexity. Genetics 182: 555564.

Haag, J. R., and C. S. Pikaard. 2011. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nat. Rev. Mol. Cell. Biol. 12: 483-492.

Hambraeus, G., C. von Wachenfeldt, and L. Hederstedt. 2003. Genome-wide survey of mRNA half-lives in *Bacillus subtilis* identifies extremely stable mRNAs. Mol. Genet. Genomics 269: 706-714.

Hentze, M. W., and A. E. Kulozik. 1999. A perfect message: RNA surveillance and nonsense-mediated decay. Cell 96: 307-310.

Hodgkin, J., A. Papp, R. Pulak, V. Ambros, and P. Anderson. 1989. A new kind of informational suppression in the nematode *Caenorhabditis elegans.* Genetics 123: 301-313.

Hopfield, J. J. 1974. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. Proc. Natl. Acad. Sci. USA 71: 4135-4139.

Hopfield, J. J., T. Yamane, V. Yue, and S. M. Coutts. 1976. Direct experimental evidence for kinetic proofreading in amino acylation of tRNA-Ile. Proc. Natl. Acad. Sci. USA 73: 1164-1168.

Hiriyanna, K. T., and T. Ramakrishnan. 1986. Deoxyribonucleic acid replication time in Mycobacterium tuberculosis H37 Rv. Arch. Microbiol. 144: 105-109.

Hughes, D. 1991. Error-prone EF-Tu reduces *in vivo* enzyme activity and cellular growth rate. Mol. Microbiol. 5: 623-630.

Hughes, D., and I. Tubulekas. 1993. Ternary complex-ribosome interaction: its influence on protein synthesis and on growth rate. Biochem. Soc. Trans. 21: 851-857.

Hussain, T., V. Kamarthapu, S. P. Kruparani, M. V. Deshmukh, and R. Sankaranarayanan. 2010. Mechanistic insights into cognate substrate discrimination during proofreading in translation. Proc. Natl. Acad. Sci. USA 107: 22117-22121.

Ieong, K. W., U. Uzun, M. Selmer, and M. Ehrenberg. 2016. Two proofreading steps amplify the accuracy of genetic code translation. Proc. Natl. Acad. Sci. USA 113: 13744-13749.

Javid, B., F. Sorrentino, M. Toosky, W. Zheng, J. T. Pinkham, N. Jain, M. Pan, P. Deighan, and E. J. Rubin. 2014. Mycobacterial mistranslation is necessary and sufficient for rifampicin phenotypic resistance. Proc. Natl. Acad. Sci. USA 111: 1132-1137.

Johansson, M., M. Lovmar, and M. Ehrenberg. 2008. Rate and accuracy of bacterial protein synthesis revisited. Curr. Opin. Microbiol. 11: 141-147.

Jones, T. E., R. W. Alexander, and T. Pan. 2011. Misacylation of specific nonmethionyl tRNAs by a bacterial methionyl-tRNA synthetase. Proc. Natl. Acad. Sci. USA 108: 6933-6938.

Kaplan, C. D. 2010. The architecture of RNA polymerase fidelity. BMC Biol. 8: 85.

Kervestin, S., and A. Jacobson. 2012. NMD: a multifaceted response to premature translational termination. Nat. Rev. Mol. Cell Biol. 13: 700-712.

Kireeva, M. L., Y. A. Nedialkov, G. H. Cremona, Y. A. Purtov, L. Lubkowska, F. Malagon, Z. F. Burton, J. N. Strathern, and M. Kashlev. 2008. Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. Mol. Cell 30: 557-566.

Kramer, E. B., and P. J. Farabaugh. 2007. The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. RNA 13: 87-96.

Kramer, E. B., and A. K. Hopper. 2013. Retrograde transfer RNA nuclear import provides a new level of tRNA quality control in *Saccharomyces cerevisiae.* Proc. Natl. Acad. Sci. USA 110: 21042-21047.

Kucukyildirim, S., H. Long, W. Sung, S. F. Miller, T. G. Doak, and M. Lynch. 2016. The rate and spectrum of spontaneous mutations in *Mycobacterium smegmatis,* a bacterium naturally devoid of the post-replicative mismatch repair pathway. G3: Genes—Genomes—Genetics 6: 2157-2163.

Kuntz, I. D., K. Chen, K. A. Sharp, and P. A. Kollman. 1999. The maximal affinity of ligands. Proc. Natl. Acad. Sci. USA 96: 9997-10002.

Kurland, C. G., and M. Ehrenberg. 1984. Optimization of translation accuracy. Prog. Nucleic Acid Res. Mol. Biol. 31: 191-219.

Kurland, C. G., and M. Ehrenberg. 1987. Growth-optimizing accuracy of gene expression. Annu. Rev. Biophys. Biophys. Chem. 16: 291-317.

Landick, R. 2009. Functional divergence in the growing family of RNA polymerases. Structure 17:

323-325.

LaRiviere, F. J., S. E. Cole, D. J. Ferullo, and M. J. Moore. 2006. A late-acting quality control process for mature eukaryotic rRNAs. Mol. Cell 24: 619-626.

Larson, M. H., J. Zhou, C. D. Kaplan, M. Palangat, R. D. Kornberg, R. Landick, and S. M. Block. 2012. Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. Proc. Natl. Acad. Sci. USA 109: 6555-6560.

Lee, H., E. Popodi, H. Tang, and P. L. Foster. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. Proc. Natl. Acad. Sci. USA 109: E2774-E2783.

Lee, J. W., K. Beebe, L. A. Nangle, J. Jang, C. M. Longo-Guess, S. A. Cook, M. T. Davisson, J. P. Sundberg, P. Schimmel, and S. L. Ackerman. 2006. Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. Nature 443: 50-55.

Leeds, P., J. M. Wood, B. S. Lee, and M. R. Culbertson. 1992. Gene products that promote mRNA turnover in *Saccharomyces cerevisiae.* Mol. Cell. Biol. 12: 2165-2177.

Leng, T., M. Pan, X. Xu, and B. Javid. 2015. Translational misreading in *Mycobacterium smegmatis* increases in stationary phase. Tuberculosis 95: 678-681.

Levine, R. L., B. S. Berlett, J. Moskovitz, L. Mosoni, and E. R. Stadtman. 1999. Methionine residues may protect proteins from critical oxidative damage. Mech. Ageing Dev. 107: 323-332.

Li, L., M. T. Boniecki, J. D. Jaffe, B. S. Imai, P. M. Yau, Z. A. Luthey-Schulten, and S. A. Martinis. 2011. Naturally occurring aminoacyl-tRNA synthetases editing-domain mutations that cause mistranslation in Mycoplasma parasites. Proc. Natl. Acad. Sci. USA 108: 9378-9383.

Lisica, A., C. Engel, M. Jahnel, É. Roldán, E. A. Galburt, P. Cramer, and S. W. Grill. 2016. Mechanisms of backtrack recovery by RNA polymerases I and II. Proc. Natl. Acad. Sci. USA 113: 2946-2951.

Long, H., M. G. Behringer, E. Williams, R. Te, and M. Lynch. 2016. Similar mutation rates but highly diverse mutation spectra in ascomycete and basidiomycete yeasts. Genome Biol. Evol. 8: 3815-3821.

Long, H., S. Kucukyildirim, W. Sung, E. Williams, M. Ackerman, T. G. Doak, and M. Lynch. 2015. Background mutational features of the radiation-resistant bacterium *Deinococcus radiodurans.* Mol. Biol. Evol. 32: 2383-2392.

Lykke-Andersen, J. 2001. mRNA quality control: marking the message for life or death. Curr. Biol. 11: R88-R91.

Lynch, M. 2007. The Origins of Genome Architecture. Sinauer Assocs., Inc., Sunderland, MA.

Lynch, M. 2012. Evolutionary layering and the limits to cellular perfection. Proc. Natl. Acad. Sci. USA 109: 18851-18856.

Mango, S. E. 2001. Stop making nonSense: the *C. elegans* smg genes. Trends Genet. 17: 646-653.

Maquat, L. E. 2004. Nonsense-mediated mRNA decay: a comparative analysis of different species. Curr. Genomics 5: 175-190.

Maquat, L. E. (ed.) 2006. Nonsense-mediated mRNA Decay. Landes Bioscience, Georgetown, TX.

Maquat, L. E., and G. G. Carmichael. 2001. Quality control of mRNA function. Cell 104: 173-176.

Martinis, S. A., and M. T. Boniecki. 2010. The balance between pre- and post-transfer editing in tRNA synthetases. FEBS Lett. 584: 455-459.

Mason, P. B., and K. Struhl. 2005. Distinction and relationship between elongation rate and processivity of RNA polymerase II *in vivo.* Mol. Cell 17: 831-840.

McCandlish, D. M., and J. B. Plotkin. 2016. Transcriptional errors and the drift barrier Proc. Natl. Acad. Sci. USA 113: 31363138.

McDonald, M. J., Y. Y. Hsieh, Y. H. Yu, S. L. Chang, and J. Y. Leu. 2012. The evolution of low mutation rates in experimental mutator populations of *Saccharomyces cerevisiae.* Curr. Biol. 22: 1235-1240.

Medghalchi, S. M., P. A. Frischmeyer, J. T. Mendell, A. G. Kelly, A. M. Lawler, and H. C. Dietz. 2001. Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. Hum. Mol. Genet. 10: 99-105.

Mendell, J. T., S. M. Medghalchi, R. G. Lake, E. N. Noensie, and H. C. Dietz. 2000. Novel Upf2p orthologues suggest a functional link between translation initiation and nonsense surveillance complexes. Mol. Cell. Biol. 20: 8944-8957.

Mendell, J. T., N. A. Sharifi, J. L. Meyers, F. Martinez-Murillo, and H. C. Dietz. 2004. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nature Genet. 36: 1073-1078.

Mikkola, R., and C. G. Kurland. 1992. Selection of laboratory wild-type phenotype from natural isolates of *Escherichia coli* in chemostats. Mol. Biol. Evol. 9: 394-402.

Min, B., M. Kitabatake, C. Polycarpo, J. Pelaschier, G. Raczniak, B. Ruan, H. Kobayashi, S. Namgoong, and D. Söll. 2003. Protein synthesis in *Escherichia coli* with mischarged tRNA. J. Bacteriol. 185: 3524-3526.

Mitrovich, Q. M., and P. Anderson. 2000. Unproductively spliced ribosomal protein mRNAs are natural targets of mRNA surveillance in *C. elegans.* Genes Dev. 14: 2173-2184.

Murugan, A., D. A. Huse, and S. Leibler. 2012. Speed, dissipation, and error in kinetic proofreading. Proc. Natl. Acad. Sci. USA 109: 12034-12039.

Musa, M., M. Radman, and A. Krisko. 2016. Decreasing translation error rate in *Escherichia coli* increases protein function. BMC Biotechnol. 16: 28.

Muzyczka, N., R. L. Poland, and M. J. Bessman. 1972. Studies on the biochemical basis of spontaneous mutation. I. A comparison of the deoxyribonucleic acid polymerases of mutator, antimutator, and wild type strains of bacteriophage T4. J. Biol. Chem. 247: 7116-7122.

Myllykallio, H., P. Lopez, P. López-García, R. Heilig, W. Saurin, Y. Zivanovic, H. Philippe, and P. Forterre. 2000. Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. Science 288: 2212-2215.

Nangle, L. A., C. M. Motta, and P. Schimmel. 2006. Global effects of mistranslation from an editing defect in mammalian cells. Chem. Biol. 13: 1091-1100.

Netzer, N., et al. 2009. Innate immune and chemically triggered oxidative stress modifies translational fidelity. Nature 462: 522-526.

Neymotin, B., R. Athanasiadou, and D. Gresham. 2014. Determination of *in vivo* RNA kinetics using RATE-seq. RNA 20: 1645-1652.

Ninio, J. 1975. Kinetic amplification of enzyme discrimination. Biochimie 57: 587-595.

Ninio, J. 1991a. Connections between translation, transcription and replication error-rates. Biochimie 73: 1517-1523.

Ninio, J. 1991b. Transient mutators: a semiquantitative analysis of the influence of translation and transcription errors on mutation rates. Genetics 129: 957-962.

Pan, T. 2013. Adaptive translation as a mechanism of stress response and adaptation. Annu. Rev. Genet. 47: 121-137.

Parker, J. 1989. Errors and alternatives in reading the universal genetic code. Microbiol. Rev. 53: 273-298.

Peltz, S. W., A. B. Hammell, Y. Cui, J. Yasenchak, L. Puljanowski, and J. D. Dinman. 1999. Ribosomal protein L3 mutants alter translational fidelity and promote rapid loss of the yeast killer virus. Mol. Cell Biol. 19: 384-391.

Pezo, V., D. Metzgar, T. L. Hendrickson, W. F. Waas, S. Hazebrouck, V. Döring, P. Marlière, P. Schimmel, and V. De Crècy-Lagard. 2004. Artificially ambiguous genetic code confers growth yield advantage. Proc. Natl. Acad. Sci. USA 101: 8593-8597.

Piepersberg, W., V. Noseda, and A. Böck. 1979. Bacterial ribosomes with two ambiguity mutations: effects of translational fidelity, on the response to aminoglycosides and on the rate of protein synthesis. Mol. Gen. Genet. 171: 23-34.

Proshkin, S., A. R. Rahmouni, A. Mironov, and E. Nudler. 2010. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. Science 328: 504-508.

Reynolds, N. M., J. Ling, H. Roy, R. Banerjee, S. E. Repasky, P. Hamel, and M. Ibba. 2010. Cell-specific differences in the requirements for translation quality control. Proc. Natl. Acad. Sci. USA 107: 4063-4068.

Ribas de Pouplana, L., M. A. Santos, J. H. Zhu, P. J. Farabaugh, and B. Javid. 2014. Protein mistranslation: friend or foe? Trends Biochem. Sci. 39: 355-362.

Rocha, R., P. J. Pereira, M. A. Santos, and S. Macedo-Ribeiro. 2011. Unveiling the structural basis for translational ambiguity tolerance in a human fungal pathogen. Proc. Natl. Acad. Sci. USA 108: 14091-14096.

Santos, M. A., C. Cheesman, V. Costa, P. Moradas-Ferreira, and M. F. Tuite. 1999. Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in *Candida* spp. Mol. Microbiol. 31: 937-947.

Savageau, M. A., and R. R. Freter. 1979. On the evolution of accuracy and cost of proofreading tRNA aminoacylation. Proc. Natl. Acad. Sci. USA 76: 4507-4510.

Schimmel, P., and M. Guo. 2009. A tipping point for mistranslation and disease. Nat. Struct. Mol. Biol. 16: 348-349.

Schwanhäusser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach. 2011. Global quantification of mammalian gene expression control. Nature 473: 337-342.

Schwartz, M. H., and T. Pan. 2016. Temperature dependent mistranslation in a hyperthermophile adapts proteins to lower temperatures. Nucleic Acids Res. 44: 294-303.

Schwartz, M. H., and T. Pan. 2017. tRNA misacylation with methionine in the mouse gut micro-

biome *in situ.* Microb. Ecol. 74: 10-14.

Schwartz, M. H., J. R. Waldbauer, L. Zhang, and T. Pan. 2016. Global tRNA misacylation induced by anaerobiosis and antibiotic exposure broadly increases stress resistance in *Escherichia coli.* Nucleic Acids Res. 44: 10292-10303.

Shepherd, J., and M. Ibba. 2014. Relaxed substrate specificity leads to extensive tRNA mischarging by *Streptococcus pneumoniae* class I and class II aminoacyl-tRNA synthetases. MBio 5: e01656-14.

Springgate, C. F., and L. A. Loeb. 1975. On the fidelity of transcription by *Escherichia coli* ribonucleic acid polymerase. J. Mol. Biol. 97: 577-591.

Stadtman, E. R., and R. L. Levine. 2003. Free radical-mediated oxidation of free amino acids and amino acid residues in proteins. Amino Acids 25: 207-218.

Stillman, B. 1996. Cell cycle control of DNA replication. Science 274: 1659-1664.

Struhl, K. 2007. Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat. Struct. Mol. Biol. 14: 103-105.

Suzuki, Y., et al. 2000. Statistical analysis of the $5'$ untranslated region of human mRNA using "Oligo-Capped" cDNA libraries. Genomics 64: 286-297.

Sydow, J. F., F. Brueckner, A. C. Cheung, G. E. Damsma, S. Dengl, E. Lehmann, D. Vassylyev, and P. Cramer. 2009. Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. Mol. Cell 34: 710-721.

Sydow, J. F., and P. Cramer. 2009. RNA polymerase fidelity and transcriptional proofreading. Curr. Opin. Struct. Biol. 19: 732-739.

Taniguchi, Y., P. J. Choi, G. W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie. 2010. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science 329: 533-538.

Thompson, R. C., and P. J. Stone PJ. 1977. Proofreading of the codon-anticodon interaction on ribosomes. Proc. Natl. Acad. Sci. USA 74: 198-202.

Traverse, C. C., and H. Ochman. 2016. Conserved rates and patterns of transcription errors across bacterial growth states and lifestyles. Proc. Natl. Acad. Sci. USA 113: 3311-3316.

Turrientes, M. C., et al. 2013. Normal mutation rate variants arise in a mutator (mut S) *Escherichia coli* population. PLoS ONE 8: e72963.

Ucuncuoglu, S., K. L. Engel, P. K. Purohit, D. D. Dunlap, and D. A. Schneider, and L. Finzi. 2016. Direct characterization of transcription elongation by RNA polymerase I. PLoS One 11: e0159527.

Walmacq, C., M. L. Kireeva, J. Irvin, Y. Nedialkov, L. Lubkowska, F. Malagon, J. N. Strathern, and M. Kashlev. 2009. Rpb9 subunit controls transcription fidelity by delaying NTP sequestration in RNA polymerase II. J. Biol. Chem. 284: 19601-19612.

Wang, X., and T. Pan. 2015. Methionine mistranslation bypasses the restraint of the genetic code to generate mutant proteins with distinct activities. PLoS Genet. 11: e1005745.

Wang, X., and T. Pan. 2016. Stress response and adaptation mediated by amino acid misincorporation during protein synthesis. Adv. Nutr. 7: 773S-779S.

Wang, Y., C. L. Liu, J. D. Storey, R. J. Tibshirani, D. Herschlag, and P. O. Brown. 2002. Precision and functional specificity in mRNA decay. Proc. Natl. Acad. Sci. USA 99: 5860-5865.

Werner, F., and D. Grohmann. 2011. Evolution of multisubunit RNA polymerases in the three domains of life. Nat. Rev. Microbiol. 9: 85-98.

Wielgoss, S., et al. 2013. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. Proc. Natl. Acad. Sci. USA 110: 222-227.

Wierzbicki, A. T., T. S. Ream, J. R. Haag, and C. S. Pikaard. 2009. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nat. Genet. 41: 630-634.

Williams, L. N., A. J. Herr, and B. D. Preston. 2013. Emergence of DNA polymerase antimutators that escape error-induced extinction in yeast. Genetics 193: 751-770.

Wiltrout, E., J. M. Goodenbour, M. Fréchin, and T. Pan. 2012. Misacylation of tRNA with methionine in *Saccharomyces cerevisiae.* Nucleic Acids Res. 40: 10494-10506.

Wilusz, C. J., W. Wang, and S. W. Peltz. 2001. Curbing the nonsense: the activation and regulation of mRNA surveillance. Genes Dev. 15: 2781-2785.

Wohlgemuth, I., C. Pohl, J. Mittelstaet, A. L. Konevega, and M. V. Rodnina. 2011. Evolutionary optimization of speed and accuracy of decoding on the ribosome. Philos. Trans. R. Soc. Lond. B Biol. Sci. 366: 2979-2986.

Yadavalli, S. S., and M. Ibba. 2013. Selection of tRNA charging quality control mechanisms that increase mistranslation of the genetic code. Nucleic Acids Res. 41: 1104-1112.

Yamane, T., and J. J. Hopfield. 1977. Experimental evidence for kinetic proofreading in the aminoacylation of tRNA by synthetase. Proc. Natl. Acad. Sci. USA 74: 2246-2250.

Yates, J. L. 1979. Role of ribosomal protein S12 in discrimination of aminoacyl-tRNA. J. Biol. Chem. 254: 11550-11154.

Yuzenkova, Y., A. Bochkareva, V. R. Tadigotla, M. Roghanian, S. Zorov, K. Severinov, and N. Zenkin. 2010. Stepwise mechanism for transcription fidelity. BMC Biol. 8: 54.

Zaher, H. S., and R. Green. 2009. Fidelity at the molecular level: lessons from protein synthesis. Cell 136: 746-762.

Zaher, H. S., and R. Green. 2010. Hyperaccurate and error-prone ribosomes exploit distinct mechanisms during tRNA selection. Mol. Cell. 39: 110-120.

Zenkin, N., Y. Yuzenkova, and K. Severinov. 2006. Transcript-assisted transcriptional proofreading. Science 313: 518-520.

Zhang, J., K. W. Ieong, H. Mellenius, and M. Ehrenberg. 2016. Proofreading neutralizes potential error hotspots in genetic code translation by transfer RNAs. RNA 22: 896-904.

Figure 18.1. The multitude of functions that must be successfully navigated for the production of a properly translated and folded protein.

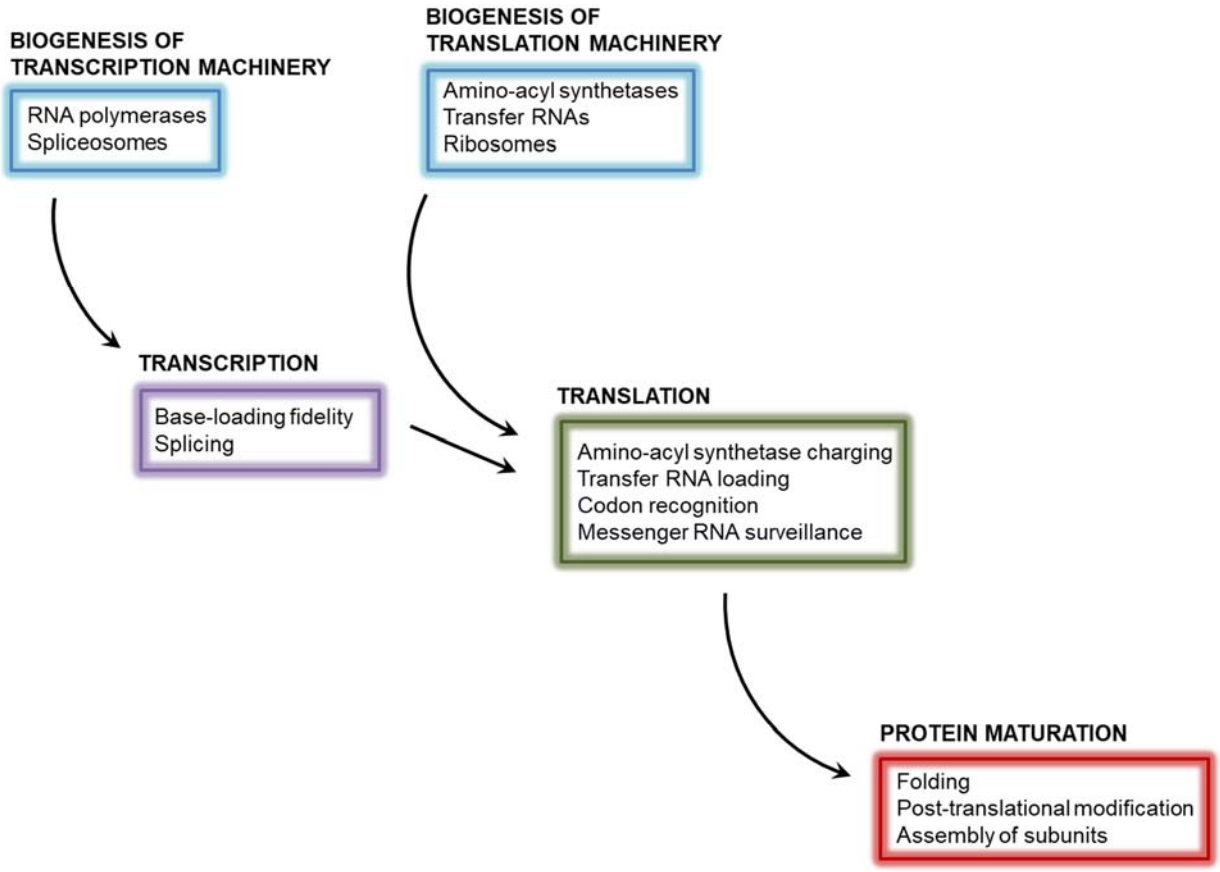

Figure 18.2. Reaction steps involving two competing Michaelis-Menten reactions involving right (R) and wrong (W) substrates.

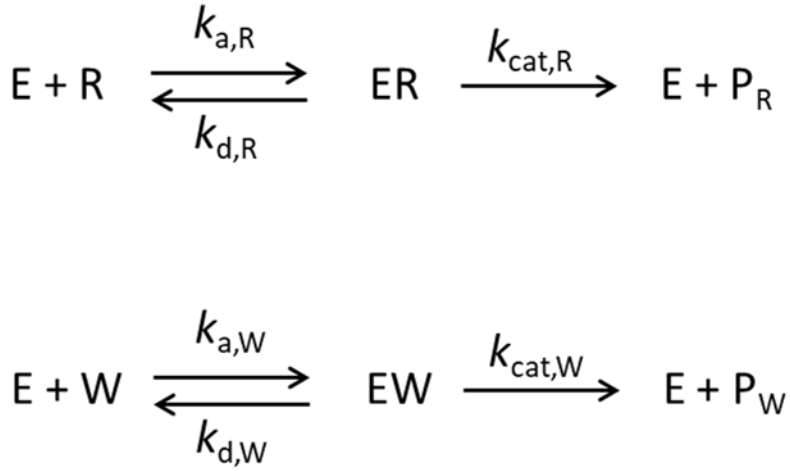

Figure 18.3. Modified reaction dynamics (compared to Figure 18.2) for the situation in which there is proofreading of an intermediate complex (denoted by asterisks), again for two competing substrates, R and W.

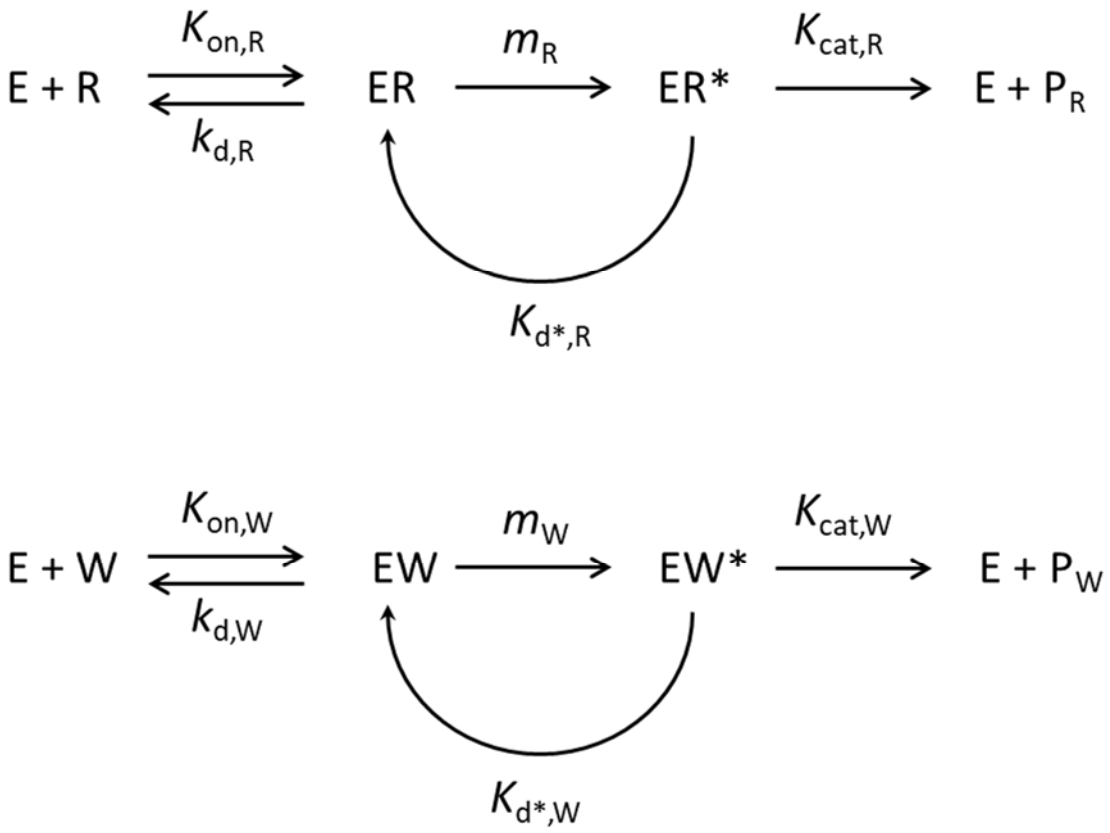

Figure 18.4. Negative phylogenetic association between transcript-error rates and a composite measure expected to be proportional to the efficiency of downward selection (based on the effective population size and the effective number of codons expected to be under selection; $N_e \cdot p\overline{L}\overline{n}\phi(\overline{n})$ as described in the text).

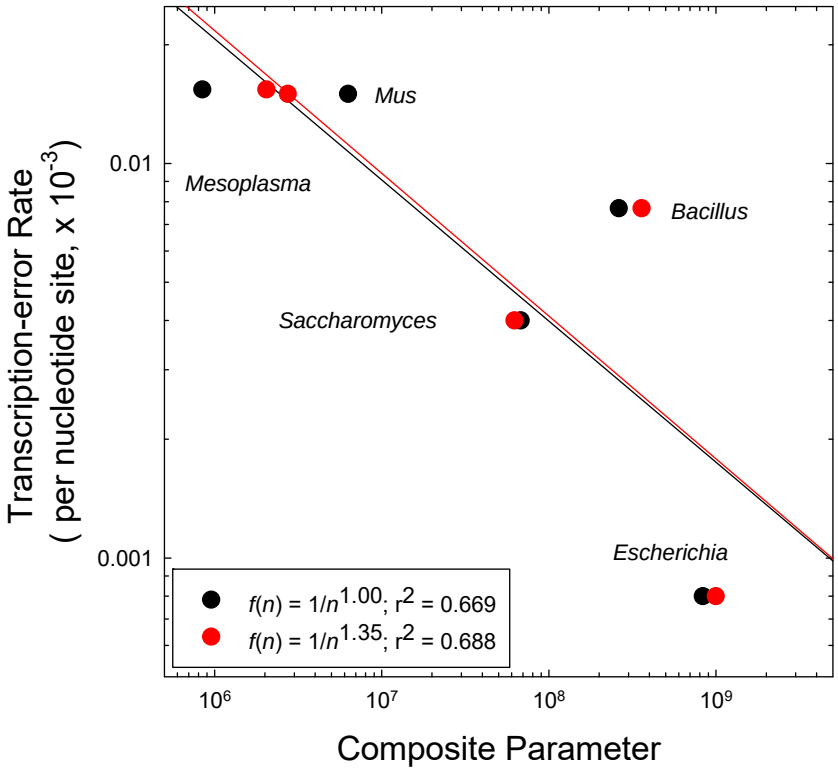

Figure 18.5. The net fitness effect of increasing accuracy can be viewed as the sum of the direct effects of errors (cellular damage), which declines with increasing accuracy, and the energetic costs of accuracy-improving mechanisms, which increase with accuracy. Depending on the exact forms of these functions, there will be an intermediate level of accuracy that minimizes the net selective disadvantage (red curve). If the minimum of this function is at a level of accuracy below the drift barrier (blue lines), it will be unachievable by natural selection.

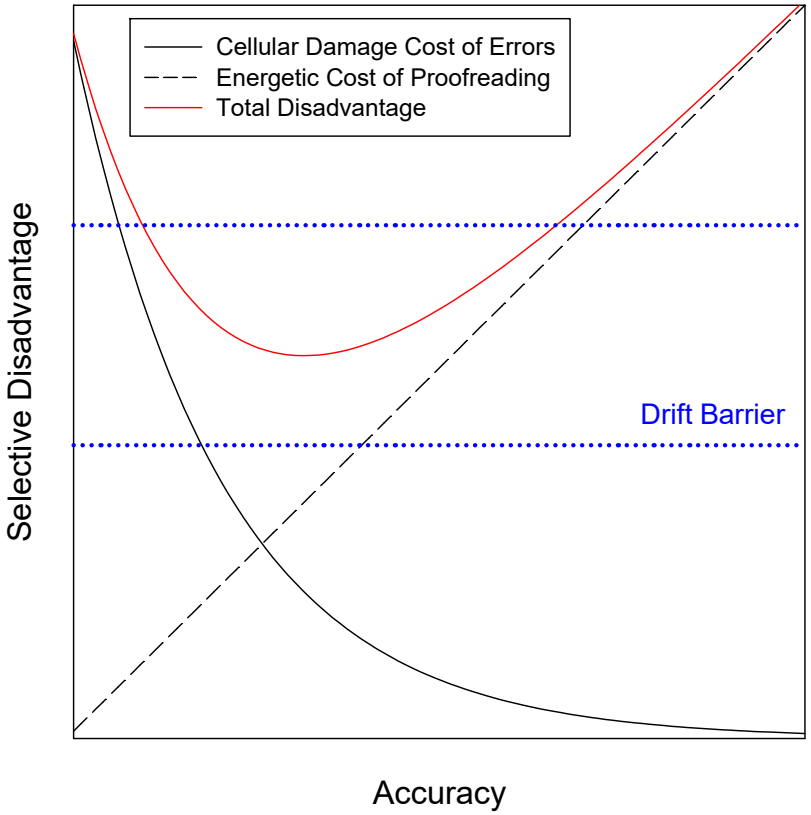

Figure 18.6. Left) Evolutionary addition of a secondary line of defense. Initially, the genome encodes for a single mechanism for error prevention (black line) leading to a level of perfection denoted by the dashed line (this is the drift barrier, the exact position of which is defined by the power of random genetic drift, and the magnitude of bias in the spectrum of mutations improving or degrading trait efficiency). A secondary adaptation (blue line) arises that then removes all but a fraction of the errors remaining after level one. The total error rate (red line) is given by the product of the error rates at each individual level, which act sequentially in the cell. Although there is an initial quantum leap in accuracy, slight imperfections (too small to be removed by selection) gradually accumulate at each level, returning the overall system to the same level as experienced prior to establishment of the second layer. **Right)** A bivariate drift barrier. If natural selection is capable of maintaining the two-layered system at a fitness level of $W = (1 - L_1)(1 - L_2)$, where L_1 and L_2 are the levels of inaccuracy at each layer, L_1 and L_2 are able to jointly move along the diagonal line of constant fitness W, with an improvement in one trait being balanced by a reduction in the other. Ultimately, the system may collapse to a single-trait state, as the less efficient trait approaches a state of nonfunctionality so closely that it is highly vulnerable to loss by further degenerative mutations and random genetic drift.