æ

20. INTRACELLULAR ERRORS

15 October 2022

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. The barrier to natural selection imposed by random genetic drift, combined with the recurrent introduction of deleterious mutations, insures that this will never happen, although such bounds may be closely approached in sufficiently large populations. Some would argue that biophysical tradeoffs with other traits prevent selection from attaining univariate optima, but this simply extends the drift barrier to two dimensions.

One consequence of these limitations is that 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 potential problems arising via dozens of cellular processes (Figure 20.1). A wide array of other cellular processes, including protein folding (Chapter 12), interactions of enzymes with inappropriate substrates (Chapter 19), and faulty assembly of proteins into higher-order structures (Chapter 13), are subject to error.

Cells have evolved multiple mechanisms that seemingly ameliorate the physiological consequences of error proliferation. 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 4). Problems arising in nonreplicating DNA are dealt with by a variety of repair mechanisms. As discussed below, 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.

This chapter focuses 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 the underlying processes. Transcription and translation errors arise at rates that are orders of magnitude higher than those incurred during replication (Chapter 4), and this is likely in part an evolutionary consequence of the transient nature of such errors. Unlike replication errors, which create cumulative damage in genomic regions linked to mutator alleles, errors arising during transcription and translation 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.

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 given this, 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, encouraging 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 idea that natural selection operates on the total performance of a system distributed over multiple traits, each of which is subject to a drift barrier.

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 most cytoplasmic ribosomal RNA subunits, a third (Pol III) for transfer RNA production, and still another for the mitochondrial genome (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 bacterial 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 subunits (Haag and Pickaard 2011). Yet, as will be noted below, despite the fact that eukaryotic RNA polymerases contain roughly twice the number of components as those from bacteria, there is no evidence that the former carry out their tasks much more efficiently.

Like replication, transcription involves phases of initiation, elongation, and termination, but several additional transcript-processing steps are involved as well (e.g., 5′-capping, intron removal, and addition of poly-A to 3′ ends of mRNAs in eukaryotes). 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 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. It is sometimes difficult to know whether all "aberrant" transcripts are actually errors as opposed to being products of various kinds of regulatory pathways, e.g., alternative transcription-initiation sites and/or RNA editing. However, most variants involve single-site changes sporadically

distributed across transcripts, often with downstream consequences with no obvious net benefits. The remaining discussion focuses on these erroneous base substitutions.

Although the mechanisms underlying transcription fidelity are not fully understood, they differ from those involved in replication (Chapter 8) in that RNA polymerases do not have separate domains set aside for proof-reading. However, the potential for error correction remains, as incorrect base incorporation results in polymerase pausing, providing time for loss of 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 molecule). Erroneous transcripts may typically arise by simple copying errors, and these may sometimes be exacerbated by base damage within the DNA template itself, as transcription often proceeds across damaged sites by substituting an inappropriate base, often an A (Brégeon et al. 2003; Clauson et al. 2010; Fritsch et al. 2020). However, as post-transcriptional errors (e.g., damaged bases) cannot be ruled out, the term transcript-error rather than transcriptional-error is adhered to here.

As the life spans of individual transcripts are generally substantially shorter than the life spans of cells (on the order of just minutes; Chapter 17), and most genes are represented in multiple mRNA copies per cell, transcription must progress at fairly high rates to meet cellular demands. Thus, unlike DNA, which only experiences one copying event per cell cycle, there can be dozens to hundreds of opportunities for error proliferation per transcribed site. Owing to their singular occurrences, transcript errors can be challenging to identify. One might think that such errors could easily be enumerated by simply comparing the sequences of RNAs to the expectations based on their genomic sources, but because the sequencing-error rate is generally substantially greater than the transcript-error rate, this approach will not work. Thus, until recently almost all information on transcript-error rates was obtained by 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} per nucleotide site (Springgate 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} per nucleotide site (de Mercoyrol et al. 1992). These early estimates, potentially subject to considerable experimental biases, suggested that transcript-error rates (in terms of sequence fidelity) fall in the broad range of 10^{-6} to 10^{-3} per nucleotide site.

More recently, it has become possible to estimate genome-wide *in vivo* error rates of transcripts by directly sequencing individual mRNA molecules multiple times (Gout et al. 2013; Traverse and Ochman 2016; Fritsch et al. 2020; Li and Lynch 2020; Li et al. 2022). These rates mostly fall in the range of 10^{-6} and 10^{-4}, and reveal no obvious phylogenetic pattern (McCandlish and Plotkin 2013; Li et al. 2022). The highest direct estimates are all from one study (Traverse and Ochman 2016), possibly a consequence of methodological issues (Gout et al. 2017;

Li and Lynch 2020; Meer et al. 2020), and if these are ignored, the range in known transcription-error rates shrinks to $\sim 10^{-6}$ to 10^{-5} (Figure 20.2a). Thus, the earlier *in vitro* estimates are anomalously high.

To put these rates into perspective, the estimated 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 30,000. For all other species for which both rates are known, the transcript-error rate is inflated by factors of 10^3 to 10^4 (Figure 20.2a). Thus, there is little question that rates of transcript-error production are substantially elevated relative to those arising during 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 on the order 0.1 to 1.0%.

Notably, these high error rates exist despite the proofreading capacity of RNA polymerases (Alic et al. 2007; Sydow and Cramer 2009). Nor are they a consequence of an exceptionally high speed of copying. The few estimates of average speeds of transcript progression exhibit only a small range of phylogenetic variation: 46 bp/sec in *E. coli* (Golding and Cox 2004; Proshkin et al. 2010); 20 to 60 in yeasts (Mason and Struhl 2005; Larson et al. 2012; Eser et al. 2016; Lisica et al. 2016; Ucuncuoglu et al. 2016); 21 in *Drosophila,* 23 in rat, 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.

Translational Fidelity

Even when an mRNA emerges error-free prior to translation, several additional challenges must be met if a faithful protein product is to be synthesized (Parker 1989; Zaher and Green 2009; Han et al. 2020). First, specific amino-acyl synthetase proteins (AARSs), each assigned to a single amino acid, 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. Most 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 proper 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. Misreads of sense codons as termination codons lead to prematurely truncated amino-acid chains, whereas misreading of termination codons as sense codons leads to termination read-through. Despite these immediate functional

consequences, translation-error rates 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-choice experiments, leaving questions as to the total error rate expected in a more natural setting in which 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 such 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 20.1). 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 is clear that the potential for translation error 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 at this stage fall in the range of 10^{-7} to 10^{-4}, all involving *E. coli* constructs (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 *in vivo* 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 devoid of codons for 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, correction needs to be made to account for the likely errors involving all amino acids not monitored. 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 20.1), substantially greater than the AARS-misloading rates. An attempt to estimate translation-error rates by mass spectrometry of individual protein fragments led to inexplicably low values in the approximate range of 10^{-4} to 10^{-3} for *E. coli* (Mordret et al. 2019).

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 protein 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 20.1). Thus, despite the estimation uncertainties involved, the translation-error rate per codon appears to be $\sim 10,000\times$ greater than the transcript-error rate per nucleotide site. This also means that only a small fraction of errors at the protein level are associated with transcription. Moreover, these high error rates arise even though the rate of translation is not much greater, and often lower, than that of transcription, e.g., 10 to 50 codons/sec in *E. coli* (with the rate increasing with increasing growth rate; Lovmar and Ehrenberg 2006), and 4 to 7 codons/sec in mouse (Gerashchenko et al. 2021).

Assuming an average translation-error rate of 0.01 per codon, a newly synthesized protein of moderate size, 300 amino acids, would contain an average of three erroneous amino acids, and assuming a Poisson distribution of errors, only 5% 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, even within a population of genetically uniform cells, each cell will harbor a statistically and transiently unique distribution of variants for most proteins. Recalling Equation 7.2a, 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 $\sim 5\times10^6$ 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 $\sim 6\times10^8$ errors, and a larger eukaryotic cell of $\sim 10^5\ \mu\text{m}^3$ harbors $\sim 4\times10^{10}$ errors.

Table 20.1. Estimated error rates associated with translation (rate of amino-acid misincorporation per codon). Standard errors are in parentheses.

Species	Synthetase Loading	Total Translation	Translation Read-through
Prokaryotes:			
Escherichia coli	0.0011 (0.0004)	0.0052 (0.0027)	0.0050 (0.0017)
Mycobacterium smegmatis		0.0168 (0.0147)	
Salmonella typhimurium			0.0034 (0.0014)
Five prokaryotic species	0.0017 (0.0009)		
Eukaryotes:			
Saccharomyces cerevisiae	0.0055 (0.0018)	0.0162 (0.0124)	0.0111 (0.0047)
Lupinus luteus	0.0009 (0.0005)		
Mus musculus		0.0152 (0.0065)	0.0042
Homo sapiens	0.0004 (0.0001)	0.0147 (0.0146)	

References: *E. coli*) Strigini and Brickman 1973; Edelmann and Gallant 1977; Fersht and Dingwall 1979; Fersht et al. 1980; Tsui and Fersht 1981; Miller and Albertini 1983; Bouadloun et al. 1983; Khazaie et al. 1984; Laughrea et al. 1987; Toth et al. 1988; Mikkola and Kurland 1992; Freist et al. 1998; Weickert and Apostol 1998; Chen et al. 2000; Beuning and Musier-Forsyth 2001; Tang and Tirrell 2002; Korencic et al. 2004; Lue and Kelley 2007; SternJohn et al. 2007; Kramer and

Farabaugh 2007; Guo et al. 2009; Reynolds et al. 2010; Zhang et al. 2013; Manickam et al. 2014. *M. smegmatis*) Javid et al. 2014; Leng et al. 2015. *S. typhimurium*) Roth 1970. Five prokaryotes) Fersht et al. 1980; Beuning and Musier-Forsyth 2001; Korencic et al. 2004; Li et al. 2011. *S. cerevisiae*) Igloi et al. 1978; Firoozan et al. 1991; Bonetti et al. 1995; Freist et al. 1998; Stansfield et al. 1998; Williams et al. 2004; Salas-Marco and Bedwell 2005; Plant et al. 2007; SternJohn et al. 2007; Kiktev et al. 2009; Reynolds et al. 2010; Kramer et al. 2010; Torabi and Kruglyak 2011; Loenarz et al. 2014. *L. luteus*) Jakubowski 1980. *M. musculus*) Mori et al. 1979; Cassan and Rousset 2001; Azpurua et al. 2013. *H. sapiens*) Beuning and Musier-Forsyth 2001; Lue and Kelley 2007; Chen et al. 2011; Zhang et al. 2013.

For the estimation of the total translation-error rate by misreading of a single codon to yield one specific amino acid, the observed error rate must be divided by the detectability of all possible misreads. Here, it is assumed that misreads only involve codon misevaluations at a single nucleotide site, so for any specific codon there are 9 possible misreads, of which some number $x < 9$ yield an amino-acid substitution (the remaining misreads leading to a stop, or owing to redundancy in the code, to no amino-acid substitution). A number $y \leq x$ of these reads lead to the misread being detected, so the detectability is y/x, and the total translation-error rate is taken to be the observed rate to the monitored amino acid divided by the detectability.

Biophysics of Substrate Discrimination and the Cost of Proofreading

As discussed in prior chapters, there are two potential limits to the evolutionary achievement of molecular perfection. The biophysical barrier represents the ultimate goal that might be attained by a supreme biochemist, capable of developing constructs constrained only by diffusion limitations and energetic features of substrate-binding mechanisms. The evolutionary barrier is the boundary set by the degree to which random genetic drift and deleterious mutation reduces the efficiency of natural selection. The biophysical issues are the subject of this section, with matters related to the drift barrier following below.

Errors arise during transcription and translation as a consequence of random diffusion and attachment of alternative substrates to catalytic sites. The frequency of inappropriate substrate usage depends on the relative concentrations and binding affinities involved. As the adhesion strengths of two alternative substrates become arbitrarily close, both substrates become equally likely to bind to an enzyme, provided the substrate concentrations are the same. As the difference in binding energies increases, the relative rate of an enzyme engaging with an inappropriate substrate declines exponentially. With unequal concentrations, the balance of competitive binding is tipped towards the more abundant substrate.

These issues can be addressed more formally as follows. Under a simple competitive binding situation, the error rate can be evaluated by considering two competing Michaelis-Menten reactions and the resultant ratio of rates of engagement with wrong (W) and right (R) substrate molecules (Foundations 20.1),

$$E = \frac{[\mathrm{W}]}{[\mathrm{R}]} \cdot \frac{k_{\mathrm{d,R}}}{k_{\mathrm{d,W}}}, \tag{20.1a}$$

where the first term represents the ratio of substrate concentrations (denoted by brackets), 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,

$$E = \frac{[\mathrm{W}]}{[\mathrm{R}]} \cdot e^{-\Delta E/(k_B T)}, \tag{20.1b}$$

where ΔE is the difference of binding strengths involving correct and incorrect substrate molecules, k_B is the Boltzmann constant, and T is the temperature in degrees Kelvin (see Foundations 7.3; and pp. 1011-1016 in Phillips et al. 2013). The actual error rate (the fraction of incorrect reactions, ϵ) is equivalent to $E/(1+E)$, but this is essentially the same as E provided $E \ll 1$. Assuming that this condition is met and that concentrations of alternative substrates are equal,

$$\epsilon \simeq e^{-\Delta E/(k_B T)}. \tag{20.2}$$

To gain an appreciation for the biological limits to accuracy under this simple model, 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. For example, a G:C pairing in DNA involves three hydrogen bonds, whereas an A:T pairing involves two. It then follows that binding-strength differentials between preferred and nonpreferred substrates (ΔE), especially those involving DNA and RNA, are generally on the order of 10 $k_B T$ or smaller, rendering most such biological processes 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, and 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 competitive binding of alternative nucleotides to single-stranded DNA.

Hopfield (1974) and Ninio (1975) realized how a form of proofreading, emerging from a simple deviation from Michaelis-Menten enzyme kinetics, can lead to a dramatic amplification in substrate fidelity (Foundations 20.1). 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 an energetic cost in the form of ATP or GTP hydrolyses, as well as in time to complete the forward reaction. 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.

The 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 an average 1.6 times prior to permanent attachment). In contrast, 270 ATPs were consumed per charged tRNA when only

valine was presented. Assuming equal substrate concentrations, these results suggest an error rate of $\sim 1.6/270 = 0.006$ resulting from the differential rejection of the two residue types.

Additional work revealed 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 (Yamane and Hopfield 1977). 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, respectively. In addition, when properly charged tRNAs encounter inappropriate codons during translation, GTP is hydrolyzed, implicating additional proofreading at the codon-anticodon recognition step on the ribosome (Thompson and Stone 1977; Yates 1979). Finally, 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 (Andersson et al. 1986).

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 baseline 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, insight into this matter can be gained by extending the example from the preceding paragraph. Assuming equal concentrations of isoleucine and valine molecules within the cytosol implies an additional consumption of $(1.6/270) \cdot (270 - 1.6) \simeq 1.6$ ATP molecules per fixed valine. Thus, assuming valine is the only erroneous substrate recognized by this particular AARS, the total cost of proofreading is approximately 3.2 ATP hydrolyses per amino-acid incorporation, i.e., a doubling of the ATP consumption were isoleucine to be the only substrate (which itself is 1.6 ATPs). The cost of engaging with any other erroneous substrate will still be $\simeq 1.6$ per substrate type, provided the number of cycles is $\gg 1.6$. These observations imply that ATP-dependent template copying mechanisms, which arose very early in cellular evolution, likely 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.

To put this in the context of a cell's total energy budget, recall from Chapter 8 that the cost of building a cell scales nearly linearly with cell volume, averaging $\sim 27V$ billion hydrolyzed ATPs (where V is cell volume in units of μm^3), whereas from Chapter 7, 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 less certain numbers, Savageau and Freter (1979) obtained an estimate of 2%. These rough calculations,

likely downwardly biased, indicate 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 20.1), but each increment in accuracy will involve additional energetic costs on the cell, hence reducing cellular rates of reproduction.

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. For example, the substantial room 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 these authors found no correlation between the growth rates and translation accuracies of different strains, this is perhaps not surprising given the difficulties of measuring both parameters to a high degree of resolution. At high-nutrient levels, wild-type *E. coli* grow more rapidly than those with hyper-accurate ribosomes, but growth rates are approximately the same under low-nutrient conditions (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 is that genomic errors remain associated with mutator alleles until separated by recombination, which reinforces the fitness effects indirectly induced by a mutator, whereas transcription and translation errors are transient. The effects of a DNA-level mutator allele gradually build up over time with the accumulation of linked deleterious mutations. In contrast, the full effects of an allele reducing transcription accuracy are immediate, but not cumulative. The half lives of individual transcripts are typically 5 to 10 minutes in bacteria, 10 to 20 minutes in yeast, and a few hours in mammalian cells, shorter than the time necessary for cell division in all cases (Chapter 17).

It need not follow, however, that the damage from transcript errors is quickly erased. Rather, a roughly steady-state density of erroneous transcripts can be expected within the cell, reflecting a balance between the decay of old mRNAs and the transcriptional production of new ones, and the same will be true of erroneous proteins. Thus, regardless of the transience of individual errors, one can expect a relatively constant number of total errors per cell, although the specific errors will vary temporally. For each expressed protein-coding gene, the total expected number of erroneous amino acids 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). As noted in Chapter 10, the conditions for equilibrium might sometimes be violated if the machinery responsible for error minimization become compromised with cell age.

An additional issue here is that 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 copy numbers of the protein free of the specific error. The cellular 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). For a gene with a steady-state number n of transcripts or proteins per cell, 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 20.2). In the case of additive effects, the dilution factor is simply $1/n$, and the net effect of errors is independent of the number of molecules per cell, as $n \cdot (1/n) = 1$.

More generally, to obtain the total burden of errors on fitness, we require the total number of erroneous amino acids within proteins, each discounted by the appropriate 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 fraction of codons with errors (which is the translation error rate in the case of proteins, and roughly three-quarters of 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 (transcription or translation) is then $uP\overline{Ln}$, where $\overline{Ln}$ is the mean number of nucleotide sites in expressed transcripts per gene. Further letting δ 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 $uP\overline{Ln\phi(n)\delta}$, where the overline denotes an average value. This implies that the selection coefficient associated with a modifier of the transcript- or translational-error rate scales with $P\overline{Ln\phi(n)\delta}$ or just $P\overline{L\delta}$ if the dilution factor is $1/n$.

As in the case of replication fidelity (Chapter 4), theory suggests that the efficiency of 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 (Foundations 20.2). If this general theoretical framework is correct, we then expect u to scale inversely with $N_eP\overline{Ln\phi(n)}$, provided δ is fairly constant across species, which further reduces to $N_eP\overline{L}$ if $\phi(n) \simeq 1/n$. Because all of the underlying cellular factors can covary with each other, as well as with N_e (e.g., organisms with small N_e tend to have relatively large P and L; Chapter 24), it is possible that no simple scaling with single parameters will emerge.

Nonetheless, the data are consistent with a negative scaling of u with the composite parameter. From Chapter 7, we know how the number of molecules n scales with cell volume, which enters $\overline{n\phi(n)}$ (Foundations 20.3). Estimates of P and $\overline{L}$ are generally available from genome sequencing data, and N_e from population-level polymorphism data (Chapter 4). For the 15 unicellular species for which data are available for these parameters, estimates of the transcript-error rate scale negatively with the composite parameter $N_eP\overline{Ln\phi(n)}$ (Figure 20.2b). The best fit to the data is obtained when $\phi(n) = 1/n$, i.e., additivity. Although the slope of the overall regression is much shallower than the expectation of -1, this might be a consequence of variation in fitness effects associated with transcript/translational fidelity, e.g., if $\overline{\delta}$ (which has been ignored in Figure 20.2b) were to decline with increasing $N_eP\overline{Ln}$ (Li et al. 2022).

Another suggested factor in the evolution of high transcription and translation error rates is the energetic cost of kinetic proofreading (Ehrenberg and Kurland 1984). The idea here is that whereas increasingly accurate transcription and translation should improve cell health, these advantages might become offset by the costs of proofreading, which generally consumes ATP/GTP and magnifies the time to progression to successful polymerization. Under this view, too high or too low of a level of accuracy leads to reduced fitness, motivating the idea that the fidelity of transcription / translation may be kept at some intermediate optimal state by natural selection (Ehrenberg and Kurland 1984; Kurland and Ehrenberg 1984, 1987). This kind of speed / accuracy tradeoff argument has been made in numerous contexts in the biophysical literature (Foundations 19.3, 20.1, and 22.2).

However, such optimization would only be possible if the supposed optimum error rate were accessible relative to the drift barrier, else the latter would take precedence. Another implicit assumption in this optimization 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, i.e., in the kinetic coefficients. Finally, this optimization hypothesis does not explain the phylogenetic range (or lack thereof) in error rates, given the substantial differences in cell-division times among study species, which might alter the premium put on the speed of transcription.

The Evolutionary Consequences of Surveillance-mechanism Layering

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 initial base misincorporations being subject to correction by subsequent proofreading, and the still smaller fraction of errors that escape proofreading being subject to mismatch repair.

Layering of defense mechanisms also occurs with transcript surveillance. In a number of organisms, the subset of erroneous mRNAs missing either 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. 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 either restricted to or substantially elaborated in the eukaryotic lineage.

The eukaryotic NMD process is mechanistically 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 leads to a downstream frameshift two-thirds of the time, typically resulting in 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. The NMD process removes many such transcripts during their first round of translation, distinguishing erroneous from true termination codons by use of information on the distance from 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; Zhang et al. 2010).

Although not all PTC-containing mRNAs are detectable by the NMD process, the importance of NMD is illustrated by experiments in which the pathway has been silenced, which show substantial increases in PTC-containing mRNAs (Mendell et al. 2004; Mitrovich and Anderson 2000; 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 the fact that some of the proteins in the NMD pathway have additional cellular functions (Maquat 2006; Isken and Maquat 2008).

Mechanisms for surveillance and removal of faulty products also exist for noncoding RNAs, such as ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs). For example, eukaryotes (but apparently not bacteria) have a decay pathway capable of detecting and removing a subset of rRNAs with erroneous sequences (LaRiviere et al. 2006; Andersen et al. 2008; Sarkar et al. 2017; Lirussi et al. 2021). Multiple mechanisms for the removal of aberrant tRNAs also exist in prokaryotes and eukaryotes (Wilusz et al. 2011; Kramer and Hopper 2013). In addition, the accuracy of translation 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 adaptive evolutionary processes, the idea being that 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 levels with the addition of more and more layers of error correction.

However, this view of ever-improving fidelity is likely incorrect. It fails to explain, for example, why eukaryotes have more elaborate surveillance mechanisms for errors in DNA, RNAs, and proteins than do bacteria, and yet still exhibit equal or higher net rates of error production. 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 result in decay back to the drift barrier, rendering the overall system no more accurate than the previous single-layered system (Figure 20.3). 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 having identical overall accuracies (Figure 20.3), 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 this view. For example, 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. However, *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 those in other microbes (Long et al. 2015; Kucukyildirim et al. 2016). In addition, mycobacteria and other species lack the DNA polymerase proof-reading domain deployed by *E. coli* and many other bacteria, but harbor an entirely different mechanism for such purposes (Rock et al. 2015). 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; Wei et al. 2022). Finally, different yeast species with very similar mutation rates have substantially different mutation spectra, implying variation in the underlying mechanisms of mutation and repair that nonetheless have the same net effects (Long et al. 2016; Nguyen et al. 2020).

From the standpoint of evolutionary layering and translational fidelity, one can also point to exemplary cases 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 simply 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. In budding yeast (*S. cerevisiae*), cytoplasmic PheRS is capable of editing a tRNA mischarged with an erroneous amino acid, 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 offset the loss of 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 fidelity, with a switch to strong reliance on just a single mechanism being conferred by no more than two amino-acid substitutions in some lineages (Martinis and Boniecki 2010).

Adaptive Significance of Errors

Given that the vast majority of amino-acid altering mutations have negative fitness effects (Chapter 12), 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 translation-fidelity mutations, Hughes (1991; Hughes and Tubulekas 1993) found that a ten-fold increase in the error rate generates a two-fold reduction in growth rate, accompanied by substantial reductions in enzyme activity and protein stability, with no associated change in protein abundance. Similarly, in comparisons of wild-type *E. coli* and a mutant line with enhanced translational fidelity, Musa et al. (2016) found an $\sim 30\%$ increase in enzyme activity in the latter.

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 labeled 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 extreme adaptationist 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 an intermediate level of mistranslation, fine-tuned by natural selection, yields populations of variant molecules, some of which will have large enough fitness-enhancing functions to offset the deleterious effects of others. This line of thinking derives from a rather liberal interpretation of multiple lines of observation.

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 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 freely available in the environment, in which case there would be no selection for phenotypic reversion. Thus, in natural settings, auxotrophic mutants are the least likely recipients

of transiently beneficial cellular errors.

Second, cells under stress often have increased translation-error rates. For example, in bacteria exposed to 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; Vargas-Rodriguez et al. 2021). The mischarging of non-methionine tRNAs with methionine is particularly common in stressful situations, with up to 10% of incorporated methionine being erroneous in mammalian, yeast, and bacterial cells under some conditions (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, and in *E. coli,* excess carbon leads to stop-codon read-through (Zhang et al. 2020). It has been argued that 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). However, cases are known in which stress leads to increased translational accuracy (Steiner et al. 2019), suggesting the possibility of reporting bias.

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 might in turn protect the proteins containing them. Moreover, a common enzyme (methionine sulfoxide reductase) confers the ability to revert the modified amino acid 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. Given that there are no known disadvantages of methionine in nonstressful conditions, if methionine serves such a useful function, why then is the cytoplasm not populated with higher free amounts of this amino acid, and why are more methionines not directly encoded into the proteome?

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). This is expected to arise naturally from the reduction in competitive binding between substrate types (Foundations 20.1). In some cases, such misloadings can provide 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 given the alternative outcome of no translation and death? In nature, a more common situation would be generic shortage of all amino acids, which as discussed in Foundations 20.1 is expected to reduce the rate of translation but not magnify the error rate.

Fourth, some organisms, such as members of the genus *Mycoplasma,* have one or more AARSs with editing defective domains (Li et al. 2011). The microsporidian *Vavraia culicis* harbors a defective leucine AARS, which misuses a variety of alter-

native amino acids up to 6% of the time (Melnikov et al. 2018). The fungal pathogen *Candida albicans* has experienced a reassignment of one particular leucine codon to serine (an alteration of the genetic code), but still misincorporates 1 to 6% leucines at such codons (Rocha et al. 2011; Bezerra et al. 2013). This codon ambiguity reduces fitness in normal environments, but by inducing the expression of stress-response proteins 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 errors is promoted by natural selection.

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 no direct evidence that error-prone translation has been promoted and/or maintained 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 immune to pathological behavior. 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. Notably, many organisms with intrinsically defective translation efficiency are pathogens, which may be highly vulnerable to random genetic drift and loss of nonessential functions.

Summary

- The production of proteins and functional RNAs with correct sequences relies on the avoidance of error proliferation at a myriad of steps, ranging from transcription and translational fidelity to mRNA splicing to protein folding and assembly.

- Transcript-error rates are universally orders of magnitude greater than rates of mutation at the DNA level, and translation-error rates are still orders of magnitude higher. Thus, for any given genetic locus, cells typically contain populations of variant molecules.

- Errors at the level of transcription and translation are mitigated by kinetic proofreading, which relies on the competitive binding and release of correct and incorrect substrate molecules. Such processes consume considerable energy and/or extend the time to complete polymerization processes, thereby leading to intrinsic tradeoffs between speed, energetic expense, and accuracy.

- Theory and observation suggest that the ability of selection to reduce transcript-error rates is a composite function of the effective population size, proteome size,

and number of transcripts per cell, but the negative scaling is weaker than that observed for mutation rates at the DNA level.

- Although evolution has resulted in multiple layers of error-correcting mechanisms, particularly in eukaryotes, this does not appear to result in a long-term increase in accuracy. Rather, stepwise increases in fidelity are eventually dissipated evolutionarily, eventually resulting in a more complex system with no overall level of improvement. This leads to the false impression that multiple lines of defense imply an evolutionary enhancement of organismal robustness to error production.

- Arguments have been made that high rates of transcript and translation errors are promoted for their positive effects, but such conclusions are extreme examples of pan-adaptationist thinking gone wild. Benefits can sometimes be found in particular settings, but significant deleterious effects are more pervasive. As the particular kinds of errors produced are idiosyncratic across species, and most species in which they have been identified are pathogens with small effective population sizes, an alternative view is that high rates of transient errors are simply a consequence of the compromised power of natural selection.

Foundations 20.1. Kinetic proofreading. Numerous cell biological processes, including steps involved in DNA replication, transcription, and mRNA translation, involve various forms of proofreading. 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 an increase in accuracy, we first consider the null situation in which two alternative substrates are engaged in the same Michaelis-Menten reaction (Chapter 19), albeit at different rates (Figure 20.4). The right and wrong substrates will be designated R and W 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 (Foundations 19.1), which then either returns to the prior state (E and X) by dissociation or proceeds to product formation, at rates $k_{\mathrm{d,X}}$ and $k_{\mathrm{cat,X}}$, respectively.

The rate of production of the correct intermediate per unit enzyme is $k_{\mathrm{a,R}}[\mathrm{R}]$, where the brackets denote concentration, and once formed the intermediate is converted to final product with probability $\lambda_{\mathrm{R}} = k_{\mathrm{cat,R}}/(k_{\mathrm{d,R}} + k_{\mathrm{cat,R}})$, 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),

$$\epsilon = \frac{k_{\mathrm{a,W}} \cdot \lambda_{\mathrm{W}} \cdot [\mathrm{W}]}{k_{\mathrm{a,R}} \cdot \lambda_{\mathrm{R}} \cdot [\mathrm{R}]}, \tag{20.2.1a}$$

which can be seen to depend on ratios involving both substrate concentrations and kinetic coefficients. Assuming 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,

$$\epsilon = \frac{\lambda_{\mathrm{W}}}{\lambda_{\mathrm{R}}} = \frac{k_{\mathrm{cat,W}} \cdot (k_{\mathrm{d,R}} + k_{\mathrm{cat,R}})}{k_{\mathrm{cat,R}} \cdot (k_{\mathrm{d,W}} + k_{\mathrm{cat,W}})}. \tag{20.2.1b}$$

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

$$\epsilon \simeq \frac{k_{\mathrm{cat,W}}}{k_{\mathrm{cat,R}} + k_{\mathrm{cat,W}}}. \tag{20.2.1c}$$

which is equal to one half when the two catalytic rates are equivalent. On the other hand, if both dissociation rates are large relative to the catalytic rates, and the latter are equivalent for both substrate molecules, Equation 20.2.1b reduces to the ratio of dissociation rates,

$$\epsilon \simeq \frac{k_{\mathrm{d,R}}}{k_{\mathrm{d,W}}}. \tag{20.2.1d}$$

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 20.5). 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 elevated 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, except for the dissociation constants, are equal for both substrates.

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,

$$\lambda_{1,\mathrm{X}} = k_{\mathrm{on}}[\mathrm{X}] \qquad \text{for E + X} \rightarrow \mathrm{EX} \tag{20.2.2a}$$

$$\lambda_{2,\mathrm{X}} = \frac{m}{m + k_{\mathrm{d,X}}} \qquad \text{for EX} \rightarrow \mathrm{EX}^* \tag{20.2.2b}$$

$$\lambda_{3,\mathrm{X}} = \frac{k_{\mathrm{d}^*,\mathrm{X}}}{k_{\mathrm{cat}} + k_{\mathrm{d}^*,\mathrm{X}}} \qquad \text{for EX}^* \rightarrow \mathrm{EX} \tag{20.2.2c}$$

$$\lambda_{4,\mathrm{X}} = \frac{k_{\mathrm{cat}}}{k_{\mathrm{cat}} + k_{\mathrm{d}^*,\mathrm{X}}} \qquad \text{for EX}^* \rightarrow \mathrm{E + P_X}\ . \tag{20.2.2d}$$

The overall forward rate of production of $\mathrm{P_X}$ can be summarized as the series

$$\Lambda_{1,\mathrm{X}} = \lambda_{1,\mathrm{X}}\lambda_{2,\mathrm{X}}\lambda_{4,\mathrm{X}}[1 + (\lambda_{2,\mathrm{X}}\lambda_{3,\mathrm{X}}) + (\lambda_{2,\mathrm{X}}\lambda_{3,\mathrm{X}})^2 + \cdots\ ] \tag{20.2.3a}$$

$$= \frac{\lambda_{1,\mathrm{X}}\lambda_{2,\mathrm{X}}\lambda_{4,\mathrm{X}}}{1 - \lambda_{2,\mathrm{X}}\lambda_{3,\mathrm{X}}}. \tag{20.2.3b}$$

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

$$\epsilon = \frac{\lambda_{2,\mathrm{W}}\lambda_{4,\mathrm{W}}(1 - \lambda_{2,\mathrm{R}}\lambda_{3,\mathrm{R}})}{\lambda_{2,\mathrm{R}}\lambda_{4,\mathrm{R}}(1 - \lambda_{2,\mathrm{W}}\lambda_{3,\mathrm{W}})}, \tag{20.2.4}$$

and substituting from above yields

$$\epsilon = \frac{mk_{\mathrm{cat}} + k_{\mathrm{d,R}}k_{\mathrm{cat}} + k_{\mathrm{d,R}}k_{\mathrm{d}^*,\mathrm{R}}}{mk_{\mathrm{cat}} + k_{\mathrm{d,W}}k_{\mathrm{cat}} + k_{\mathrm{d,W}}k_{\mathrm{d}^*,\mathrm{W}}} \tag{20.2.5a}$$

The assumption that m and k_{cat} are the same for both substrate molecules can be relaxed to allow for inequalities in Equations 20.2.3b and 20.2.4, but this will not be pursued here.

Comparison to Equation 20.2.1c 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 dissociation coefficients at both steps are the same, i.e., $k_{\mathrm{d,X}} \propto k_{\mathrm{d}^*,\mathrm{X}}$,

$$\epsilon \simeq \left(\frac{k_{\mathrm{d,R}}}{k_{\mathrm{d,W}}}\right)^2. \tag{20.2.5b}$$

Thus, in this limiting case, a proofreading step can reduce the error rate down to the square of that expected in the absence of proofreading, Equation 20.2.1.d. These specific results, 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 also increases the reaction time (along with the energetic cost), implying an intrinsic tradeoff between accuracy and speed. Expanding from the logic underlying Equation 20.2.3a, the average number of intermediate steps that are cycled through before reaching the final product is

$$\begin{aligned} n_c &= \sum_{i=1}^{\infty} i \cdot (\lambda_{2,\mathrm{X}}\lambda_{3,\mathrm{X}})^{i-1} \\ &= \frac{1}{[1-\lambda_{2,\mathrm{X}}\lambda_{3,\mathrm{X}}]^2} \end{aligned} \tag{20.2.6}$$

Note that $\lambda_{2,\mathrm{X}}\lambda_{3,\mathrm{X}}$ is a measure of the probability of transition from $\mathrm{EX} \rightarrow \mathrm{EX}^* \rightarrow \mathrm{EX}$, i.e., of recycling. The time to complete a reaction is proportional to n_c.

This being said, it can also 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 20.5, some of which are jointly favorable for both traits and others of which are not. A lucid examination of these models, including an extension to an arbitrary number of correction steps is provided by Qian (2006).

Foundations 20.2. The evolutionary bounds on the transcription-error rate. Given that natural selection relentlessly promotes the sequences of protein-coding genes toward their optimal functional state, most errors in transcripts are expected to be deleterious (Chapter 12). Here, we consider the expected selective advantage of a genomic variant that improves transcription 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) at the amino-acid sequence level (where fitness effects are manifest); 2) the total number of molecules per cell associated with each gene; and 3) the fitness effects of such errors.

Letting u be the rate of missense/nonsense transcript error per codon, and L_i be the number of amino acids in a protein of type i, the number (j) of errors in individual protein molecules of this type will be Poisson distributed with expectation uL_i,

$$P(j|uL_i) = \frac{e^{-uL_i}(uL_i)^j}{j!}. \tag{20.2.1}$$

From the standpoint of translation errors, u is defined to be the error-rate per codon, whereas if μ were to be the transcript-error rate per nucleotide site, as it is here, because

there are three nucleotide sites per codon, and $\sim 3/4$s of nucleotide substitutions cause an amino-acid substitution, $u \simeq 3\mu\cdot(3/4) = 9\mu/4$, i.e., the error rate per codon $\simeq 2.25\times$ the error rate per nucleotide site.

Because transcription and translation errors are singular events, individual variant proteins within a cell 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 genetic 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 effects is

$$f(n_i) = \frac{1}{n_i^x}, \tag{20.2.2}$$

which equals 1.0 when $n_i = 1$ (effects are fully penetrant), 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). Positive synergistic effects are implied by $x < 0$, whereas $x > 1$ results in a relatively rapid decline in $f(n_i)$ (i.e., relatively recessive effects of errors).

From Chapter 12, the average reduction in fitness associated with single, fully expressed deleterious mutations in proteins is generally < 0.1, and based on known transcript- and translation-error rates and typical gene lengths, the number of errors per protein will generally be $\ll 10$. Thus, letting δ_i be the average fitness loss per single error in a single molecule if fully revealed, unless there are very strong epistatic effects, the total reduction in fitness associated with a protein containing j errors $\simeq j\delta_i$. Provided $j\delta_i \ll 1$, it is useful to use the further approximation, $j\delta_i \simeq 1 - e^{-j\delta_i}$.

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 $P(j|uL_i)$,

$$s_i = n_i \cdot f(n_i) \cdot \sum_{j=1}^{L_i} P(j|uL_i) \cdot (1 - e^{-j\delta_i}), \tag{20.2.3a}$$

$$\simeq n_i \cdot f(n_i) \cdot \left[1 - \exp\left(-\frac{uL_i\delta_i}{1+\delta_i}\right)\right]. \tag{20.2.3b}$$

Further simplification is obtained by noting that because δ_i and $uL_i\delta_i \ll 1$,

$$s_i \simeq uL_i \cdot n_i \cdot [\, f(n_i) \cdot \delta_i \,]. \tag{20.2.3c}$$

Equation 20.2.3c shows that the fitness consequences of mistranslation are a function of three quantities: 1) the total mistranslation rate per expressed protein (uL_i); 2) the steady-state number of proteins per gene (n_i); and 3) the average effect of an amino-acid substitution (the final term in brackets, which is a function of the deleterious effect of a fully penetrant mutation and the dilution effect).

Finally, we require an expression for the entire burden of errors throughout the proteome for cell fitness. Here, letting P denote the number of protein-coding loci, we

will assume that the effects of errors arising at each genetic locus act independently, such that total mean fitness can be written as

$$W(u) = \prod_{i=1}^{P}(1-s_i) \simeq \exp\left(-\sum_{i=1}^{P} s_i\right) \simeq 1 - \left\{(uP\overline{L}) \cdot [\overline{n \cdot f(n)}] \cdot \overline{\delta}\right\}, \quad (20.2.4)$$

where an overline denotes an average over loci. The first approximation follows from $s_i \ll 1$, and the second from the assumption of independent n_i, L_i, and δ_i. 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 4, 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, 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 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

$$N_e = \frac{1}{W[(1-\Delta)u^*] - W[u^*]}. \quad (20.2.5)$$

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

$$u^* \simeq \frac{1}{\Delta \cdot N_e \cdot P\overline{L} \cdot \overline{n}\phi(\overline{n}) \cdot \overline{\delta}}. \quad (20.2.6)$$

where $\phi(\overline{n}) = \{1 - [x(1-x)C_n^2/2]\}/\overline{n}^x$ is the average dilution factor, with $\overline{n}$ and C_n denoting the mean and coefficient of variation in expression level (obtained by Taylor-series expansion of Equation 20.2.2). This shows that, all other things being equal, there is an expected inverse relationship between the drift barrier and N_e. However, there is also an inverse scaling with the total number of codons in the proteome $P\overline{L}$, with the copy-number effect $\overline{n}\phi(\overline{n})$, and with the average fitness effect of mutations $\overline{\delta}$. 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.

Literature Cited

Alic, N., N. Ayoub, E. Landrieux, E. Favry, P. Baudouin-Cornu, M. Riva, and C. Carles. 2007. Selectivity and proofreading both contribute significantly to the fidelity of RNA polymerase III transcription. Proc. Natl. Acad. Sci. USA 104: 10400-10405.

Andersen, K. R., T. H. Jensen and D. E. Brodersen. 2008. Take the "A" tail – quality control of ribosomal and transfer RNA. Biochim. Biophys. Acta 1779: 532-537.

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.

Azpurua, J., Z. Ke, I. X. Chen, Q. Zhang, D. N. Ermolenko, Z. D. Zhang, V. Gorbunova, and A. Seluanov. 2013. Naked mole-rat has increased translational fidelity compared with the mouse, as well as a unique 28S ribosomal RNA cleavage. Proc. Natl. Acad. Sci. USA 110: 17350-17355.

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.

Bacher, J. M., W. F. Waas, D. Metzgar, V. de Crécy-Lagard, and P. Schimmel. 2007. Genetic code ambiguity confers a selective advantage on *Acinetobacter baylyi*. J. Bacteriol. 189: 6494-6496.

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.

Beuning, P. J., and K. Musier-Forsyth. 2001. Species-specific differences in amino acid editing by class II prolyl-tRNA synthetase. J. Biol. Chem. 276: 30779-30785.

Bezerra, A. R., J. Simões, W. Lee, J. Rung, T. Weil, I. G. Gut, M. Gut, M. Bayés, L. Rizzetto, D. Cavalieri, 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.

Bonetti, B., Fu L, Moon J, and Bedwell DM. 1995. The efficiency of translation termination is determined by a synergistic interplay between upstream and downstream sequences in *Saccharomyces cerevisiae*. J. Mol. Biol. 251: 334-345.

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

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.

Cassan, M., and J. P. Rousset. 2001. UAG readthrough in mammalian cells: effect of upstream and downstream stop codon contexts reveal different signals. BMC Mol. Biol. 2: 3.

Chen, J. F., N. N. Guo, T. Li, E. D. Wang, and Y. L. Wang. 2000. CP1 domain in *Escherichia coli* leucyl-tRNA synthetase is crucial for its editing function. Biochemistry 39: 6726-6731.

Chen, X., J. J. Ma, M. Tan, P. Yao, Q. H. Hu, G. Eriani, and E. D. Wang. 2011. Modular pathways for editing non-cognate amino acids by human cytoplasmic leucyl-tRNA synthetase. Nucleic Acids Res. 39: 235-247.

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.

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.

Edelmann, P., and J. Gallant. 1977. Mistranslation in *E. coli.* Cell 10: 131-137.

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

Eser, P., L. Wachutka, K. C. Maier, C. Demel, M. Boroni, S. Iyer, P. Cramer, and J. Gagneur. 2016. Determinants of RNA metabolism in the *Schizosaccharomyces pombe* genome. Mol. Syst. Biol. 12: 857.

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

Fersht, A. R., and C. Dingwall. 1979. An editing mechanism for the methionyl-tRNA synthetase in the selection of amino acids in protein synthesis. Biochemistry 18: 1250-1256.

Fersht, A. R., J. S. Shindler, and W. C. Tsui. 1980. Probing the limits of protein-amino acid side chain recognition with the aminoacyl-tRNA synthetases. Discrimination against phenylalanine by tyrosyl-tRNA synthetases. Biochemistry 19: 5520-5524.

Firoozan, M., C. M. Grant, J. A. Duarte, and M. F. Tuite. 1991. Quantitation of readthrough of termination codons in yeast using a novel gene fusion assay. Yeast 7: 173-183.

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

Freist, W., H. Sternbach, I. Pardowitz, and F. Cramer. 1998. Accuracy of protein biosynthesis: quasi-species nature of proteins and possibility of error catastrophes. J. Theor. Biol. 193: 19-38.

Fritsch, C., J.-F. Gout, S. Haroon, A. Towheed, X. Zhang, Y. Song, S. Simpson, D. Wallace, K. Thomas, M. Lynch, et al. 2020. Genome-wide surveillance of transcription errors in response to genotoxic stress. Proc. Natl. Acad. Sci. USA 118: e2004077118.

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.

Gerashchenko, M. V., Z. Peterfi, S. H. Yim, and V. N. Gladyshev. 2021. Translation elongation rate varies among organs and decreases with age. Nucleic Acids Res. 49: e9.

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, et al. 2017. The landscape of transcription errors in eukaryotic cells. Science Advances 3: e1701484.

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: 555-564.

Guo, M., Y. E. Chong, R. Shapiro, K. Beebe, X. L. Yang, and P. Schimmel. 2009. Paradox of mistranslation of serine for alanine caused by AlaRS recognition dilemma. Nature 462: 808-812.

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.

Han, N. C., P. Kelly, and M. Ibba. 2020. Translational quality control and reprogramming during stress adaptation. Exp. Cell Res. 394: 112161.

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

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

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.

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.

Igloi, G. L., F. von der Haar, and F. Cramer. 1978. Aminoacyl-tRNA synthetases from yeast: generality of chemical proofreading in the prevention of misaminoacylation of tRNA. Biochemistry 17: 3459-3468.

Isken, O., and L. E. Maquat. 2008. The multiple lives of NMD factors: balancing roles in gene and genome regulation. Nat. Rev. Genet. 9: 699-712.

Jakubowski, H. 1980. Valyl-tRNA synthetase form yellow lupin seeds: hydrolysis of the enzyme-bound noncognate aminoacyl adenylate as a possible mechanism of increasing specificity of the aminoacyl-tRNA synthetase. Biochemistry 19: 5071-5508.

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.

Khazaie, K., J. H. Buchanan, and R. F. Rosenberger. 1984. The accuracy of Q beta RNA translation. 1. Errors during the synthesis of Q beta proteins by intact *Escherichia coli* cells. Eur. J. Biochem. 144: 485-489.

Kiktev, D., S. Moskalenko, O. Murina, A. Baudin-Baillieu, J. P. Rousset, and G. Zhouravleva. 2009. The paradox of viable sup45 STOP mutations: a necessary equilibrium between translational readthrough, activity and stability of the protein. Mol. Genet. Genomics 282: 83-96.

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.

Korencic, D., I. Ahel, J. Schelert, M. Sacher, B. Ruan, C. Stathopoulos, P. Blum, M. Ibba, and D. Söll. 2004. A freestanding proofreading domain is required for protein synthesis quality control in Archaea. Proc. Natl. Acad. Sci. USA 101: 10260-10265.

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.

Kramer, E. B., H. Vallabhaneni, L. M. Mayer, and P. J. Farabaugh. 2010. A comprehensive analysis

of translational missense errors in the yeast *Saccharomyces cerevisiae*. RNA 16: 1797-1808.

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 (Bethesda) 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.

Laughrea, M., J. Latulippe, A. M. Filion, and L. Boulet. 1987. Mistranslation in twelve *Escherichia coli* ribosomal proteins. Cysteine misincorporation at neutral amino acid residues other than tryptophan. Eur. J. Biochem. 169: 59-64.

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.

Li, W., and M. Lynch. 2020. Universally high transcript error rates in bacteria. eLife 9: e54898.

Li, W., S. Baehr, M. Marasco, L. Reyes, D. Brister, C. S. Pikaard, J.-F. Gout, M. Vermulst, and M. Lynch. 2022. Evolution of transcript-error rates. (in prep.)

Lirussi, L., Ö. Demir, P. You, A. Sarno, R. E. Amaro, and H. Nilsen. 2021. RNA metabolism guided by RNA modifications: the role of SMUG1 in rRNA quality control. Biomolecules 11:

76.

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.

Loenarz, C., R. Sekirnik, A. Thalhammer, W. Ge, E. Spivakovsky, M. M. Mackeen, M. A. McDonough, M. E. Cockman, B. M. Kessler, P. J. Ratcliffe, et al. 2014. Hydroxylation of the eukaryotic ribosomal decoding center affects translational accuracy. Proc. Natl. Acad. Sci. USA 111: 4019-4024.

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.

Lovmar, M., and M. Ehrenberg. 2006. Rate, accuracy and cost of ribosomes in bacterial cells. Biochimie 88: 951-961.

Lue, S. W., and S. O. Kelley. 2007. A single residue in leucyl-tRNA synthetase affecting amino acid specificity and tRNA aminoacylation. Biochemistry 46: 4466-4472.

Luo, J., and B. D. Hall. 2007. A multistep process gave rise to RNA polymerase IV of land plants. J. Mol. Evol. 64: 101-112.

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.

Manickam, N., N. Nag, A. Abbasi, K. Patel, and P. J. Farabaugh. 2014. Studies of translational misreading *in vivo* show that the ribosome very efficiently discriminates against most potential errors. RNA 20: 9-15.

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: 3136-3138.

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.

Meer, K. M., P. G. Nelson, K. Xiong, and J. Masel. 2020. High transcriptional error rates vary as a function of gene expression level. Genome Biol. Evol. 12: 3754-3761.

Melnikov, S. V., K. D. Rivera, D. Ostapenko, A. Makarenko, N. D. Sanscrainte, J. J. Becnel, M. J. Solomon, C. Texier, D. J. Pappin, and D. Söll. 2018. Error-prone protein synthesis in parasites with the smallest eukaryotic genome. Proc. Natl. Acad. Sci. USA 115: E6245-E6253.

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.

Miller, J. H., and A. M. Albertini. 1983. Effects of surrounding sequence on the suppression of nonsense codons. J. Mol. Biol. 164: 59-71.

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.

Mordret, E., O. Dahan, O. Asraf, R. Rak, A. Yehonadav, G. D. Barnabas, J. Cox, T. Geiger, A. B. Lindner, and Y. Pilpel. 2019. Systematic detection of amino acid substitutions in proteomes reveals mechanistic basis of ribosome errors and selection for translation fidelity. Mol. Cell 75: 427-441.e5.

Mori, N., D. Mizuno, and S. Goto. 1979. Conservation of ribosomal fidelity during ageing. Mech. Ageing Dev. 10: 379-398.

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 hyperther-

mophilic 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., J. M. Goodenbour, A. David, K. A. Dittmar, R. B. Jones, J. R. Schneider, D. Boone, E. M. Eves, M. R. Rosner, J. S. Gibbs, et al. 2009. Innate immune and chemically triggered oxidative stress modifies translational fidelity. Nature 462: 522-526.

Nguyen, D. T., B. Wu, H. Long, N. Zhang, C. Patterson, S. Simpson, K. Morris, W. K. Thomas, M. Lynch, and W. Hao. 2020. Variable spontaneous mutation and loss of heterozygosity among heterozygous genomes in yeast. Mol. Biol. Evol. 37: 3118-3130.

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.

Phillips, R., J. Kondev, J. Theriot, and H. Garcia. 2013. Physical Biology of the Cell, 2nd Ed. Garland Science, New York, NY.

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.

Plant, E. P., P. Nguyen, J. R. Russ, Y. R. Pittman, T. Nguyen, J. T. Quesinberry, T. G. Kinzy, and J. D. Dinman. 2007. Differentiating between near- and non-cognate codons in *Saccharomyces cerevisiae*. PLoS One 2: e517.

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.

Qian, H. 2006. Reducing intrinsic biochemical noise in cells and its thermodynamic limit. J. Mol. Biol. 362: 387-392.

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.

Rock, J. M., U. F. Lang, M. R. Chase, C. B. Ford, E. R. Gerrick, R. Gawande, M. Coscolla, S. Gagneux, S. M. Fortune, and M. H. Lamers. 2015. DNA replication fidelity in *Mycobacterium* tuberculosis is mediated by an ancestral prokaryotic proofreader. Nat. Genet. 47: 677-681.

Roth, J. R. 1970. UGA nonsense mutations in *Salmonella typhimurium.* J. Bacteriol. 102: 467-475.

Salas-Marco, J., and D. M. Bedwell. 2005. Discrimination between defects in elongation fidelity and termination efficiency provides mechanistic insights into translational readthrough. J. Mol. Biol. 348: 801-815.

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.

Sarkar, A., M. Thoms, C. Barrio-Garcia, E. Thomson, D. Flemming, R. Beckmann, and E. Hurt. 2017. Preribosomes escaping from the nucleus are caught during translation by cytoplasmic quality control. Nat. Struct. Mol. Biol. 24: 1107-1115.

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.

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 microbiome *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.

Stansfield, I., K. M. Jones, P. Herbert, A. Lewendon, W. V. Shaw, and M. F. Tuite. 1998. Missense translation errors in *Saccharomyces cerevisiae.* J. Mol. Biol. 282: 13-24.

Steiner, R. E., A. M. Kyle, and M. Ibba. 2019. Oxidation of phenylalanyl-tRNA synthetase positively regulates translational quality control. Proc. Natl. Acad. Sci. USA 116: 10058-10063.

SternJohn, J., S. Hati, P. G. Siliciano, and K. Musier-Forsyth. 2007. Restoring species-specific

posttransfer editing activity to a synthetase with a defunct editing domain. Proc. Natl. Acad. Sci. USA 104: 2127-2132.

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

Strigini, P., and E. Brickman. 1973. Analysis of specific misreading in *Escherichia coli*. J. Mol. Biol. 75: 659-672.

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

Suzuki, Y., D. Ishihara, M. Sasaki, H. Nakagawa, H. Hata, T. Tsunoda, M. Watanabe, T. Komatsu, T. Ota, T. Isogai, 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.

Tang, Y., and D. A. Tirrell. 2002. Attenuation of the editing activity of the *Escherichia coli* leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins *in vivo*. Biochemistry 41: 10635-10645.

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

Torabi, N., and L. Kruglyak. 2011. Variants in SUP45 and TRM10 underlie natural variation in translation termination efficiency in *Saccharomyces cerevisiae*. PLoS Genet. 7: e1002211.

Toth, M. J., E. J. Murgola, and P. Schimmel. 1988. Evidence for a unique first position codon-anticodon mismatch in vivo. J. Mol. Biol. 201: 451-454.

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.

Tsui, W. C., and A. R. Fersht. 1981. Probing the principles of amino acid selection using the alanyl-tRNA synthetase from *Escherichia coli*. Nucleic Acids Res. 9: 4627-4637.

Turrientes, M. C., F. Baquero, B. R. Levin, J.-L. Martínez, A. Ripoll, J.-M. González-Alba, R. Tobes, M. Manrique, M.-R. Baquero, M.-J. Rodríguez-Domínguez, 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.

Vargas-Rodriguez, O., A. H. Badran, K. S. Hoffman, M. Chen, A. Crnković, Y. Ding, J. R. Krieger, E. Westhof, D. Söll, and S. Melnikov. 2021. Bacterial translation machinery for deliberate mistranslation of the genetic code. Proc. Natl. Acad. Sci. USA 118: e2110797118.

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.

Wei, W., W.-C. Ho, M. Behringer, and M. Lynch. 2022. Rapid evolution of mutation rate and spectrum in response to environmental and population-genetic challenges. Nature Comm. 13: 4752.

Weickert, M. J., and I. Apostol. 1998. High-fidelity translation of recombinant human hemoglobin in *Escherichia coli*. Appl. Environ. Microbiol. 64: 1589-1593.

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., J. E. Barrick, O. Tenaillon, M. J. Wiser, W. J. Dittmar, S. Cruveiller, B. Chane-Woon-Ming, C. Médigue, R. E. Lenski, and D. Schneider. 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, I., J. Richardson, A. Starkey, and I. Stansfield. 2004. Genome-wide prediction of stop codon readthrough during translation in the yeast *Saccharomyces cerevisiae*. Nucleic Acids Res. 32: 6605-6616.

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.

Wilusz, J. E., J. M. Whipple, E. M. Phizicky, and P. A. Sharp. 2011. tRNAs marked with CCACCA are targeted for degradation. Science 334: 817-821.

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, H., Z. Lyu, Y. Fan, C. R. Evans, K. W. Barber, K. Banerjee, O. A. Igoshin, J. Rinehart, and J. Ling. 2020. Metabolic stress promotes stop-codon readthrough and phenotypic heterogeneity. Proc. Natl. Acad. Sci. USA 117: 22167-22172.

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.

Zhang, Z., B. Shah, and P. V. Bondarenko. 2013. G/U and certain wobble position mismatches as possible main causes of amino acid misincorporations. Biochemistry 52: 8165-8176.

Zhang, Z., L. Zhou, L. Hu, Y. Zhu, H. Xu, Y. Liu, X. Chen, X. Yi, X. Kong, and L. D. Hurst. 2010. Nonsense-mediated decay targets have multiple sequence-related features that can inhibit translation. Mol. Syst. Biol. 6: 442.

Figure 20.1. The multitude of functions that must be successfully navigated for the production of a properly translated and folded protein.

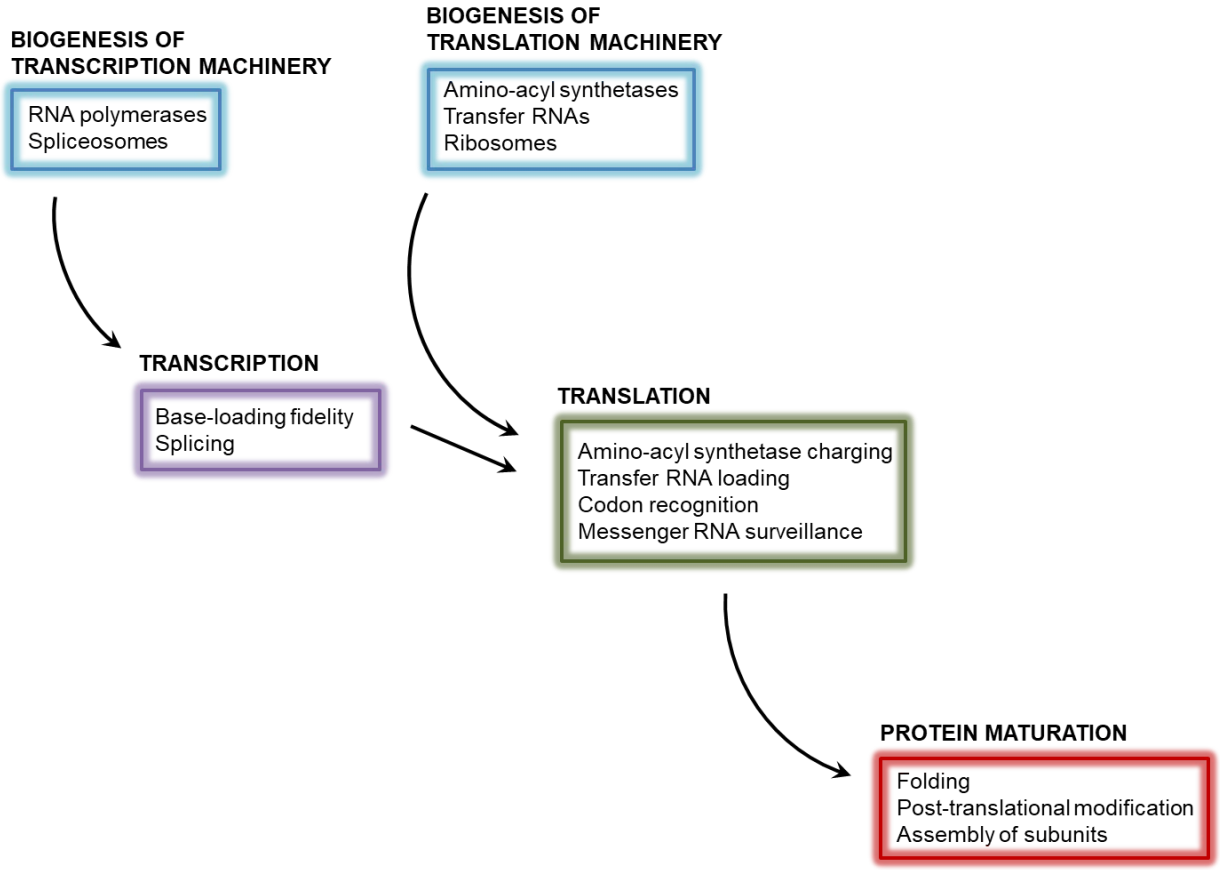


Figure 20.2. The distribution of transcript-error rates across the Tree of Life. **Left**) In comparison to mutation rates at the DNA level (per generation), transcript-error rates are orders of magnitude higher and relatively constant. Diagonal dashed lines denote where data would lie with three different levels of inflation. **Right**) Negative 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, the effective number of codons expected to be under selection, and the degree of expression of individual errors, as described in the text). Only the data for unicellular species are used in the regression, and the line shown denotes the best fit, which is obtained when $x = 1.0$, i.e., additivity of effects (inset shows how the fit varies with x). Data are from Li et al. (2022).

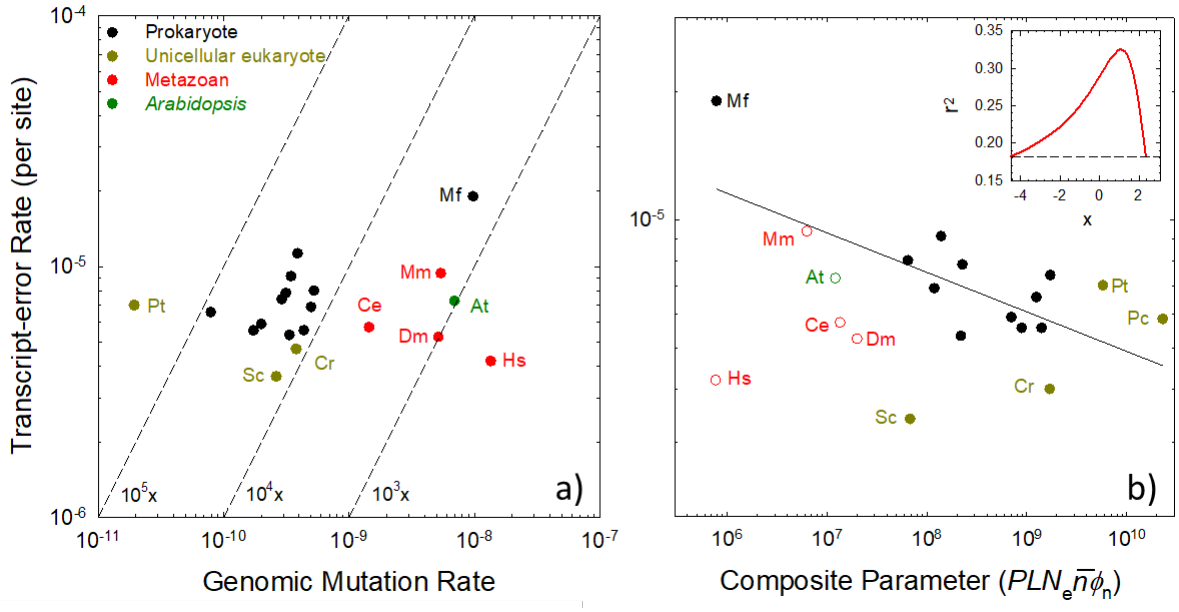
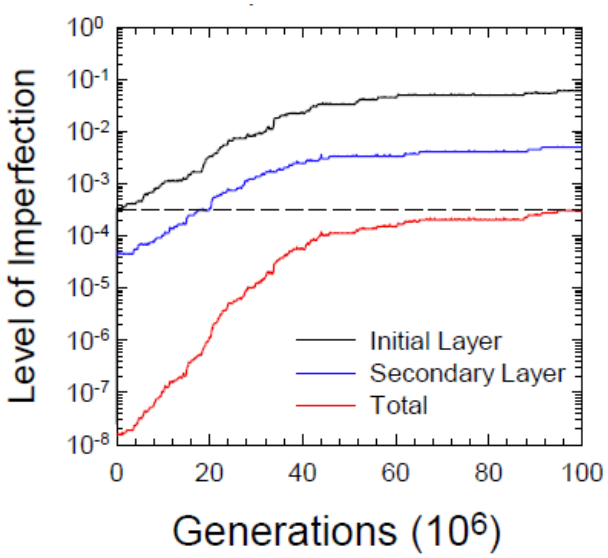


Figure 20.4. Reaction steps involving two competing Michaelis-Menten reactions involving right (R) and wrong (W) substrates. E denotes an enzyme. EX an enzyme-substrate complex, and P_X a product generated by action on substrate X. The coefficients (rates) on the arrows are defined in Foundations 20.1.



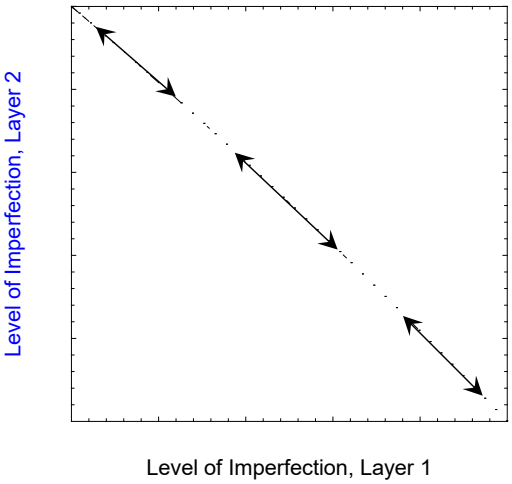


Figure 20.4. Reaction steps involving two competing Michaelis-Menten reactions involving right (R) and wrong (W) substrates. E denotes an enzyme. EX an enzyme-substrate complex, and P_X a product generated by action on substrate X. The coefficients (rates) on the arrows are defined in Foundations 20.1.

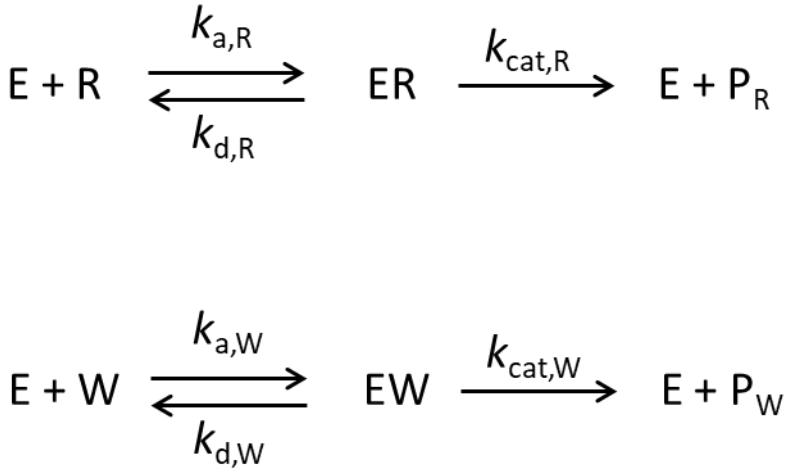


Figure 20.5. Modified reaction dynamics (compared to Figure 20.4) for the situation in which there is proofreading of an intermediate complex (denoted by asterisks), again for two competing substrates, R and W.