Published in final edited form as:
Nat Rev Genet. 2009 October ; 10(10): 715–724. doi:10.1038/nrg2662.

The evolutionary consequences of erroneous protein synthesis

D. Allan Drummond[1] and **Claus O. Wilke**[2]

[1]FAS Center for Systems Biology, Harvard University.

[2]Center for Computational Biology and Bioinformatics, Institute for Cell and Molecular Biology, and Section of Integrative Biology, The University of Texas at Austin.

Abstract

Errors in protein synthesis disrupt cellular fitness, cause disease phenotypes, and shape gene and genome evolution. Experimental and theoretical results on this topic have accumulated rapidly in disparate fields such as neurobiology, protein biosynthesis and degradation, and molecular evolution, yet with limited communication between disciplines. Here, we review studies of error frequencies, their cellular and organismal consequences, and attendant long-range evolutionary responses. Measurements of error frequencies, from transcription through protein folding, remain in their infancy; we emphasize major areas where little is known, such as the failure rate of protein folding, or where technological innovations may enable imminent gains, such as translational missense error frequencies. Evolutionary responses to errors fall into two broad categories: adaptations that minimize errors and their attendant costs, and adaptations which exploit errors for the organism's benefit. Given this wide spectrum of effects, it may be more useful to refer to synthesis outcomes as beneficial and deleterious rather than correct and erroneous.

Synthesis of a functional protein from genetic information is strikingly error-prone. For example, amino-acid misincorporations during translation are estimated to occur once in every 1,000 to 10,000 codons translated[1,2]. At this error rate, 15% of average-length protein molecules will contain at least one misincorporated amino acid. Polypeptide errors can induce protein misfolding, aggregation, and cell death (e.g. Ref. 3). Misfolded proteins underlie a broad array of neurogenerative diseases, and misincorporation of amino acids during translation may be a causative factor in the pathology of multiple sclerosis and ALS[4,5]. Conversely, global defects in protein synthesis produce tissue-specific neurodegeneration linked to production of misfolded proteins[3,6].

We define erroneous protein synthesis as any disruption in the conversion of a coding sequence into a functioning protein. Besides amino-acid misincorporations, sources of errors are transcription errors, aberrant splicing, premature termination, faulty posttranslational modifications, and kinetic missteps during folding (Figure 1). This definition explicitly includes correctly synthesized polypeptides that fail to fold into a functional protein.

We have previously hypothesized that major patterns of coding sequence evolution, conserved from bacteria to humans, arise from the selective pressure to minimize the cost of erroneous protein synthesis, including the failure of properly synthesized polypeptides to fold[5]. Such selection would act most strongly on highly expressed genes and, in animals, on genes expressed in neural tissues. Mathematical modeling and computer simulations predict biophysical adaptations that reduce this cost[5,7–9], and several of these predictions have now been verified in a recent experimental evolution study[10].

Together, these studies illuminate a pathway leading from the fidelity of protein production through cellular dysfunction and organismal fitness defects—exemplified by

neurodegeneration—to adaptations whose imprints are visible in the evolution of coding sequences across taxa.

Here, we first review what is known about the frequencies of errors in the production of functional proteins, from transcription to protein folding. We do not attempt a comprehensive review of all measurements. Instead, we aim to create perspective and to motivate much-needed future studies by highlighting the diverse set of approaches taken. We then review the many ways in which organisms may have evolved to cope with errors in synthesis, either by selectively reducing error rates or by evolving tolerance to errors. Next, we examine how organisms exploit errors in synthesis to achieve biological and evolutionary ends that are inaccessible when synthesis is error-free. We conclude with a discussion of implications for future research.

Erroneous protein synthesis

Errors arise at all steps of protein synthesis, from transcription to protein folding, and have widespread phenotypic consequences. Yet surprisingly little is known about the exact error rates and error spectra.

Error rates in protein synthesis

The science of measuring error rates associated with protein synthesis remains in its infancy, even though the first attempts go back more than 45 years (e.g. Ref. 11). For example, the literature contains experimental measurements for the frequency of less than 5% of the 1,216 (64×19) possible codon-to-amino-acid errors in translation, with only a handful of estimates from the same species. Recent studies have made substantial progress on measuring error rates in specific cases (see e.g. Ref. 12), but current technological developments will likely soon give us the first comprehensive view of translation error frequencies in normal cells (Box 1).

Table 1 provides estimates of error rates from transcription through protein folding, emphasizing the heterogeneous experimental approaches used and the patchy knowledge that has resulted. The central observation is that synthesis errors are orders of magnitude more frequent than DNA-replication errors. The *E. coli* genome is 4.6×10^6 base-pairs long, such that at the typical mutation rate of approximately 10^{-9} per base pair, one bacterium in 200 will bear a mutation in its genome. By contrast, the average *E. coli* coding sequence is 335 codons long, and at a canonical per-codon missense error rate of 5×10^{-4}, 15% of protein molecules will contain at least one error. At the bacterial scale, perfectly replicated genomes are commonplace, but perfectly synthesized proteomes never occur. The available evidence suggests that eukaryotes are no more or less accurate at protein synthesis than are prokaryotes[13]. All else equal, longer proteins necessarily accumulate more errors, leading to astonishing predictions: if canonical missense error rates hold, each molecule of the giant human muscle protein titin, consisting of 34,350 amino acids, would contain an average of 17 missense errors, and an average human sarcomere would contain no error-free titin molecules at all.

Errors in posttranslational modification are likely important but their frequency and effects remain largely unknown. One of the most common modifications, glycosylation, is performed on more than 50% of proteins in a human cell[14]. Glycosylation is not template-driven and shows remarkable heterogeneity[15]. Oligosaccharides attached to glycosylation sites tend to vary from copy to copy of the same protein, and occupancy rates of glycosylation sites also vary, making it unclear to what extent heterogeneity in glycosylation should be considered erroneous. That not all heterogeneity is functionally normal is demonstrated by the often highly deleterious effects of glycosylation-altering mutations, which usually affect the efficiency of glycosylation or the composition of glycans without disrupting glycosylation altogether[16]. The extent and importance of misphosporylation also remains poorly understood despite potentially

major consequences. For example, misphosphorylation of the microtubule-binding protein tau is a pathological signature of all cases of Alzheimer's disease and apparently contributes to tau misfolding and aggregation[17].

Perhaps surprisingly, we know even less about the error rate in producing functional proteins. An early study reported that up to 30% of newly synthesized proteins were rapidly degraded, most of which were believed to be defective ribosomal products (DRiPs)[18]. Yet a later study using similar techniques found that most newly synthesized proteins were largely protected from degradation, even when unable to fold correctly due to misincorporation errors, and "at most a few percent" of newly synthesized proteins were rapidly degraded[19]. Thus, the failure rate of functional protein production, the ultimate expression of failures in protein synthesis, remains essentially unknown. Correspondingly, the proportion of those failures due to upstream synthesis errors versus errors in folding of correctly synthesized proteins also remains unclear.

Deleterious effects of synthesis errors

Plentiful evidence demonstrates that errors in protein synthesis reduce organism fitness: disruption of translational fidelity with common antibiotics such as streptomycin and kanamycin kills bacteria; cells with impaired translational proofreading ability display altered morphologies[20] and suffer severe fitness defects[21], as do cells with elevated rates of transcription errors in an essential gene[10]; defects in translational fidelity and in protein folding cause disease phenotypes in mouse models[3,6].

A single amino-acid substitution in the editing domain of an alanyl-tRNA synthetase—a mutation which causes misacylation, subsequent widespread translation errors, and protein misfolding—causes degeneration of Purkinje cells in the mouse cerebellum, ataxia, and death[3]. This result supports the possibility that disease conditions involving tissue-specific dysfunction arise from global errors in protein synthesis[20]. Neurons may be unusually sensitive to synthesis errors because of their long lifetimes, large surface-area-to-volume ratios with correspondingly abundant sites for membrane-induced aggregation[22], branched morphologies which impede transport and damage responses[23], fluctuating cell polarization, and protein quality control systems more likely to be overloaded by misfolded proteins[17] (cf. Refs. 24, [25]).

Fitness costs can arise by multiple different mechanisms. Protein synthesis errors will often lead to *loss of function* of the protein. A recent study demonstrated that disruption of folding and function of the antibiotic-resistance protein β-lactamase by transcriptional errors reduced cellular fitness, but could be compensated by increased expression and by stabilizing mutations in the protein sequence[10].

Protein synthesis errors may also produce polypeptides displaying a *gain of toxic function*. In rare cases, the error may confer an alternate or pathological function on an otherwise normal, folded protein. More often, errors disrupt folding, and the misfolded molecule may be toxic. In this context, "toxic" simply means harmful and does not specify the modality or severity of the harm. Misfolded proteins may destabilize membranes[26], steal quality-control bandwidth from essential proteins[24,25], and induce chronic stress. The toxic effects of aminoglycoside antibiotics, which befoul ribosomes and lead to production of misfolded proteins, have been traced in part to misfolded-protein-induced signaling through the membrane receptor *cpxA*. The ultimate consequence is increased radical formation, membrane depolarization, and cell death[27]. Misfolded protein cytotoxicity has been studied extensively as a contributor to neurodegenerative disease. It has become increasingly clear that at the molecular level, misfolding-associated disease phenotypes often reflect gains of toxic function rather than losses of function[3,17,22,23,25,26,28].

Synthesis and degradation of non-functional proteins may also be costly without being obviously harmful (*clean-up costs*, see e.g. Ref. 29). Ribosomal throughput dedicated to a polypeptide that will ultimately fail to function represents an opportunity cost, particularly for fast-growing organisms[30]. Expression of quality control systems, such as chaperones, to assist, rescue, or degrade polypeptides represents a further fitness cost acting in *trans*. Toxicity and clean-up costs may coexist: even if quality control systems ultimately detect and either degrade or refold all misfolded proteins, the latter may still wreak substantial toxic havoc, just as crime does not cease to be a problem even if all criminals are eventually caught.

Errors in the proteins responsible for reproduction of genetic and non-genetic material, particularly in translation and replication, may lead to reduced fidelity and subsequent dysfunction in succeeding generations. Such an effect, originally conceived as an *error catastrophe* by Orgel[31], has been demonstrated in bacteria, where heritable mutations can arise from an editing defect in translation[21,32].

Effect of gene expression level

To the extent that protein-synthesis errors produce harmful molecular species or waste valuable cellular resources, the severity of the resulting phenotypic effects will depend on the expression level of that gene. The more highly expressed a gene, the larger the amount of erroneously synthesized proteins produced, and thus the bigger the influence of these proteins on the organism's phenotype. For example, the clean-up costs due to synthesis of non-functional protein will be proportional to the amount of protein produced. Many forms of misfolded protein toxicity, such as aggregation and interference with membranes, increase with absolute protein concentration and therefore with gene expression level. Note that if synthesis errors primarily act by reducing protein function, an effect from gene expression level is not expected; errors in a low-expression, but functionally critical, protein such as a DNA polymerase or transcription factor need not contribute any less to organism fitness than disruption of the activity of a high-abundance enzyme or structural molecule[5,7].

Adaptations for cost minimization

Faced with costly protein-synthesis errors, organisms may evolve two high-level cost-reduction strategies: reduction of error frequencies (increased accuracy), and reduction of the costs of the remaining errors (increased tolerance or robustness). Because costs tend to increase with gene expression level, selection for cost reduction is often visible in differences between genes of low and high expression level.

Reduction of error frequencies

The primary source of missense substitutions during protein synthesis is misincorporation of non-cognate tRNAs during translation. Codons corresponding to low-abundance tRNAs tend to be more error-prone than other codons[12]. Consequently, codon usage affects translation error frequencies. Selection pressure to use codons with low error rates is commonly referred to as selection for *translational accuracy* 33. Accuracy selection should not, however, cause uniform usage of accurate codons along the gene. Instead, it should disproportionately affect those sites at which translation errors would have particularly severe effects on protein folding or function[33]. A common test for translational accuracy selection therefore assesses whether preferred codons associate with evolutionarily conserved sites[5,33,34] or with sites that are known to be important for protein structure or function[33,35]. In general, these analyses show a moderate but highly significant tendency for preferred codons to coincide with sites at which translation errors are expected to be important, consistent with weak selection for increased translational accuracy.

Although changes in codon usage to improve accuracy come at little cost, ribosomes may also be made more accurate. However, increased ribosomal accuracy comes often at the cost of translation speed and energy efficiency[36], due in part to the intrinsic physical implementation of increased accuracy through increased energy-dependent rejection of tRNAs. Consequently, organisms may evolve to balance ribosome speed, ribosome accuracy, and energetic costs.

A second codon-level selection pressure penalizes codons that have a high probability of being mistranslated into radically different amino acids. We refer to this selection pressure as selection for *error mitigation*. While not leading to a reduction of error frequencies *per se*, this selection pressure reduces the frequency of the most costly errors at the expense of a larger number of more benign errors. Several bioinformatics studies have found evidence for selection for error mitigation[37–39]. The genetic code itself also has error-mitigating properties[40], and may have evolved specifically to minimize the effects of translation errors[41].

It is likely that selection limits error frequencies at all steps of protein synthesis. Some simple predictions have yet to be tested, such as whether high-expression genes have lower transcriptional error rates. But aside from accuracy selection and error mitigation, little is known about the signatures that would indicate such selection pressures. One exception is the efficiency of splicing in fission yeast, as estimated by the proportion of intron-exon junctions retained in cellular mRNAs. It increases markedly with gene expression level[42], presumably because missplicing becomes more costly when incorrectly spliced mRNAs are abundant.

Increased tolerance or robustness

Errors need not be eliminated altogether if instead organisms can tolerate a certain amount of errors without paying a significant fitness cost (Figure 2). Some tolerance is inherent in protein biochemistry. *In vitro*, proteins can be robust to many individual or multiple mutations[43–48], although most mutations tend to reduce protein stability. Robustness can itself be modulated by mutations in the protein[47,48]. These observations suggest that proteins can evolve robustness to typical errors arising under translation, termed *translational robustness*[7,8]. Proteins that possess translational robustness can fold and function properly even if mistranslated. Mathematical and computational modeling predicts that this selection pressure will cause proteins to be more thermostable and to also be more tolerant to genetic mutations[5,7–9]. Recent experimental results confirm these predictions[10].

But even if a gene is translated without any errors, the resulting protein may misfold, because of interactions with other proteins (e.g., other misfolded or aggregated proteins) or properties of the protein itself. Key among protein properties are thermodynamic stability, measured by the free energy of unfolding, and folding kinetics, measured by the rate of folding or unfolding. For most proteins, thermodynamics dictate whether a protein can ever attain a stable folded state, whereas kinetics determine how likely a thermodynamically stable protein is to complete folding before other processes, such as aggregation and degradation, derail it. Rapid folding and high stability tend to be correlated. We have previously hypothesized that selection reduces the propensity of proteins to misfold even when translated without errors[5,7], but this hypothesis has not yet been tested experimentally. Because of the close relationship between thermodynamic stability and tolerance to mutations[47–50], more translationally robust proteins may also be more kinetically stable and vice versa. A key difficulty at present is distinguishing stochastic misfolding from mistranslation-induced misfolding, as translation errors remain difficult to detect. Consistent with either translational-robustness selection or selection against stochastic misfolding is the observation that highly expressed genes are less aggregation-prone than genes of low expression level[51–53].

Other adaptations beside robust protein folding may reduce the cost of synthesis errors. One is the efficient detection and degradation of mis-spliced products. The nonsense-mediated

decay pathway degrades mRNAs that contain premature stop codons[54]. The introns of eukaryotes tend to either contain stop codons or alter the translational reading frame to reveal a downstream stop codon, leading to mRNA degradation of mis-spliced transcripts by the nonsense-mediated decay pathway[55].

Genome-wide signatures of cost reduction

Broad patterns of coding-sequence evolution, such as the tendency for highly expressed proteins to evolve slowly, may reflect selection to reduce costs of protein misfolding[5]. Genome-wide analyses of evolutionary rates have consistently found that expression level is a major predictor of both synonymous and non-synonymous divergence in bacteria, fungi, plants, and animals[5,56]. Multivariate analyses find that quantities related to translation frequency make stronger contributions to evolutionary rate than do quantities linked primarily to gene function[57–61]. We have hypothesized that selection against protein misfolding, including misfolding of error-free polypeptides, imposes a strong constraint on coding-sequence evolution[5]. Many genomic patterns—covariation between evolutionary rates, expression level, codon-usage bias, and the transition–transversion ratio, as well as an association between optimal codons and evolutionarily conserved sites—can be reproduced in a model involving only selection against mistranslation-induced misfolding[5]. New genome-wide signals are needed to allow disentangling of selection pressures against costs of error-free versus error-induced protein misfolding.

In all extant models, we do not expect a one-to-one relationship between gene expression level and evolutionary conservation. Selection acts only on the deleterious outcomes of erroneous synthesis, which may vary from protein to protein. For example, protein alleles that are less likely to become toxic, or are more rapidly detected and shuttled toward degradation or refolding, should experience less evolutionary constraint. This pressure, and the resulting constraints on sequence change, should intensify with increasing expression level or for genes expressed in sensitive tissue types. Likewise, if a particular protein fold is highly tolerant to synthesis errors, then genes encoding proteins of this fold will experience little selection pressure to reduce costs, even if expressed at a relatively high level. By contrast, sensitive folds will experience much stronger selection pressure at comparable expression levels. Consistent with this reasoning, biophysical properties of the protein fold also influence the rate of sequence divergence[62–65], and the relative contributions of expression level and protein structure to evolutionary conservation seem to be of comparable magnitude[66].

Beneficial synthesis errors

Even though errors in protein synthesis tend to be deleterious on average, in numerous cases they can have direct benefits for organism fitness.

Error-dependent protein expression

A wide array of organisms, from viruses to mammals, have evolved certain genes that depend on errors in protein synthesis. The best-known example is programmed frameshift, where the elongating ribosome shifts forward (+1) or back (−1) by a single nucleotide to enter a new reading frame[67]. *E. coli* DNA polymerase III subunits τ and γ, and eukaryotic ornithine decarboxylase antizyme, depend on frameshifting for proper protein expression[68,69].

Programmed frameshifts can control gene expression (Figure 3A)[70]. Ornithine decarboxylase antizyme (OAZ) noncompetitively inhibits ornithine decarboxylase (ODC), an enzyme which catalyzes the first step in polyamine synthesis. Polyamines such as spermidine stimulate +1 frameshifting. In eukaryotes from fission yeast to mammals, the OAZ gene normally terminates at an early stop codon, yielding only a short peptide with no inhibitory activity, but a +1

frameshift yields full-length antizyme which inhibits ODC. At low polyamine levels, frameshifting occurs infrequently and little antizyme is produced, more ODC is active, and more polyamines accumulate[70]. As polyamine levels rise, frameshifting is stimulated, yielding more full-length antizyme which inhibits ODC and reduces polyamine production[70]. Thus, the polyamine-controlled frequency of a translation "error" has evolved to implement feedback regulation of polyamine levels.

In baker's yeast, gene expression is partly regulated by splicing efficiency[71]. Under amino-acid starvation, splicing is inhibited for the majority of intron-containing ribosomal genes[71]. The consequence of this regulation is likely a reduction in the number of functional ribosomes, and thus inhibition of translation. Some bacteriophages use spontaneous readthrough of stop codons in specific contexts to regulate gene expression of phage proteins[72].

Certain picornaviruses carry, within a long polypeptide, a short sequence (~19 amino-acids) that induces eukaryotic ribosomes to skip a peptide bond[73]. This skip-inducing *2A sequence* allows these viruses to encode multiple proteins using a single, compact sequence without paying the price of encoding a protease. Such ribosome skipping, in essence a bug in the translational hardware uncovered and exploited by viruses, is now being coopted by human biological engineers[74].

Suppression of deleterious mutations

Synthesis errors can suppress otherwise deleterious mutations, such as reading through the stop codon in an important gene. Such nonsense suppression had been long studied in bacterial and cell-culture systems[75]. Recently, it has taken on increased importance as a therapy for genetic diseases in which a premature stop-codon mutation causes a disease phenotype, such as in cystic fibrosis and Duchenne muscular dystrophy[76].

Drugs that interfere with translational fidelity in bacteria are commonly used as antibiotics. Bacterial mutants that depend on streptomycin for viability are readily isolated[77] and tend to have hyperaccurate but slow ribosomes[36]. Streptomycin independence is often regained by mutations that decrease ribosomal fidelity[78].

Exploration of alternative or novel phenotypes

Synthesis errors can reveal cryptic genetic variation or produce novel phenotypes and thus allow an organism to either switch epigenetically between different phenotypes or pre-screen potentially beneficial mutations[79, 80].

The yeast prion [*PSI*$^+$] is an amyloid-forming conformation of the translation-termination factor Sup35p. It sequesters Sup35p and impairs translation termination (Figure 3B). The [*PSI*$^+$]-state is self-propagating[81], arising and resolving spontaneously with low probability[82]. In many environmental conditions [*PSI*$^+$] strains grow better than strains without the prion[83], due to genetic variation revealed upon readthrough of stop codons[84]. The prion domain of Sup35p is evolutionarily conserved and seems to have a beneficial effect over evolutionary timescales[85]. It may have adaptive value unrelated to [*PSI*$^+$]. But the [*PSI*$^+$] state alone is sufficient to maintain the prion domain if environmental conditions under which [*PSI*$^+$] reveals adaptive genetic variation are encountered at least once every few million generations[86,87].

More generally, organisms can take advantage of beneficial phenotypes generated by errors[79,80]. A mutation which increases the production of the erroneous but beneficial product will increase in frequency in the population. As a consequence, organisms may derive a direct benefit from mutations that increase the likelihood of additional beneficial mutations in the future. This mechanism has been termed the *look-ahead effect*[80]. A variant of the look-ahead

mechanism may be the cause of *adaptive mutability*, i.e., an increase in beneficial mutations under conditions of environmental stress, observed in *E. coli*[88]. One of the experiments reported in Ref. 88required a reversion of an inactivating frameshift in the *lac* operon. But even in the absence of the reversion mutation, rare accidental frameshifts during translation provide residual function to the inactivated gene; duplication of the inactivated gene leads to a commensurate increase in fitness derived from the residual function; at the same time, it increases the probability that a mutation corrects the frameshift[89].

Implications for future research

Our understanding of the fidelity of transcription, translation, and protein folding remains sketchy (Box 2). No comprehensive, or even representative, error spectra exist for cells under normal physiological conditions. Technological innovations such as single-molecule nucleic acid sequencing have given us a surprising portrait of rampant splicing errors in a eukaryotic genome[42], and this technology in combination with deep-coverage quantitative mass spectrometry[90] may soon provide a similar breakthrough in our understanding of transcriptional and translational error spectra (see Box 1). However, the frequency and types of errors in common posttranslational modifications such as glycosylation and phosphorylation remain almost completely unknown, as do the consequences of these errors for protein folding and function. Moreover, the relative fitness costs of loss of protein function, quality control, and gain of toxic function remain unknown, and considerable effort will be required to determine these as well (Box 2). Yet whatever the results of such studies, the existing evidence shows that protein synthesis is surprisingly error-prone, and that erroneous protein synthesis can differentially affect specific tissue types, impose substantial cellular fitness costs, and modulate the evolution of whole genomes.

In stark contrast with the rarity of DNA replication errors, the extraordinary frequency of protein synthesis errors in normal cells urges a different, perhaps unfamiliar, view of cellular operations. Cells are inherently noisy statistical ensembles, and the genotype is best understood as encoding the frequency of different outcomes rather than a single so-called *correct* state that is disrupted by errors. Notions of correct and erroneous may be subsumed by the more useful notions of beneficial and deleterious, with the important difference that supposed errors may be beneficial, even essential. For example, programmed +1 frameshifts and translational hops seem to have evolved by amplification of low-frequency translation errors[67].

Recent single-molecule studies underscore the need to embrace the extraordinary molecular diversity arising from a single genotype. In fission yeast, the frequency of retained introns appears to exceed 90% for the vast majority of transcripts[42]. Are all these retained introns technical artifacts, errors whose deleterious effects are too small to be eliminated by natural selection, errors in transcripts destined for degradation by nonsense-mediated decay[55], or an uneasy compromise resulting from energetic or kinetic costs associated with increased splicing fidelity? Or do some of these retained introns confer important benefits on the organism which would be suppressed by higher-fidelity splicing? Similarly, for some high-expression proteins, certain mistranslation-generated, biochemically similar molecular species are expected to exist at cellular abudances of 10–100 molecules per cell, sufficient for action as regulatory proteins. It seems unlikely that nature always fails to exploit the existence of these molecular subspecies, but they are difficult to hunt down; perhaps high-expression genes which change expression markedly in cells with hyperaccurate ribosomes may point to autoregulatory systems maintained by mistranslation. We believe that erroneous synthesis with its attendant modifiers and resulting adaptations, far from being a negligible nuisance, will play a central role in our understanding of molecular evolution.

Measuring translational error rates

Translation is the most error-prone step of protein synthesis. Therefore, accurate measurements of amino-acid misincorporation rates are crucial for a thorough understanding of synthesis errors. We can write all possible missense errors in the form of a 64×19 matrix with 1,216 independent entries. To date, only a small percentage of these entries has been measured, and only in a handful of organisms. The challenge in measuring missense error rates is that in a given sample, the abundance of error-free molecules is several orders of magnitude higher than that of any species of error-containing molecules, overloading most unbiased detection methods and forcing investigators to employ clever, but strongly biased, schemes to obtain any result at all.

Historical methods used to measure translational error rates fall into three broad categories. First, some groups have measured the amount of a specific amino acid in a protein that should not contain this amino acid. For example, Edelmann and Gallant measured the amount of cysteine in the normally cysteine-free protein flagellin of *E. coli*[91]. Second, some groups have measured the change in a protein's isoelectric point due to amino-acid misincorporation[92,93]. Both of these approaches share the drawback that they average over many different elements in the ribosomal error matrix. A third approach builds on special reporter systems that produce a signal when a specific codon is mistranslated. For example, Kramer and Farabaugh studied misincorporation of lysine at various codons using fused luciferases, F-luc and R-luc, whose luminescence can be determined independently and with extreme accuracy. In F-luc, they replaced the codon for the essential lysine at position 529 by all near-cognate and several other codons[12,94]. With these constructs, they measured the frequency of mistranslation of specific codons into lysine by assaying the F-luc activity relative to R-luc activity.

Could an estimate of the entire 64×19 error matrix be obtained in a single experiment? In principle, yes. Massive gains in the sensitivity of quantitative tandem mass spectrometry (MS/MS) (e.g. Ref. 90) offer the tantalizing potential for detecting low-frequency errors against a background of wild-type molecules. Deep quantitative MS/MS probing of peptides generated from a purified target protein or proteins, using a detection database including all possible single amino-acid substitutions as well as the DNA-encoded sequence, could in principle detect both the type and position of amino-acid substitutions introduced by mistranscription and mistranslation. By encoding the target protein(s) with multiple instances of all 64 codons, each codon's error spectrum could be estimated in multiple contexts, and single-molecule RNA sequencing the target gene's transcripts could be used to assess the frequency and position of transcription errors, allowing translation errors due to misacylation and misreading to be disentangled. While such an experiment is technically demanding, it is within the reach of present-day methods and would, at a stroke, provide the first comprehensive view of the translation error spectrum in any organism.

Open questions

- What are the exact error rates for transcription, splicing, translation, and postranslational modifications? How do these rates vary between genes, and covary with key variables such as gene expression levels?
- What is the failure rate of protein folding? What proportion of folding failures result from upstream synthesis errors versus stochastic and/or *trans*-acting factors?
- What is the genome-wide distribution of posttranslational modifications (PTMs) such as glycosylation and phosphorylation? What proportion of PTM events are deleterious under normal conditions?

- What are the main mechanisms by which protein synthesis errors produce fitness costs?
- How large are the fitness costs associated with erroneous protein synthesis compared to other, unrelated fitness costs?
- Why are error rates so high even though the associated fitness costs seem substantial?
- How important are synthesis errors for the proper functioning of the cellular machinery?
- How else have organisms evolved to exploit the molecular diversity present in their cells due to errors in protein synthesis?
- What do signatures of natural selection against the consequences of protein synthesis errors reveal about the spectrum of human neurodegenerative diseases, particularly their prevalence, severity, and cellular manifestations?

At a glance

- Protein synthesis is a complex, multi-stage process, and many things can go wrong along the way.
- At present, our knowledge of protein-synthesis error rates is limited. But existing measurements indicate that protein synthesis is very error-prone in comparison to DNA replication.
- A protein that hasn't been synthesized correctly may be non-functional or toxic. However, it may also have a new, beneficial function.
- The synthesis of non-functional and toxic proteins imposes fitness costs on the organism. These costs generally increase with the expression level of a gene.
- Many mechanisms seem to have evolved to minimize costs of erroneous protein synthesis.
- In some cases, organisms also take advantage of synthesis errors. For example, programmed frameshifts are sometimes used for expression regulation.
- Cellular life is an inherently noisy process. Every single gene produces a spectrum of different protein variants, and organisms optimize and take advantage of the properties of the entire spectrum.

Acknowledgments

This work was funded in part by NIH grants P50 GM068763 and R01 GM088344.

References

1. Parker J. Errors and alternatives in reading the universal genetic code. Microbiol. Reviews 1989;53:273–298.
2. Ogle JM, Ramakrishnan V. Structural insights into translational fidelity. Annu. Rev. Biochem 2005;74:129–177. [PubMed: 15952884]
3. Lee JW, et al. Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. Nature 2006;443:50–55. [PubMed: 16906134] Global mistranslation-induced protein misfolding leads to cell-type-specific neurodegeneration in a mouse. A exciting study that links the fidelity of translation to a disease phenotype.

4. Rubinstein E. Misincorporation of the proline analog azetidine-2-carboxylic acid in the pathogenesis of multiple sclerosis: a hypothesis. J. Neuropathol. Exp. Neurol 2008;67:1032–1034. [PubMed: 18978540]

5. Drummond DA, Wilke CO. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell 2008;134:341–352. [PubMed: 18662548] A bioinformatics and modeling study demonstrating that genome-wide patterns of molecular evolution are shared from bacteria to mammals. The study also showed that in a simple model of protein evolution selection against mistranslation-induced protein misfolding is sufficient to reproduce these patterns.

6. Zhao L, Longo-Guess C, Harris BS, Lee JW, Ackerman SL. Protein accumulation and neurodegeneration in the woozy mutant mouse is caused by disruption of SIL1, a cochaperone of BiP. Nature Genet 2005;37:974–979. [PubMed: 16116427]

7. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. Why highly expressed proteins evolve slowly. Proc. Natl. Acad. Sci. USA 2005;102:14338–14343. [PubMed: 16176987]

8. Wilke CO, Drummond DA. Population genetics of translational robustness. Genetics 2006;173:473–481. [PubMed: 16489231]

9. Willensdorfer M, Bürger R, Nowak MA. Phenotypic mutation rates and the abundance of abnormal proteins in yeast. PLoS Comput. Biol 2007;3:e203. [PubMed: 18039025]

10. Goldsmith M, Tawfik DS. Potential role of phenotypic mutations in the evolution of protein expression and stability. Proc. Natl. Acad. Sci. USA 2009;106:6197–6202. [PubMed: 19339491] One of the first studies to investigate the effect of transcription errors on protein evolution in an experimental system. TEM-1 β-lactamase expressed under an errror-prone RNA polymerase evolved increased expression level, increased thermostability, and increased mutational robustness.

11. Loftfield RB. The frequency of errors in protein synthesis. Biochem. J 1963;89:82–92. [PubMed: 14097371]

12. Kramer EB, Farabaugh PJ. The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. RNA 2007;13:87–96. [PubMed: 17095544] A highly accurate measurement of specific amino-acid misincorporation frequencies under translation.

13. Stansfield I, Jones KM, Herbert P, Shaw ALWV, Tuite MF. Missense translation errors in *Saccharomyces cerevisiae*. J. Mol. Biol 1998;282:13–24. [PubMed: 9733638]

14. Wong C-H. Protein glycosylation: new challenges and opportunities. J. Org. Chem 2005;70:4219–4225. [PubMed: 15903293]

15. Mahal LK. Glycomics: Towards bioinformatic approaches to understanding glycosylation. Anti-Cancer Agents in Medicinal Chemistry 2008;8:37–51. [PubMed: 18220504]

16. Freeze HH. Genetic defects in the human glycome. Nature Reviews Genetics 2006;7:537–551.

17. Winklhofer K, Tatzelt J, Haass C. The two faces of protein misfolding: gain- and loss-of-function in neurodegenerative diseases. The EMBO Journal 2008;27:336–349. [PubMed: 18216876]

18. Schubert U, et al. Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. Nature 2000;404:770–774. [PubMed: 10783891]

19. Vabulas R, Hartl F. Protein synthesis upon acute nutrient restriction relies on proteasome function. Science 2005;310:1960–1963. [PubMed: 16373576]

20. Nangle L, Motta C, Schimmel P. Global effects of mistranslation from an editing defect in mammalian cells. Chemistry and Biology 2006;13:1091–1100. [PubMed: 17052613]

21. Bacher JM, de Crécy-Lagard V, Schimmel PR. Inhibited cell growth and protein functional changes from an editing-defective tRNA synthetase. Proc. Natl. Acad. Sci. USA 2005;102:1697–1701. [PubMed: 15647356]

22. Stefani M. Generic cell dysfunction in neurodegenerative disorders: Role of surfaces in early protein misfolding, aggregation, and aggregate cytotoxicity. Neuroscientist 2007;13:519–531. [PubMed: 17901260]

23. Malgaroli A, Vallar L, Zimarino V. Protein homeostasis in neurons and its pathological alterations. Current Opinion in Neurobiology 2006;16:270–274. [PubMed: 16713705]

24. Gidalevitz T, Ben-Zvi A, Ho K, Brignull H, Morimoto RI. Progressive disruption of cellular protein folding in models of polyglutamine diseases. Science 2006;311:1471–1474. [PubMed: 16469881] This study creates insight into the nature of cellular costs due to protein misfolding, demonstrating

how aggregation-prone proteins can induce cytotoxic effects by destabilizing marginally stable essential proteins.

25. Gidalevitz T, Krupinski T, Garcia S, Morimoto RI. Destabilizing protein polymorphisms in the genetic background direct phenotypic expression of mutant SOD1 toxicity. PLoS Genetics 2009;5:e1000399. [PubMed: 19266020]
26. Stefani M, Dobson CM. Protein aggregation and aggregate toxicity: New insights into protein folding, misfolding diseases and biological evolution. J. Mol. Med 2003;81:678–699. [PubMed: 12942175]
27. Kohanski MA, Dwyer DJ, Wierzbowski J, Cottarel G, Collins JJ. Mistranslation of membrane proteins and two-component system activation trigger antibiotic-mediated cell death. Cell 2008;135:679–690. [PubMed: 19013277]
28. Bucciantini M, et al. Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. Nature 2002;416:507–511. [PubMed: 11932737] One of the first demonstrations that misfolded proteins can be cytotoxic. Aggregates of the N-terminal domain of the *E. coli* HypF protein reduce the viability of mouse fibroblasts in a concentration-dependent manner.
29. Bürger R, Willensdorfer M, Nowak MA. Why are phenotypic mutation rates much higher than genotypic mutation rates? Genetics 2006;172:197–206. [PubMed: 16143614]
30. Stoebel D, Dean A, Dykhuizen D. The cost of gene expression of *E. coli lac* operon proteins is in the process, not in the products. Genetics 2008;178:1653–1660. [PubMed: 18245823]
31. Orgel L. The maintenance of the accuracy of protein synthesis and its relevance to ageing. Proc. Natl. Acad. Sci. USA 1963;49:517–521. [PubMed: 13940312]
32. Bacher JM, Schimmel P. An editing-defective aminoacyl-tRNA synthetase is mutagenic in aging bacteria via the SOS response. Proc. Natl. Acad. Sci. USA 2007;104:1907–1912. [PubMed: 17264207]
33. Akashi H. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. Genetics 1994;136:927–935. [PubMed: 8005445]
34. Stoletzki N, Eyre-Walker A. Synonymous codon usage in *Escherichia coli*: Selection for translational accuracy. Mol. Biol. Evol 2007;24:374–381. [PubMed: 17101719]
35. Zhou T, Weems M, Wilke CO. Translationally optimal codons associate with structurally sensitive sites in proteins. Mol. Biol. Evol 2009;26:1571–1580. [PubMed: 19349643]
36. Ruusala T, Andersson D, Ehrenberg M, Kurland CG. Hyper-accurate ribosomes inhibit growth. The EMBO Journal 1984;3:2575–2580. [PubMed: 6391914]
37. Archetti M. Selection on codon usage for error minimization at the protein level. J. Mol. Evol 2004;59:400–415. [PubMed: 15553093]
38. Archetti M. Genetic robustness and selection at the protein level for synonymous codons. J. Evol. Biol 2006;19:353–365. [PubMed: 16599911]
39. Higgs PG, Hao W, Golding GB. Identification of selective effects on highly expressed genes. Evolutionary Bioinformatics 2007;2:1–13.
40. Freeland SJ, Hurst LD. The genetic code is one in a million. J. Mol. Evol 1998;47:238–248. [PubMed: 9732450]
41. Woese CR. On the evolution of the genetic code. Proc. Natl. Acad. Sci. USA 1965;54:1546–1552. [PubMed: 5218910]
42. Wilhelm BT, et al. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature 2008;453:1239–1243. [PubMed: 18488015] A detailed study of the transcriptome of *Schizosaccharomyces pombe* under multiple conditions, using both high-throughput sequencing and tiling arrays. The study found widespread transcription of non-coding regions and frequent intron retention in mRNAs.
43. Loeb DD, et al. Complete mutagenesis of the HIV-1 protease. Nature 1989;340:397–400. [PubMed: 2666861]
44. Shafikhani S, Siegel RA, Ferrari E, Schnellenberger V. Generation of large libraries of random mutants in *Bacillus subtilis* by PCR-based plasmid multimerization. Biotechniques 1997;23:304–310. [PubMed: 9266088]
45. Daugherty PS, Chen G, Iverson BL, Georgiou G. Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies. Proc. Natl. Acad. Sci. USA 1999;97:2029–2034. [PubMed: 10688877]

46. Guo HH, Choe J, Loeb LA. Protein tolerance to random amino acid change. Proc. Natl. Acad. Sci. USA 2004;101:9205–9210. [PubMed: 15197260]
47. Bloom JD, et al. Thermodynamic prediction of protein neutrality. Proc. Natl. Acad. Sci. USA 2005;102:606–611. [PubMed: 15644440]
48. Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. Nature 2006;444:929–932. [PubMed: 17122770]
49. Taverna DM, Goldstein RA. Why are proteins so robust to site mutations? J. Mol. Biol 2002;315:479–484. [PubMed: 11786027]
50. Wilke CO, Bloom JD, Drummond DA, Raval A. Predicting the tolerance of proteins to random amino acid substitution. Biophys. J 2005;89:3714–3720. [PubMed: 16150971]
51. Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M. Life on the edge: A link between gene expression levels and aggregation rates of human proteins. Trends Biochem. Sci 2007;32:204–206. [PubMed: 17419062]
52. Vendruscolo M, Tartaglia GG. Towards quantitative predictions in cell biology using chemical properties of proteins. Mol. BioSyst 2008;4:1170–1175. [PubMed: 19396379]
53. Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M. A relationship between mRNA expression levels and protein solubility in *E. coli*. J. Mol. Biol 2009;388:381–389. [PubMed: 19281824]
54. McGlincy NJ, Smith CWJ. Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? Trends Biochem. Sci 2008;33:385–393. [PubMed: 18621535]
55. Jaillon O, et al. Translational control of intron splicing in eukaryotes. Nature 2008;451:359–362. [PubMed: 18202663] A study of the importance of the nonsense-mediated decay (NMD) pathway in limiting the amount of misspliced introns in eukaryotes. Many introns in eukaryotes contain stop codons that trigger the NMD pathway in case of missplicing.
56. Pál C, Papp B, Lercher M. An integrated view of protein evolution. Nature Reviews Genetics 2006;7:337–348.
57. Rocha EPC, Danchin A. An analysis of determinants of amino acids substitution rates in bacterial proteins. Mol. Biol. Evol 2004;21:108–116. [PubMed: 14595100]
58. Agrafioti I, et al. Comparative analysis of the *Saccharomyces cerevisiae* and *Caenorhabditis elegans* protein interaction networks. BMC Evol. Biol 2005;5:23. [PubMed: 15777474]
59. Wolf YI, Carmel L, Koonin EV. Unifying measures of gene function and evolution. Proc. R. Soc. B 2006;273:1507–1515.
60. Drummond DA, Raval A, Wilke CO. A single determinant dominates the rate of yeast protein evolution. Mol. Biol. Evol 2006;23:327–337. [PubMed: 16237209]
61. Xia Y, Franzosa EA, Gerstein MB. Integrated assessment of genomic correlates of protein evolutionary rate. PLoS Comp. Biol 2009;5:e1000413.
62. Bloom JD, Drummond DA, Arnold FH, Wilke CO. Structural determinants of the rate of protein evolution in yeast. Mol. Biol. Evol 2006;23:1751–1761. [PubMed: 16782762]
63. Hartling J, Kim J. Mutational robustness and geometrical form in protein structures. J. Exp. Zoology B: Mol. Devel. Evol 2007;310B:216–226.
64. Choi SC, Hobolth A, Robinson DM, Kishino H, Thorne JL. Quantifying the impact of protein tertiary structure on molecular evolution. Mol. Biol. Evol 2007;24:1769–1782. [PubMed: 17522088]
65. Zhou T, Drummond DA, Wilke CO. Contact density affects protein evolutionary rate from bacteria to animals. J. Mol. Evol 2008;66:395–404. [PubMed: 18379715]
66. Wolf MY, Wolf YI, Koonin EV. Comparable contributions of structural-functional constraints and expression level to the rate of protein sequence evolution. Biology Direct 2008;3:40. [PubMed: 18840284]
67. Farabaugh PJ. Programmed translational frameshifting. Ann. Rev. Genetics 1996;30:507–528. [PubMed: 8982463]
68. Blinkowa AL, Walker JR. Programmed ribosomal frameshifting generates the *Escherichia coli* DNA polymerase III gamma subunit from within the tau subunit reading frame. Nucleic Acids Res 1990;18:1725–1729. [PubMed: 2186364]
69. Matsufuji S, et al. Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. Cell 1995;80:51–60. [PubMed: 7813017]

70. Ivanov IP, Matsufuji S, Murakami Y, Gesteland RF, Atkin JF. Conservation of polyamine regulation by translational frameshifting from yeast to mammals. EMBO Journal 2000;19:1907–1917. [PubMed: 10775274]
71. Pleiss JA, Whitworth GB, Bergkessel M, Guthrie C. Rapid, transcript-specific changes in splicing in response to environmental stress. Mol. Cell 2007;27:928–937. [PubMed: 17889666]
72. Engelberg-Kulka H, Dekel L, Israeli-Reches M, Belfort M. The requirement of nonsense suppression for the development of several phages. Mol. Gen. Genet 1979;170:155–159. [PubMed: 372760]
73. Donnelly MLL, et al. Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. Journal of General Virology 2001;82:1013–1025. [PubMed: 11297676]
74. Funston GM, Kallioinen SE, de Felipe P, Ryan MD, Iggo RD. Expression of heterologous genes in oncolytic adenoviruses using picornaviral 2A sequences that trigger ribosome skipping. Journal of General Virology 2008;89:389–396. [PubMed: 18198369]
75. Gorini L. Informational suppression. Ann. Rev. Genetics 1970;4:107–134. [PubMed: 4950057]
76. Welch EM, et al. PTC124 targets genetic disorders caused by nonsense mutations. Nature 2007;447:87–91. [PubMed: 17450125]
77. Bjare U, Gorini L. Drug dependence reversed by a ribosomal ambiguity mutation, *ram*, in *Escherichia coli*. J. Mol. Biol 1971;57:423–435. [PubMed: 4931677]
78. Björkman J, Samuelsson P, Andersson D, Hughes D. Novel ribosomal mutations affecting translational accuracy, antibiotic resistance and virulence of *Salmonella typhimurium*. Molecular Microbiology 1999;31:53–58. [PubMed: 9987109]
79. Masel J. Cryptic genetic variation is enriched for potential adaptations. Genetics 2006;172:1985–1991. [PubMed: 16387877]
80. Whitehead DJ, Wilke CO, Vernazobres D, Bornberg-Bauer E. The look-ahead effect of phenotypic mutations. Biology Direct 2008;3:18. [PubMed: 18479505]
81. Wickner RB, Masison DC, Edskes HK. [*PSI*] and [*URE3*] as yeast prions. Yeast 1995;11:1671–1685. [PubMed: 8720070]
82. Chernoff YO, Newnam GP, Kumar J, Allen K, Zink AD. Evidence for a protein mutator in yeast: Role of the Hsp70-related chaperone Ssb in formation, stability, and toxicity of the [*PSI*] prion. Mol. Cell. Biol 1999;19:8103–8112. [PubMed: 10567536]
83. True HL, Lindquist SL. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. Nature 2000;407:477–483. [PubMed: 11028992] An influential paper demonstrating that the yeast prion [PSI^+] can uncover hidden genetic variation and produce new heritable phenotypes.
84. True HL, Berlin I, Lindquist SL. Epigenetic regulation of translation reveals hidden genetic variation to produce complex traits. Nature 2004;431:184–187. [PubMed: 15311209]
85. Jensen MA, True HL, Chernoff YO, Lindquist S. Molecular population genetics and evolution of a prion-like protein in *Saccharomyces cerevisiae*. Genetics 2001;159:527–535. [PubMed: 11606530]
86. King OD, Masel J. The evolution of bet-hedging adaptations to rare scenarios. Theor. Popul. Biol 2007;72:560–575. [PubMed: 17915273]
87. Griswold CK, Masel J. Complex adaptations can drive the evolution of the capacitor [PSI^+], even with realistic rates of yeast sex. PLoS Genet 2009;5:e1000517. [PubMed: 19521499]
88. Cairns J, Overbaugh J, Miller S. The origin of mutants. Nature 1988;335:142–145. [PubMed: 3045565]
89. Andersson DI, Slechta ES, Roth JR. Evidence that gene amplification underlies adaptive mutability of the bacterial *lac* operon. Science 1998;282:1133–1135. [PubMed: 9804552]
90. de Godoy LMF, et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature 2008;455:1251–1254. [PubMed: 18820680]
91. Edelmann P, Gallant J. Mistranslation in *E. coli*. Cell 1977;10:131–137. [PubMed: 138485]
92. Parker J, Friesen JD. "two out of three" codon reading leading to mistranslation in vivo. Mol. Gen. Genet 1980;177:439–445. [PubMed: 6768967]
93. Ellis N, Gallant J. An estimate of the global error frequency in translation. Mol Gen Genetics 1982;188:169–172.

94. Toth MJ, Murgola EJ, Schimmel P. Evidence for a unique first position codon-anticodon mismatch in vivo. J. Mol. Biol 1988;201:451–454. [PubMed: 3262166]

95. Kireeva ML, et al. Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. Mol. Cell 2008;30:557–566. [PubMed: 18538654]

96. Fox-Walsh KL, Hertel KJ. Splice-site pairing is an intrinsically high fidelity process. Proc. Natl. Acad. Sci. USA 2009;106:1766–1771. [PubMed: 19179398]

97. Daviter T, Gromadski K, Rodnina M. The ribosome's response to codon-anticodon mismatches. Biochimie 2006;88:1001–1011. [PubMed: 16716484]

98. Curran J, Yarus M. Base substitutions in the tRNA anticodon arm do not degrade the accuracy of reading frame maintenance. Proc. Natl. Acad. Sci. USA 1986;83:6538–6542. [PubMed: 2428035]

99. Jorgensen F, Kurland CG. Processivity errors of gene expression in *Escherichia coli*. J. Mol. Biol 1990;215:511–521. [PubMed: 2121997]

100. Arava Y, Boas F, Brown P, Herschlag D. Dissecting eukaryotic translation and its control by ribosome density mapping. Nucleic Acids Research 2005;33:2421–2432. [PubMed: 15860778]

Glossary

Clean-up costs, Any fitness costs related to the production and degradation of non-functional protein.; Gain of toxic function, Any event that causes a protein to generate a deleterious effect on the cell that expresses it. For example, a mutation that causes a protein to aggregate and become cytotoxic would be called a gain-of-toxic-function mutation.; Kinetic misfolding, Failure of an error-free protein to assume its proper ground-state conformation, or spontaneous loss of the ground-state conformation.; Look-ahead effect, The ability of organisms to sense the effect of potential future mutations if these mutations arise as errors under protein synthesis.; Programmed frameshift, A frameshift that is required for the proper expression of a specific functional protein. The frequency with which ribosomes change the reading frame at programmed-frameshift sites is often tightly regulated.; Ribosome skipping, A mechanism employed by certain picornaviruses to produce multiple peptides from a single open reading frame. A specific sequence (the 2A sequence) causes the translating ribosome to skip the formation of a peptide bond at the junction of the 2A sequence and the downstream sequence.; Stochastic misfolding, See *kinetic misfolding*; Selection for error mitigation, A selection pressure that causes genes or specific sites in genes to be encoded by codons that, when mistranslated, lead to the substitution of amino acids with limited deleterious effects.; Translational-accuracy selection, A selection pressure that causes genes or specific sites in genes to be encoded by high-fidelity codons, i.e., codons corresponding to highly abundant tRNAs.; Translational-robustness selection, A selection pressure that causes proteins to be tolerant to missense errors under translation. Translationally robust proteins fold and function even when mistranslated..

Biographies

D. Allan Drummond is a Bauer Fellow at Harvard University's FAS Center for Systems Biology. He obtained his Ph.D. in 2006 at The California Institute of Technology, USA. His research interests include protein quality control and molecular evolution.

Claus O. Wilke is an Assistant Professor at the The University of Texas at Austin in the Section of Integrative Biology. He is also affiliated with the Center of Computational Biology and Bioinformatics and the Institute of Cell and Molecular Biology at The University of Texas at Austin. Wilke obtained his Ph.D. in Theoretical Physics from the Ruhr-University Bochum, Germany in 1999, and was a postdoctoral scholar at The California Institute of Technology from 2000 to 2004. Wilke works on computational evolutionary biology, molecular evolution, and protein folding.

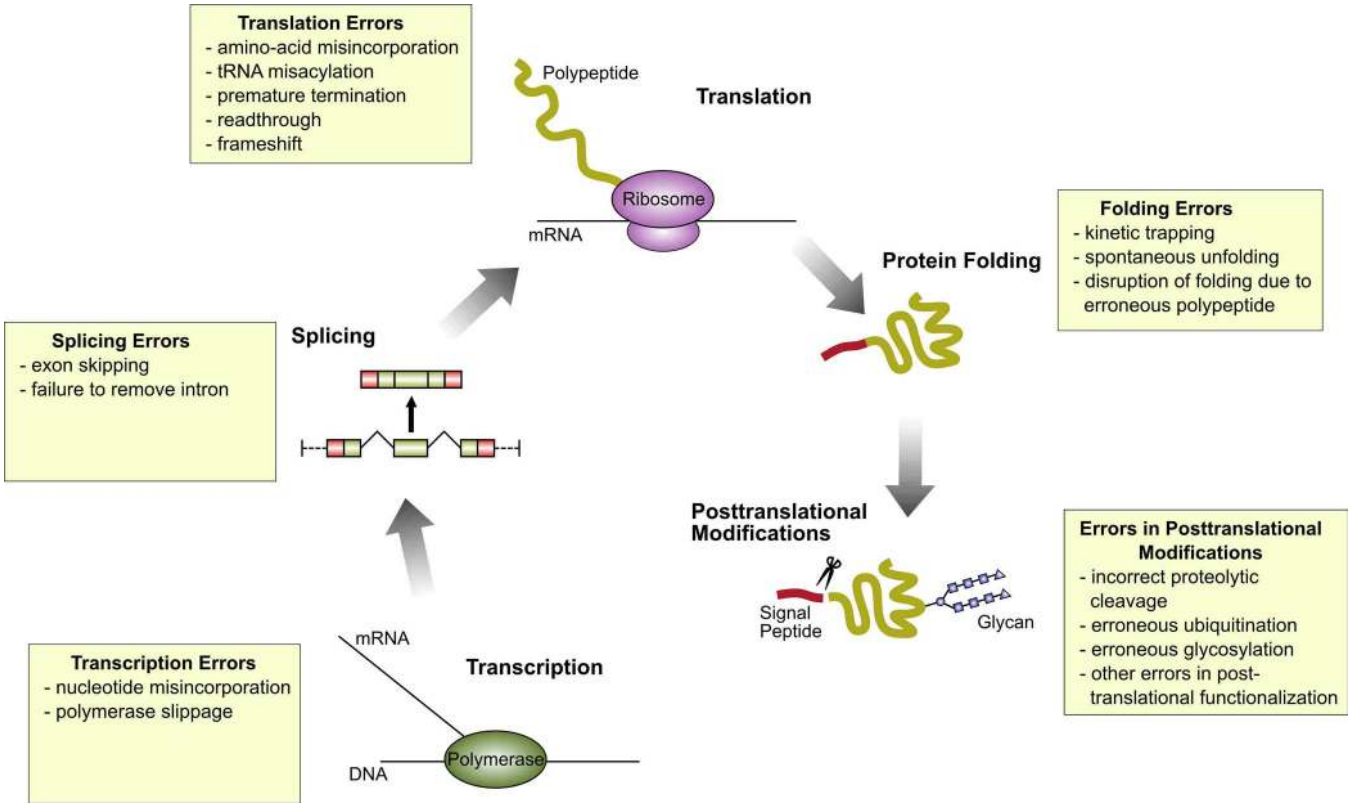

Figure 1.
Sources of errors in eukaryotic gene expression.

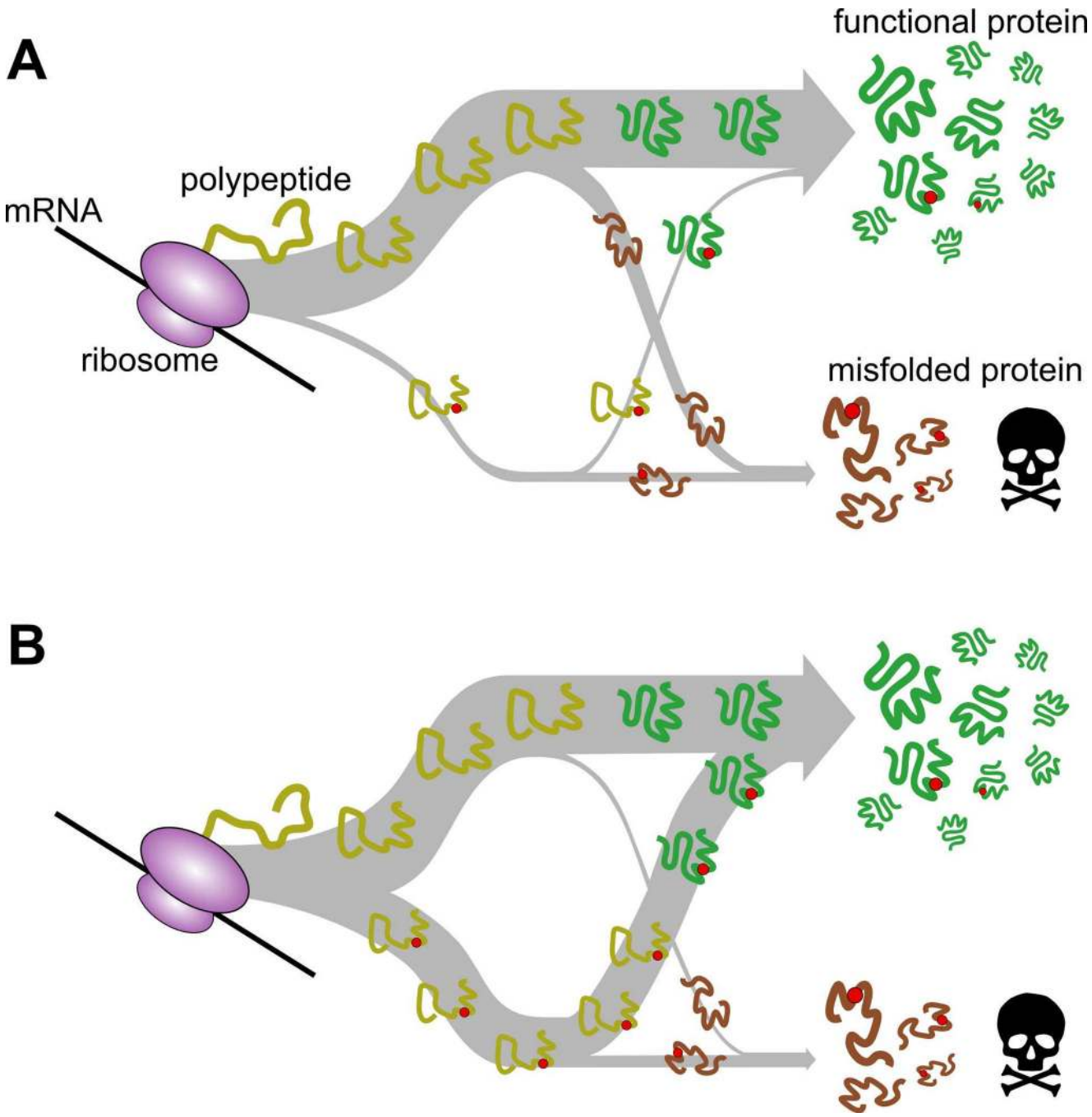

Figure 2.
Alternative strategies to reduce protein misfolding. (A) Proteins are poor folders and misfold readily. A highly accurate translational apparatus produces few proteins with translation errors and thus limits the total amount of misfolded protein. (B) An error-prone translation system produces many proteins with errors. But proteins are excellent folders and tend not to misfold even when mistranslated.

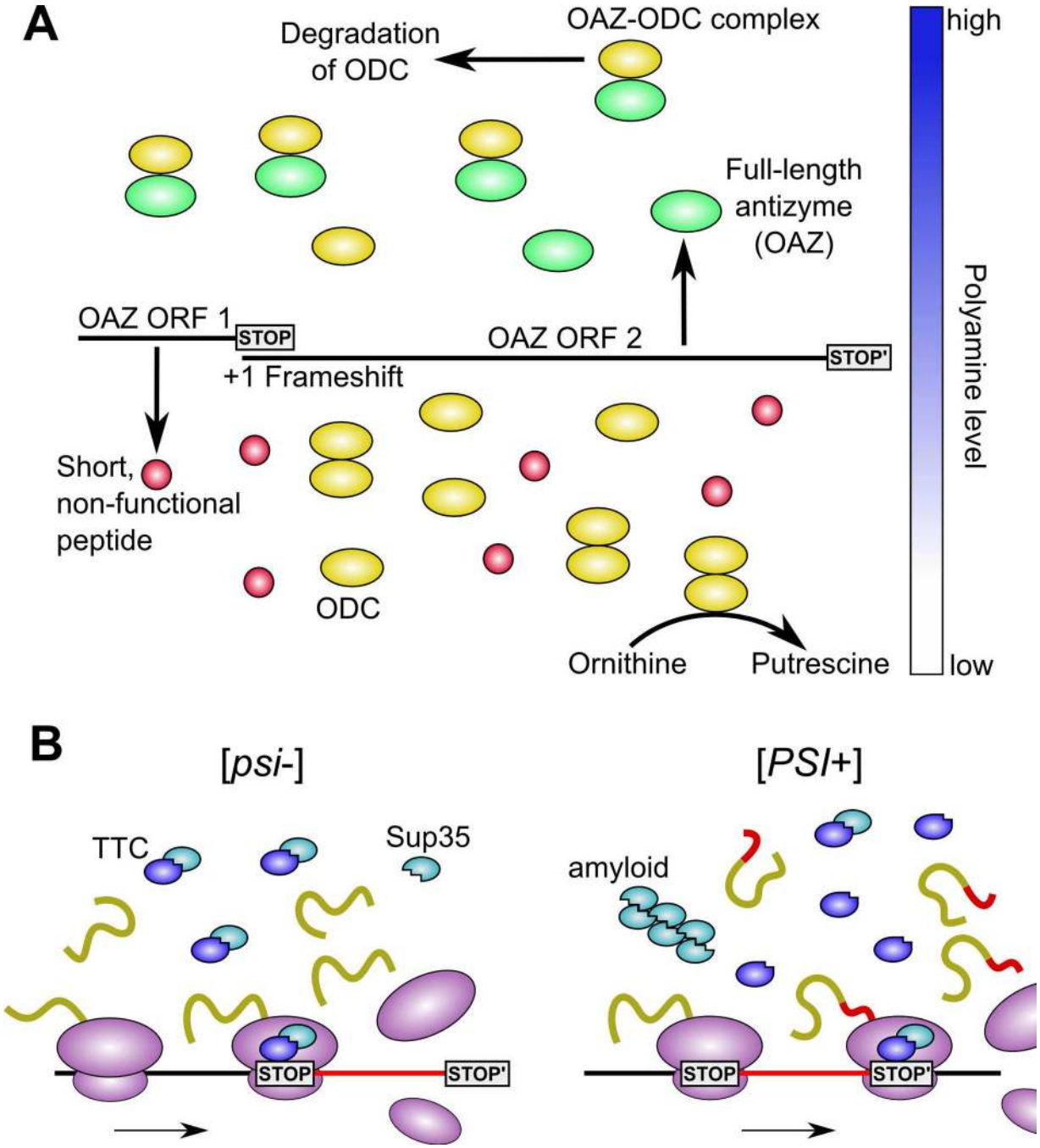

Figure 3.

Evolutionary exploitation of synthesis errors. (A) A frameshift regulates gene expression. Ornithine decarboxylase (ODC) catalyzes the synthesis of putrescine. ODC is inhibited by ornithine decarboxylase antizyme (OAZ), which binds to ODC and causes it to be degraded by the proteasome. Proper expression of functional (OAZ) requires a +1 frameshift. The frameshift occurs readily at high concentration of polyamines such as putrescine and its derivatives spermine and spermidine. (B) Translation of cryptic genetic variation. In the absence of the yeast prion ([*psi*−]), the protein Sup35p is readily available to form translation-termination complexes (TTCs). Consequently, stop codons are recognized reliably. In the presence of the yeast prion ([*PSI*+]), much Sup35p is sequestered, and there are too few TTCs

for reliable translation termination. As a consequence, cryptic genetic variation (indicated in red) is expressed.

Table 1

Representative average error rates at various steps of protein synthesis.

Error rate	Type of error	Basis of estimate	Organism	Reference
		Transcriptional misincorporation		
2×10^{-6}	Rate of UTP incorporation at G	Rate constants for polymerization and dissociation	*in vitro*	95
		Splicing errors		
5.7×10^{-6} to 2.3×10^{-2}	Proportion of retained vs. spliced introns	Quantitative PCR against exon-exon boundaries	*H. sapiens* (HeLa cells)	96
≈0.95	Proportion of retained introns	Quantitative sequencing of intron-exon boundaries compared to exon-exon boundaries	*S. pombe*	42
		Translational misincorporation		
6×10^{-5} to 2×10^{-4}	Arg CGU/C codons → Cys	Appearance of Cys in Cys-free flagellin	*E. coli*	91
3.4×10^{-4}	24 codons → Lys	Restoration of activity to inactive K529 mutants of firefly luciferase	*E. coli*	12
2×10^{-4}	Errors yielding positive charge changes in three proteins	2D gel quantitation	*E. coli*	93
5×10^{-6}	UAC → His	Restoration of activity to inactive H195Y mutant chloramphenicol acetyl transferase	*S. cerevisiae*	13
1×10^{-5} to 1×10^{-3}	Leu, Val, Ser codons → Phe	Kinetic modeling from experimentally determined rate constants	*in vitro*	97
		Frameshift		
1.5×10^{-5}	−1 frameshift at amber (UAG) stop codon	Restoration of *beta*-galactosidase activity to stop-containing variant	*E. coli*	98
3×10^{-5}	+1 frameshift at amber (UAG) stop codon	Restoration of *beta*-galactosidase activity to stop-containing variant	*E. coli*	98
		Premature termination		
2.7×10^{-4}	Premature termination rate per codon	0.76 success probability for completing 1,021-amino-acid β-galactosidase polypeptide	*E. coli*	99
0 to 2×10^{-3}	Premature termination rate per codon	Comparison of ribosome density on 5' vs. 3' of mRNA	*S. cerevisiae*	100
		Post-translational modification errors		
		Unknown (error-free state not well-defined)		
		Failure of protein folding		
		Unknown (no reliable estimates)		