REVIEW

Phenotypic mutations contribute to protein diversity and shape protein evolution

Maria Luisa Romero Romero[1,2] | **Cedric Landerer**[1,2] | **Jonas Poehls**[1,2] |
Agnes Toth-Petroczy[1,2,3]

[1]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

[2]Center for Systems Biology Dresden, Dresden, Germany

[3]Cluster of Excellence Physics of Life, TU Dresden, Dresden, Germany

Correspondence
Agnes Toth-Petroczy, Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany.
Email: toth-petroczy@mpi-cbg.de and tothpet@mpi-cbg.de

Funding information
Max-Planck-Gesellschaft

Review Editor: John Kuriyan

Abstract

Errors in DNA replication generate genetic mutations, while errors in transcription and translation lead to phenotypic mutations. Phenotypic mutations are orders of magnitude more frequent than genetic ones, yet they are less understood. Here, we review the types of phenotypic mutations, their quantifications, and their role in protein evolution and disease. The diversity generated by phenotypic mutation can facilitate adaptive evolution. Indeed, phenotypic mutations, such as ribosomal frameshift and stop codon readthrough, sometimes serve to regulate protein expression and function. Phenotypic mutations have often been linked to fitness decrease and diseases. Thus, understanding the protein heterogeneity and phenotypic diversity caused by phenotypic mutations will advance our understanding of protein evolution and have implications on human health and diseases.

KEYWORDS

frameshifts, protein evolution, transcriptional errors, translational errors

1 | INTRODUCTION

Evolution is limited by the accuracy of information transfer in the cell. The evolution of a highly accurate machinery for replication, transcription, and translation was crucial to the emergence of cellular life as we know it today.[1] However, the information transfer is not perfect, and every time DNA, RNA, or a protein is synthesized, errors may happen, resulting in mutations. Errors during transcription and translation lead to so-called "phenotypic mutations." Sources of phenotypic mutations include misincorporation of nucleotides or amino acids as well as RNA polymerase slippage, ribosomal slippage, premature termination, and stop codon readthrough (Figure 1a).

Phenotypic mutations are orders-of-magnitude more frequent than genetic mutations (10^{-2} to 10^5 vs. 10^{-6}–10^{-8} respectively).[2,3] In fact, protein synthesis from genetic information is so error-prone that ~15% of an average-length protein contains at least one mis-incorporated amino acid.[4] Thus, each gene corresponds to a population of proteins, also called "statistical proteins",[1] that differ from the canonical, error-free protein (Figure 1b).

"Biology is messy," as Dan Tawfik outlined in an intriguing concept article,[5] and wherever the cost of inaccuracy proves bearable, biological systems produce heterogeneity. This heterogeneity can facilitate physiological and evolutionary adaptation to new environments and challenges.[5] Phenotypic mutations explore a vast

mutational space from a single gene while maintaining the wild type sequence. Thus, Tawfik suggested that phenotypic mutations allow tinkering for novel functions, as illustrated by a seminal example of a frameshift that controls cytosolic and peroxisomal localization.[6]

Here, we first review the types and magnitude of phenotypic mutations that originate from transcriptional and translational errors. Then we discuss their evolutionary impact shaping protein robustness and their potential in adaptation, such as the evolution of new functions via frameshifts. Finally, we highlight how phenotypic mutations can alter phenotypes and review associations with human disease phenotypes and therapeutic applications.

2 | DETECTING PHENOTYPIC MUTATIONS

2.1 | The nature and magnitude of transcriptional errors

Transcriptional errors include nucleotide misincorporations in the mRNA and slippage of the RNA polymerase (RNAP), giving rise to single amino acid changes, long regions of altered amino acid sequences, and deletions (Figure 1a and Table 1).[7] Single nucleotide misincorporations occur >100-fold more frequently than DNA replication errors[8,9] and, since individual mRNAs can be translated up to 40 times, they significantly impact the downstream protein population.[10]

One of the earliest studies of the error rate of RNAP was presented by Springgate and Loeb in 1975. They purified *Escherichia coli* RNA polymerase and showed that during the *in vitro* transcription of synthetic polydeoxynucleotides, it inserted noncomplementary nucleotides.[11] This work described erroneous base-pairing during transcription as an infrequent and nonuniform event. Substitution of U to C was the most common error, occurring with a frequency of 4×10^{-4} error per base pair, two orders of magnitude less accurate than the fidelity of DNA replication in vitro (for *E. coli*, 2.5×10^{-6} error per base pair[12]). The outcome of such a transcriptional error rate would be that ∼5% of the *E. coli* proteome differ from the canonical sequence.

About a decade later, an assay measuring *in vivo* transcriptional error rate in *E. coli* revealed a rate of 1.4×10^{-4} errors per base. The assay relies on introducing nonsense mutations in lacZ and then assaying for residual β-galactosidase activity.[13] Functional mRNA LacZ will only be synthesized upon translational errors that revert the nonsense mutation.[14] Reporter-construct assays have long provided valuable information about RNA fidelity and transcriptional errors,[15–19] yet these studies have mainly focused on individual loci.

Only recently it became possible to study transcriptional errors globally in the transcriptome. Detecting incorrect transcripts by conventional next-generation sequencing is not possible because the reverse transcription itself is at least as error-prone as transcription itself. Barcoding individual RNA fragments before multiple rounds of the reverse transcription step overcame the obstacle and allowed the detection of natural transcripts at transcriptome-wide level. Application of this method to *Caenorhabditis elegans* showed a base misincorporation rate in mRNAs of 4×10^{-6}
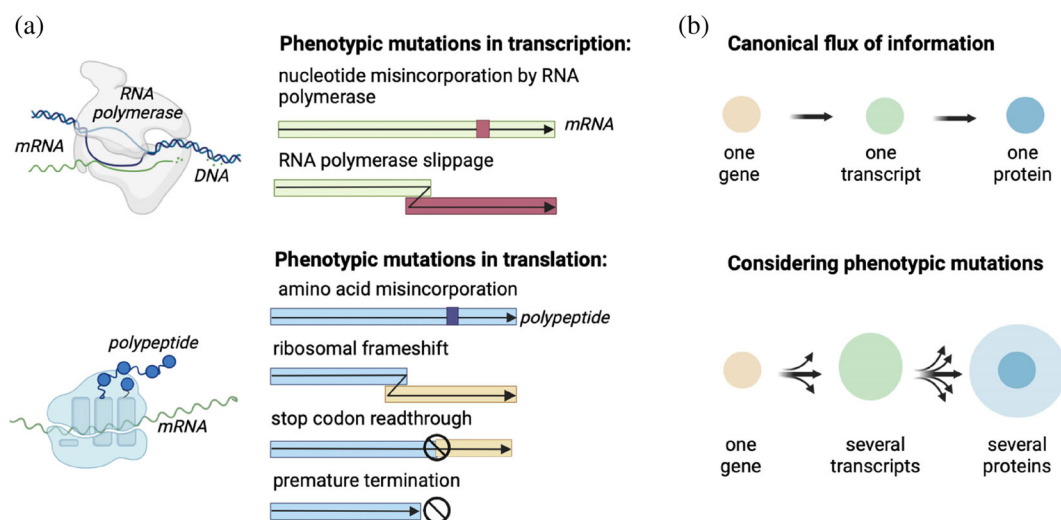

FIGURE 1 Phenotypic mutations contribute to protein diversity. (a) Transcriptional errors, such as nucleotide misincorporation and RNA polymerase slippage, and translational errors, such as amino acid misincorporation, ribosomal frameshift, stop codon readthrough, and premature termination, lead to protein sequence heterogeneity. (b) Phenotypic mutations can generate several transcripts and proteins from a single gene sequence

TABLE 1 Representative transcriptional error rates of several species

Nucleotide misincorporation in transcription		
Organism	Error rate (per nucleotide)	Method
Escherichia coli	4×10^{-4}	In vitro transcription of purified RNAP[11]
E. coli	1.4×10^{-4}	In vivo reporter-construct assay to detect mutations in the lacZ transcript[13]
E. coli	$1 - 2 \times 10^{-5}$	In vivo reporter-construct assay to detect mutations in the lacZ transcript[15]
E. coli	$\sim 10^{-6}$	In vivo reporter-construct assay to detect G to A errors[16]
E. coli	8.2×10^{-5}	In vivo transcriptome-wide detection by CirSeq[9]
rpb9Δ yeast	1.7×10^{-3}	In vivo reporter-construct assay to detect mutations in the lacZ transcript[17]
Yeast	2.8×10^{-4}	In vivo reporter-construct assay to detect mutations in the lacZ transcript[18]
C. elegans	4×10^{-6}	In vivo transcriptome-wide detection by NGS[20]
Yeast	3.9×10^{-6}	In vivo transcriptome-wide detection of RNAP II error by CirSeq[25]
Yeast	4.3×10^{-6}	In vivo transcriptome-wide detection of RNAP I errors by CirSeq[25]
Yeast	9.3×10^{-6}	In vivo transcriptome-wide detection of mitochondrial mRNA errors by CirSeq[25]
Yeast	1.7×10^{-5}	In vivo transcriptome-wide detection of RNAP III errors by CirSeq[25]
Yeast	4.2×10^{-5}	In vivo transcriptome-wide detection by CirSeq[132]
M. florum	1.8×10^{-5}	In vivo transcriptome-wide detection by CirSeq[24]
E. coli	5.8×10^{-6}	In vivo transcriptome-wide detection by CirSeq[24]
B. subtilis	5.8×10^{-6}	In vivo transcriptome-wide detection by CirSeq[24]
A. tumefaciens	7.3×10^{-6}	In vivo transcriptome-wide detection by NGS[24]
Buchnera aphidicola	4.7×10^{-5}	In vivo transcriptome-wide detection by CirSeq of a specie that has lost transcription fidelity factors[9]
Carsonella ruddii	5.1×10^{-5}	In vivo transcriptome-wide detection by CirSeq of a specie that has lost transcription fidelity factors[9]

errors per base,[20] about 10 times lower than previously reported for prokaryotes and unicellular eukaryotes (10^{-5}–10^{-3} errors per codon[21]). Another powerful method, Circular Sequencing (Cir-Seq), is based on circularized genomic RNA fragments to generate tandem repeats for next-generation sequencing to distinguish transcriptional errors from the noise produced by technical errors.[22,23] Different variants of this method have been applied to assess error rates, revealing that eukaryotes have a more accurate transcriptional machinery than prokaryotes in general. Yet, there are great differences in the transcriptional error rate even among prokaryotes (5.8×10^{-6} to 1.8×10^{-5} Table 1 and Figure 2).[24]

Transcriptional error rates vary among species,[20] within the same species, and even within the same organism.[25] In yeast, for example, the overall error rate was measured to be 3.9×10^{-6} errors per base, however different RNA polymerases have different error rates: mRNA synthesized by RNAPI had 4.3×10^{-6} errors per base, mRNA synthesized by mitochondrial RNAP had 9.3×10^{-6} errors per base, and RNA synthesized by RNAPIII had 1.7×10^{-5} errors per base.[25] Furthermore, nucleotide misincorporations differ in frequency depending on the sequence context; for example, *in vitro* transcripts synthesized by *E. coli* RNAP have a strong bias in errors toward AMP misincorporations instead of GMP.[26] Transcriptional error rates also differ when comparing nascent to total RNA due to the proof-reading machinery.[27] Despite these differences in the magnitude of the transcriptional error rate, all these studies agree that RNA polymerases do not operate as accurately as DNA polymerases (Figure 2a and Table 1). Lynch proposed that, since individual loci produce numerous transcripts, and mRNAs have a short lifespan, transcriptional errors likely have milder consequences than genomic errors. Thus, the lower selection pressure on the fidelity of the transcriptional machinery might be one of the reasons for the higher transcriptional error rates.[28]

2.2 | The nature and magnitude of translational errors

Estimating translational error rates, especially on a large scale, has proven challenging. A significant constraint to accurately measure translational errors is the lack of
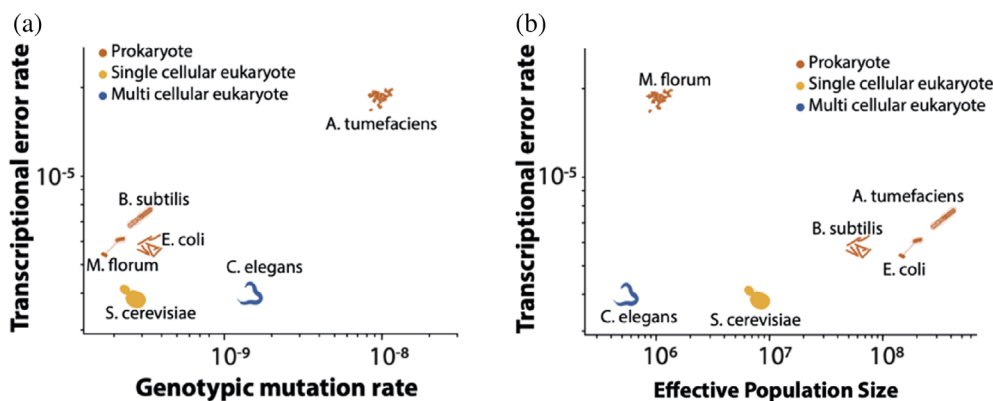
FIGURE 2 Transcription error rates vary across the tree of life. (a) Transcriptional error rates in the form of nucleotide misincorporations are orders of magnitude higher than genomic mutations across the tree of life (Table 1). (b) Increasing effective population size correlates with higher transcriptional error rates. Higher effective population sizes potentially increase the efficacy of selection for local error rate reduction. Effective population size estimates were taken from Sung et al.[143]

high-throughput methods to detect these rare and transient events proteome-wide. Thus, most of our knowledge regarding translational error rates comes from the design of reporter-construct assays of a single locus. Another challenge, once the protein synthesis error is identified, is to distinguish transcriptional from translational errors. Overcoming these challenges, many creative works, some of them listed below (Table 2), have shown that: (a) translational errors are orders of magnitude more frequent than genetic mutations (Figure 2a)[29]; (b) organisms have hotspot sequences, prone to translational recoding, in their adaptive toolbox.[30] For example, programmed ribosomal frameshifting (PRF)[6,31,32] and programmed stop codon readthrough[33] are used to encode novel functions, as we will review in the following chapters (Table 3).

Probably the first work that addressed translational errors, then called "*nongenetic errors*" in protein biosynthesis, dates from 1963. Combining radioactive measurements with chromatography techniques, amino acid misincorporations were found in ovalbumin as frequently as 3×10^{-4} errors per codon.[34] Later, engineering a reporter for lysine misincorporation in the active site of luciferase in *E. coli*, Kramer and Farabaugh estimated that error rates vary from 10^{-4} to 10^{-3} errors per codon and proposed that the frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition.[35]

The first methodology to detect and quantify translational errors proteome-wide was presented in 2019. Shotgun mass spectrometry was applied to assess amino acid misincorporation in the proteome of *E. coli* and *Saccharomyces cerevisiae*.[36] Mass spectrometry is, so far, the most powerful method to study protein forms, yet it lacks sensitivity to detect rare, low abundance proteins.

The above methods alone cannot discern whether the errors in the protein sequence have transcriptional or translational origin; however, they could be combined with RNA-Seq techniques to pinpoint the source of the error. Taken together, the findings on translational error rates show that the synthesis of a functional protein from mRNA is strikingly error-prone and that error rates differ among organisms (Table 2).

2.3 | The nature and magnitude of frameshifts

Investigations of translational errors often focus on single amino acid misincorporations.[36,37] However, another type of error that is hard to investigate proteome-wide is ribosomal frameshift. In ribosomal frameshifting, the ribosome slips on the mRNA, usually by −1 or +1, and continues translation in the new reading frame.[38–40] As a consequence, a completely changed protein sequence is synthesized. Like other phenotypic mutations, frameshifting is a stochastic process: at a given site, some proportion of ribosomes will shift, and some will continue translation in the normal frame.[41] Several studies have shown that these errors happen at low rates *in vivo*, with estimated rates per codon ranging from 10^{-5} to more than 10^{-2}.[42–44] Specific sites show much higher frameshifting rates (Table 3), for example, ~30% in SARS-CoV-2,[45] and in some cases exceeding 80%.[46] This phenomenon is called programmed ribosomal frameshifting (PRF)[47–49] and, in many cases, serves a specific function in protein production.

The most widespread use of PRF is found in viruses, possibly due to constraints on their genome size and limited availability of other mechanisms to regulate protein expression.[32]

TABLE 2 Representative translational error rates of several species

Errors in translation		
Organism	**Error rate**	**Method**
Amino acid misincorporation		
E. coli	3×10^{-4} per base	In vitro radioactive assay combined with chromatography techniques to nonsense mutations in ovalbumin[34]
E. coli	10^{-4} to 10^{-3} per base	In vivo reporter-construct assay for lysine misincorporation in the active site of luciferase[35]
E. coli	10^{-4} to 10^{-3} per base	Proteome-wide methodology based on mass spectrometry to identify and quantify amino acid mismatches in vivo[36]
Stop codon readthrough		
E. coli	50%	In vivo reporter-construct assay for stop codon readthrough based on RF2-lacZ gene fusion[101]
Frameshifting		
E. coli	$1.1 - 2.2 \times 10^{-4}$ per base	In vitro laboratory drift to assess frameshifting of homonucleotide sequences[72,91]
E. coli	1.5×10^{-5} per base	Restoration of β-galactosidase activity to stop-containing variant[144]
E. coli	3×10^{-5} per base	Restoration of β-galactosidase activity to stop-containing variant[144]
Premature termination		
E. coli	2.7×10^{-4} per codon	76% success probability for completing the fully amino-acid β-galactosidase polypeptide[145]
Yeast	$0 - 2 \times 10^{-3}$ per codon	Comparison of ribosome density on 5′ versus 3′ end of mRNA[146]

TABLE 3 Representative functional frameshifts of various genes

Functional frameshifts				
Organism	**Gene**	**Frameshift type**	**Frameshift frequency**	**Description**
E. coli	copA	FS -1	0.5	PRF generates a copper transporter and a copper chaperone from the same gene[110]
E. coli	dnaX	FS −1	0.7	Translational frameshift generates the gamma subunit of DNA polymerase III[46]
B. subtilis	cdd	FS −1	0.1–0.2	a frameshift near the stop codon results in a CDA subunit extended by 13 amino acids[147]
Bacteria	prfB	FS +1	Regulated by RF2	PRF regulates the expression of peptide chain release factor 2[101,102]
S. cerevisiae	ABP140	FS +1	Unknown	PRF leads to the production of fully length ABP140[108]
S. cerevisiae	EST3	FS +1	0.08–0.4	PRF leads to the production of fully length EST3 required for yeast telomere replication[107]
S. cerevisiae	Ty1	FS +1	0.2–0.6	The primary translation product of TYB is a TYA/TYB fusion protein expressed by translational frameshifting[148]
E. gossypii	IDP2	FS +1	0.28	PRF controls the localization of NADP-dependent isocitrate dehydrogenases (IDP) in the yeast[6]
P. anserina	PaYIP3	FS −1	0.01	PRF controls the expression of Rab-GDI complex dissociation factor[149]
H. sapiens	CCR5	FS −1	0.1	The NMD pathway and miRNAs regulate the framshifts in the CCR5 mRNA[150]
H. sapiens	PEG10	FS +1	0.3	Mammalian gene PEG10 expresses two reading frames by high PRF in embryonic-associated tissues[52]
Mammals	EDR1	FS −1	0.3	A developmentally regulated mammalian gene utilizes PRF[48]
Eukaryotes	OAZ1	FS +1	Regulated by polyamines	A negative feedback loop by a frameshifting controls the expression of the mammalian ornithine decarboxylase antizyme[55]

In prokaryotes and eukaryotes, the extent of PRF is less clear. Computational studies showed that a known −1 FS site occurs in dozens of *E. coli* genes,[50] and the *S. cerevisiae* genome contains thousands of sequences resembling −1 sites.[51] However, only a handful of genes have been experimentally confirmed to frameshift with high frequency (Table 3).

Even if a frameshift occurs with high frequency, it is difficult to infer if it is simply an error or harbors a function. Indications for a functional frameshift are, for example: (a) conservation of the frameshift site and the sequence of the frameshifted ORF; (b) binding or enzymatic activity in the shifted ORF; and (c) a phenotypic effect due to abolition of frameshifting, for example, by synonymous mutations of the frameshift site. For instance, in the human gene PEG10, a −1 frameshift was confirmed in cell culture. The frameshift site and the shifted ORF are conserved, and the frameshift produces a working protease. However, it is still unclear if this frameshift has an actual function in the living organism.[52] In other cases, frameshift sites were found to work *in vitro*, but the product could not be identified *in vivo*,[53] or a frameshift was only inferred *in silico* by similarity.[54] The only human cases of ribosomal frameshifting with a known function are the antizyme genes, OAZ1, 2, and 3, where +1 frameshifting is widely conserved.[55–57] Recently, the programmed frameshifting in human CCR5 was found to be an artifact.[58]

An exceptional case are the ciliates of the *Euplotes* genus, which use +1 or +2 frameshifting in more than 10% of their genes.[59,60] Invariably, frameshifting happens at in-frame stop codons, and it is suspected that *Euplotes* evolved to use additional signals for genuine termination, while the default behavior at a stop codon is frameshifting and continued translation.[59]

Unlike single amino acid misincorporations, frameshifting can generate an entirely new protein sequence and can occur at orders of magnitude higher frequencies. Below, we review how frameshifting can lead to novel functions (Sections 3.4–3.6).

3 | EVOLUTIONARY IMPACT OF PHENOTYPIC MUTATIONS

The evolution of an accurate transcriptional and translational machinery was crucial to the emergence of cellular life.[1] It was proposed that, at the advent of life, evolution of the codon table was driven by the necessity to minimize the effects of translational errors[1] and maximize resistance against single base mutations.[61] However, while the standard genetic code is highly robust in this regard, genetic codes with even higher robustness are possible.[62–65] As Maseshiro and Kimura pointed out, the genetic code not only has to be robust, but also allow for change not to limit evolvability—two conflicting constraints that are hard to satisfy at the same time.[62] The trade-offs between robustness and evolvability were extensively studied in the context of genetic mutations.[66–70] Below, we highlight work regarding robustness to phenotypic mutations.

3.1 | The cost and benefit of phenotypic mutations

Similar to genetic mutations, phenotypic mutations impose a cost as well as may provide benefits to the organism. The metabolic production costs of a correct and erroneous protein are similar. However, the erroneous protein incurs these production costs, often without benefiting the cell. This is because the erroneous protein will most likely have reduced or no function, as most exchanges of amino acids are destabilizing and deleterious,[71–74] or even impose cytotoxic stress on the cell due to misfolding[75] (Figure 3a).

Cells and organisms that are robust against mutations have an evolutionary advantage, as they reduce the potential costs of mutations.[66,76] Likewise, cells and organisms with reduced cost of phenotypic mutations might have an evolutionary advantage. Since only individual proteins within a population of proteins are affected by phenotypic mutations, it is difficult to compare evolutionary effects between genetic and phenotypic mutations.

Two mechanisms were proposed for coping with protein synthesis errors[37]: (a) increasing robustness of proteins to mutations; or (b) increasing the fidelity of transcription and translation (Figure 3b). Why and when these different options are preferred is still not fully understood. There is a trade-off between the cost of phenotypic mutations and mechanisms to cope with them. The observed rate of phenotypic mutations results from a balance between the fitness cost imposed by their deleterious effects[77–79] and the cost of decreasing their incidence. Any potential benefits that phenotypic mutations may provide could influence the rate.

Species in nature provide insights into the evolution of error rates. So far, no experiments have been designed to test whether higher translational fidelity could evolve under laboratory conditions. However, theoretical models predict that higher translational fidelity could be achieved by either optimizing codon usage[35] or evolving the translational machinery toward higher fidelity.[80] The

(a)

(b)

Avoiding deleterious effects by reducing error rate

Compensating increased error rate by modulating protein production

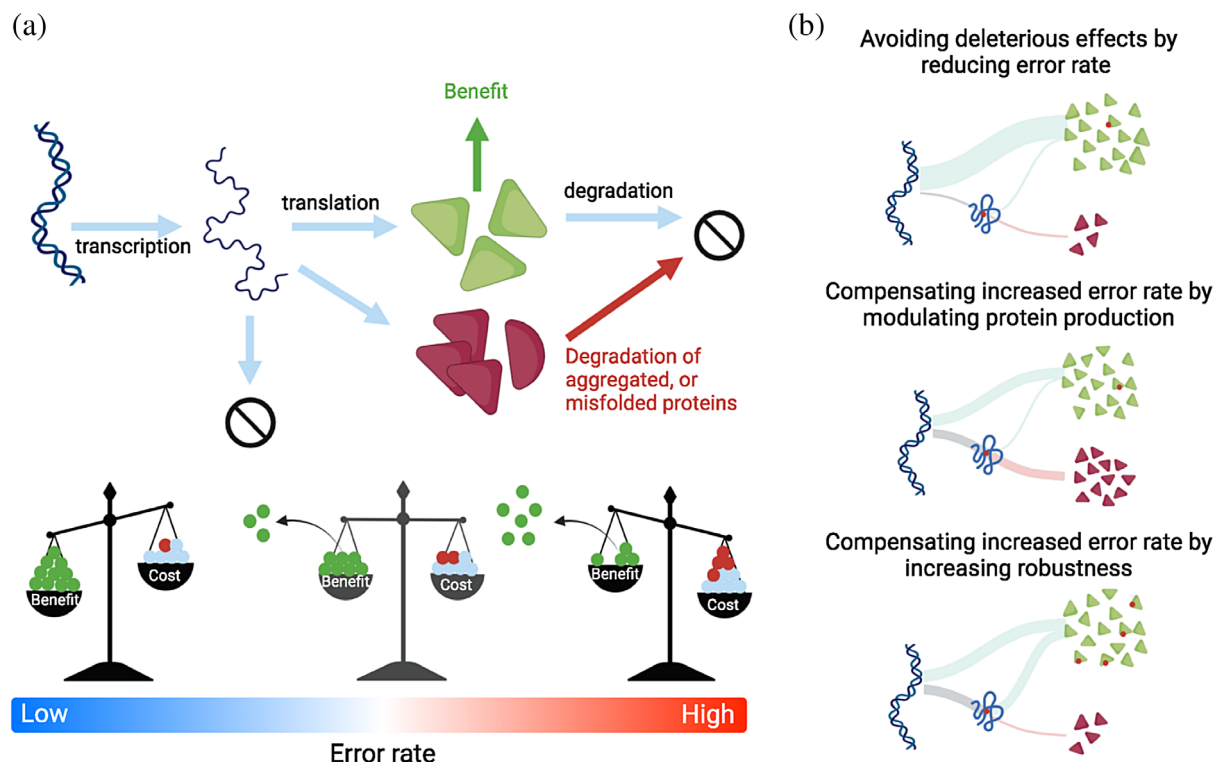Compensating increased error rate by increasing robustness

FIGURE 3 Errors during protein production can be costly, and coping mechanisms are diverse. (a) Most metabolic costs incurred during protein production do not differ between functional and nonfunctional protein variants. Other costs, however, such as opportunity costs may differ at much larger scales. Transcriptional and translational errors are likely deleterious and diminish benefits of the functional protein. In case of strongly deleterious errors, additional cytotoxic costs may occur due to protein misfolding or aggregation. A cell might tolerate additional metabolic or opportunity costs until the error rate causes the cost to exceed the benefit of the remaining functional protein. (b) To minimize the fitness effects of errors in protein production, a cell can either minimize the error rate, modulate protein production level, or increase protein tolerance to errors.[37] Minimizing global error rates are costly as it requires ATP-dependent proof-reading. Local reduction of error rates could be achieved by adapting the codon usage.[35] Protein production can either be increased to compensate for the missing erroneous proteins[83] or decreased to minimize cytotoxic costs[84]

question arises what the limits are for protein production fidelity. Bürger et al. derived a lower and upper bound for the phenotypic error rate.[80] The upper bound can be directly linked to the cost of nonfunctional proteins, as at some point, the cost of producing a fraction of nonfunctional proteins exceeds the benefit of the fraction of canonical, error-free proteins. The lower bound is closely linked to the proof-reading cost of the RNA polymerase and the ribosome. It has been shown that increasing the fidelity of ribosome proof-reading reduces bacterial growth rates.[42]

What is the reason for the vastly different error rates across species (Tables 1 and 2)? It was proposed that species with high effective population sizes have higher overall error rates because reducing error rates by kinetic proof-reading is a costly global solution. Increased efficacy of selection due to the large effective population size allows the evolution of "local" solutions on a gene-by-gene basis to reduce the effects of errors.[81] Thus, there is

no need to decrease globally the error rate by increasing the fidelity of the kinetic proof-reading. Local solutions to mis-transcription are: (a) local robustness to the consequences of mis-transcription when it occurs, and (b) locally reduced mis-transcription rates at the sites that are the most sensitive. Transcriptional error rates of non-C to U errors are lower in highly expressed genes in *E. coli* but not in *S. cerevisiae*, suggesting that *E. coli* reduces "local" error rates in proteins where selection is strongest.[82] Still, local mis-transcription rates, even of highly expressed *E. coli* genes are higher than the global mis-transcription rate in *S. cerevisiae*. While this trend holds for *S. cerevisiae* and *E. coli*, it does not generalize for other species when comparing effective population sizes and transcriptional error rate estimates from RNA-Seq experiments (Figure 2b). We note that *M. florum* is an outlier, and notably more experiments and error rate estimates are needed to validate theoretical predictions and understand how error rates evolve.

3.2 | Phenotypic mutations shape protein properties

Increased robustness to phenotypic mutations was observed by a pioneering study by Goldsmith and Tawfik, who explored the evolutionary response to increased transcriptional errors.[83] They used a bacterial system where the antibiotic resistance gene TEM-1 beta lactamase was transcribed either by a normal, high fidelity RNA polymerase or an error-prone, mutant polymerase. They studied the response of the bacteria to ampicillin selection pressure and identified two mechanisms allowing *E. coli* to cope with the erroneous TEM-1 beta lactamase proteins due to the higher transcriptional error rate: (a) increasing expression level and (b) fixation of stabilizing mutations (Figure 4). Increasead expression levels could compensate for the large amount of mistranscribed mRNA and consequently nonfunctional TEM-1 variants. Interestingly, the more stable TEM-1 exhibited increased tolerance to both phenotypic and genetic mutations.
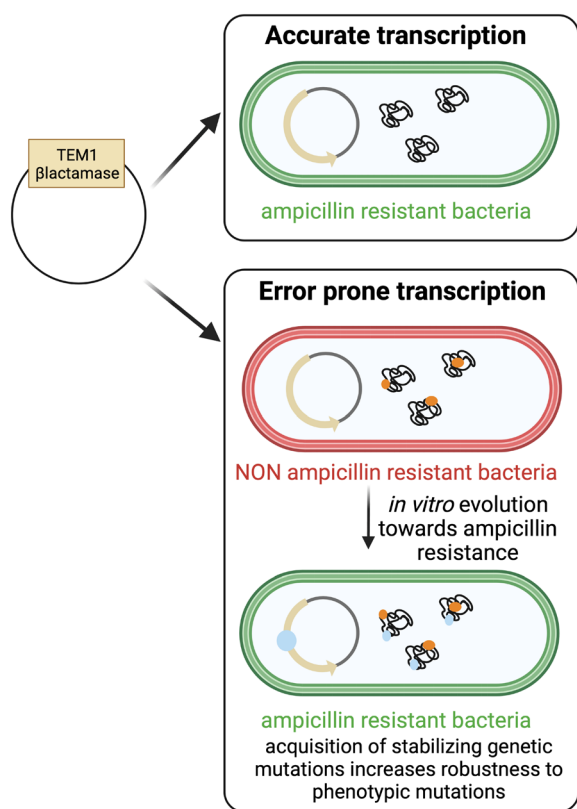

Accurate transcription

ampicillin resistant bacteria

Error prone transcription

NON ampicillin resistant bacteria

in vitro evolution towards ampicillin resistance

ampicillin resistant bacteria
acquisition of stabilizing genetic mutations increases robustness to phenotypic mutations

FIGURE 4 Phenotypic mutations play a role in shaping protein traits and tolerance to genetic mutations. TEM-1 beta-lactamase was transcribed with a high-fidelity RNA polymerase and its error-prone mutant. Higher transcriptional error rates promoted enhanced TEM-1 expression levels and stabilized enzyme variants[83]

A similar follow-up work on TEM-1 used an error-prone ribosome to study the evolutionary effects of translational errors.[84] Under weak selection, they observed mutations toward alternative start codons, decreasing translation initiation, and thereby decreasing the cytotoxic costs of misfolding and aggregation. Under strong selection, however, when reducing gene expression would be detrimental, they observed the purging of deleterious and the fixation of stabilizing mutation, which increases robustness to (phenotypic) mutations.

The evolutionary response observed in these experiments[83,84] mitigated the effects of phenotypic mutations by adjusting the expression level and by increasing the robustness of proteins (Figure 3b). Goldsmith and Tawfik surmised that phenotypic mutations might play a role in shaping protein properties such as expression levels, stability, and tolerance to genetic mutations.[83]

Kalapis et al. studied the response to protein synthesis errors by evolving a mutant yeast strain that was engineered to translate a codon ambiguously, leading to amino acid misincorporations.[85] Unlike the previous studies that focused on the response at the level of a single gene, that is, TEM-1, Kalapis et al. explored genetic changes across the whole genome of the organism. These laboratory evolution experiments revealed that fitness loss due to mistranslation was mitigated by large chromosomal duplication and deletion events that alter the dosages of numerous, functionally related proteins simultaneously. They also observed faster degradation rates by the ubiquitin-proteasome system that led to the elimination of erroneous protein products, thereby reducing the extent of toxic protein aggregation in mistranslating cells. However, they observed a trade-off between adaptation to mistranslation and survival upon starvation. As a response to an enhanced energy demand of accelerated protein turnover, the evolved lines exhibited increased glucose uptake by selective duplication of hexose transporter genes. This means that, while adaptive mechanisms to sudden and catastrophic level of mistranslation evolved rapidly, they affected cellular homeostasis.

3.3 | Phenotypic mutations promote evolvability

Changing environments may favor higher error rates which results in diverse protein populations.[44,86] The protein diversity that is generated by protein synthesis errors can occasionally be beneficial and promote evolvability.

One determinant of evolvability is the set of possible mutations available to a genotype, or "mutational neighborhood." Nelson and Masel found that selection on a better phenotypic mutational neighborhood has as its

byproduct the creation of a better genetic mutational neighborhood.[81] A benign mutational neighborhood arises as a byproduct of transiently elevated error rates via a mechanism termed emergent evolutionary capacitance[87] (similar to evolutionary capacitance but without a capacitor). Capacitance results in a higher error rate that promotes evolvability.[87] Theoretical studies on stop codon readthrough suggested the purging of the mutational neighborhood of catastrophically bad options enrich the remainder for potential adaptations through a process of elimination.[88]

So far, no experimental work has tackled how increased global error rates could facilitate adaptation. Meyerovich et al. showed how higher local error rates could be advantageous in a reporter gene. They compared bacterial survival toward antibiotic resistance using a specific antibiotic resistance gene (*cat*—encoding the chloramphenicol resistance protein), as a reporter, and showed how higher levels of errors could be selectively advantageous. They found that low temperature induced stop codon readthrough and frameshifts. Accordingly, bacteria harboring the mutated *cat* gene (i.e., inactivated via frameshift or nonsense mutation) grown at low temperature, showed higher survival rate under antibiotic selection than when grown at optimal growth temperature.[44] In this case, errors in gene expression enabled survival of strains carrying a mutated antibiotic resistance gene. The authors proposed that high error levels, in general, could facilitate the expression of pseudogenes containing frameshift and nonsense mutations. Pseudogenes are common in bacterial genomes, and they may be expressed in low amounts and become beneficial under changing environments.

3.4 | Phenotypic mutations as bridging evolutionary intermediates

Masel showed that cryptic (or hidden) genetic variation could be enriched for adaptation, and this enrichment is stronger when multiple changes are needed simultaneously to generate a potentially adaptive phenotype. While these examples were inspired by alternative splicing and stop codon readthrough due to [PSI+] prion state in yeast. It could be generalized that phenotypic mutations might act as bridging evolutionary intermediates[89] and provide a powerful mechanism for adaptation when multiple mutations are needed.

Based on theoretical modeling and simulations, Whitehead et al. showed that if two mutations are needed for a novel trait, after the first one was acquired as a genetic mutation, the second mutation can be introduced into the phenotype via a transcriptional or translational error.[90] If the novel trait is advantageous enough, the allele with only one mutation will spread through the population, even though the gene sequence does not yet code for both alleles. Thus, phenotypic mutations allow "look-ahead" for a two mutation path to evolve a novel trait.

Rockah-Shmuel et al. showed how phenotypic mutations can act as compensatory mutations and hypothesized that they can serve as evolutionary intermediates.[91] They observed that short 1–2 nt long frameshifting InDels (insertions and deletions that result from DNA polymerase slippage and alter the reading frame) persisted during a laboratory genetic drift experiment of a gene coding for the DNA methyltransferase M.HaeIII.[72] Since frameshifting InDels usually lead to loss of function, these InDels should have been purged from the population. Surprisingly, the authors found that many frameshifting InDels within homonucleotide repeats of 3–8 nt were bypassed by compensatory frameshifts by the RNA polymerase or the ribosome. Intriguingly, the genetic occurrence of InDels and the transcriptional–translational bypass to correct the induced frameshifts seem mechanistically related, and their frequencies are correlated. The longer the repeat, the higher was the frequency of InDels in the gene, and the more frequent was their bypass. Thus, the same sequence context might be slippery for the DNA polymerase and the RNA polymerase as well. It should be noted that, even when functional rescue is 100%, the bypass is not, thus truncated protein forms are also present and impose a cost on the cell.

In general, InDels could speed up sequence divergence since they can drastically alter the protein's length and sequence, unlike nonsynonymous point mutations that result in exchange of single side chains.[92] For example, an InDel that induces a frameshift and is compensated by another InDel downstream in the sequence could result in a new sequence stretch.[91] More than 600 such compensatory frameshifting InDel candidates were recently detected, including in human genes RAB36, ARHGAP6, and NCR3LG1.[93] Compensatory frameshifting InDels represent a previously overlooked source of protein variations. The evolutionary trajectory of their appearance and the role of intermediates via transcriptional or translational bypass is an exciting future research direction.

3.5 | Protein expression regulation by programmed ribosomal frameshifting

The evolutionary relevance of frameshifting depends on their frequency of occurrence. Most frameshifts are rare, as reviewed in Section 2.3. Programmed ribosomal frameshifts with very high rates may be employed to regulate protein expression. Many viruses, like HIV-1[94] and
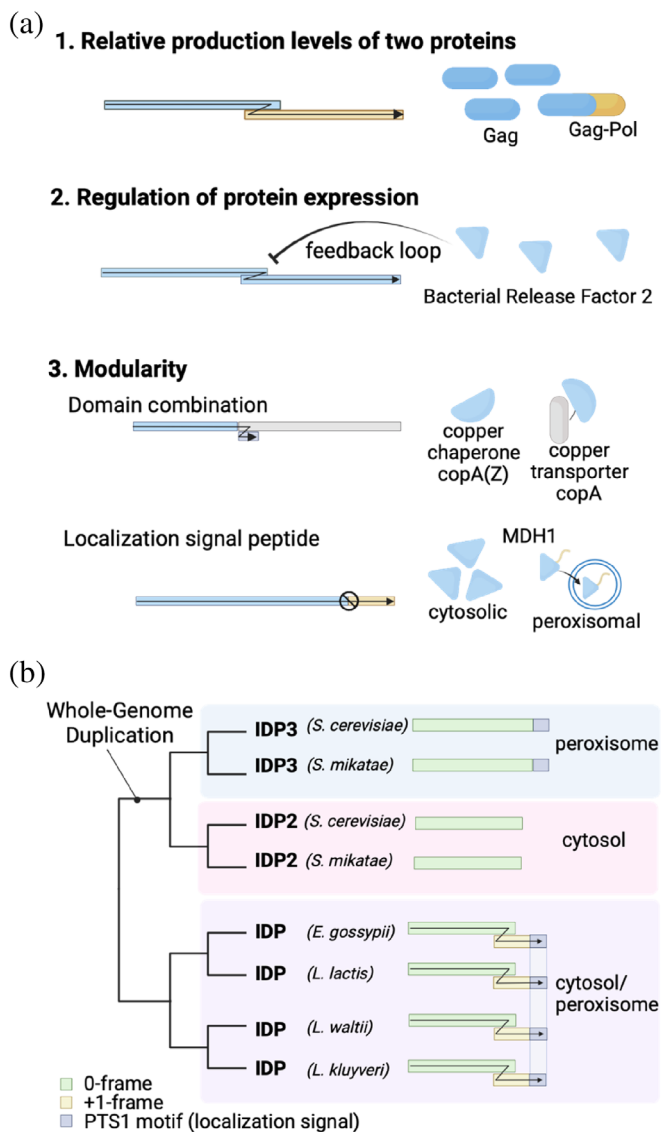
(a)

(b)

FIGURE 5 Functional innovations by ribosomal frameshifting and stop codon readthrough events. (a) (1) Frameshifting is used to regulate the relative production of proteins in many RNA viruses; for example, the polyproteins Gag and Pol are encoded in one RNA and are separated by a frameshift.[94] (2) Frameshifting is used to regulate protein expression by turning genes on or off; for example, bacterial release factor 2 (RF2) is only produced when an internal stop codon is bypassed via frameshifting.[101] With higher RF2 levels, accurate termination becomes more likely, and production of RF2 is decreased. (3) Phenotypic mutations can regulate modular function. In the copA gene (E. coli), the 5′ part of the mRNA encodes a copper-binding domain. This domain is used as part of the transporter produced by normal translation and as a soluble chaperone produced by early termination after a frameshift.[110] In the human gene MDH1, stop codon readthrough leads to synthesis of a peroxisomal signal peptide.[114] (b) Phenotypic mutation preceded a genetic solution for dual targeting. In several budding yeasts, there is only one IDP gene; however, it contains a peroxisomal signal peptide (PTS1) in the +1-frame of the 3′ UTR. In E. gossypii, frameshifting leads to producing a signal peptide and targeting the peroxisome only in a fraction of proteins.[6] Species that underwent Whole-Genome Duplication have two IDP paralogues: IDP3 bears PTS1 in the 0-frame and is fully targeted to the peroxisome, while IDP2 has no PTS1, and it localizes to the cytosol

SARS-CoV-2[95] use −1 frameshifts to control the relative production of large polycistronic sequences (Gag-Pol and Rep, respectively) (Figure 5a).

Interrupted genes or shifted open reading frames are "fixed" by a readthrough mechanism that generates functional proteins.[96–99] Under fluctuating environments or within small populations, frameshifting InDels in sequence repeats provide a rapid means of switching genes on and off.[100]

A unique case is the regulation of the expression of bacterial release factor 2 (RF2), which is responsible for translation termination. Regulation occurs via a negative feedback loop (Figure 5a). The RF2 gene (prfB in E. coli) contains a premature stop codon. High RF2 levels will lead to early termination and a truncated, nonfunctional product. With decreasing RF2 concentration, a +1 frameshift at the stop codon and continued translation becomes more likely, leading to full-length RF2.[101] This frameshift and regulatory mechanism is conserved in around 70% of bacteria.[102]

A well-studied example of eukaryotic frameshifting is the ornithine decarboxylase antizyme (OAZ).[55,103] The antizyme, which is needed to degrade polyamines, requires a +1 frameshift to be produced. Frameshifting in the encoding genes increases with polyamine concentrations, creating a negative feedback loop.[55] This mechanism is found in yeast, rats, drosophila, and humans.[104–106]

Other S. cerevisiae genes where +1 frameshifting in effect reduces translation include EST3 and ABP140, both of which require the frameshift to be produced full-length, with the truncated sequence having no known function.[107,108]

3.6 | Functional modularity via phenotypic mutations

Frameshifts are used to diversify protein function by providing functional modularity, where the full-length protein and a truncated form have overlapping functions. In E. coli, programmed ribosomal frameshifting is used in the dnaX gene, which, when translated fully, encodes the Tau subunit of DNA polymerase III. A −1 frameshift after the first 430 amino acids leads to termination shortly afterwards and production of the shorter Gamma subunit.[46] Tau is essential and Gamma is part of the Pol III holoenzyme, assumed to be functional as well.[109]

The copA gene, in contrast, encodes for a copper transporter when fully translated, while a −1 frameshift shortly after the first domain produces a copper chaperone (Figure 5a). Possibly, this PRF acts to produce the copper-binding domain in two forms: membrane-bound (as part of the transporter) and as a soluble chaperone.[110] The copper chaperone resulting from the frameshift contributes to copper tolerance by scavenging copper, thus helping cells survive toxic copper concentrations.

Another form of functional modularity is producing an extended protein harboring a signal peptide by frameshifts or stop codon readthrough (Figure 5a). Dual localization, that is, cytosolic and peroxisomal, was observed via cryptic peroxisomal signals revealed by alternative splicing and stop codon readthrough.[111] Interestingly, different organisms use different molecular mechanisms to generate the peroxisomal proteins. In the plant pathogen *Ustilago maydis*, peroxisomal targeting of Pgk1 and Gadph are due to stop codon readthrough and alternative splicing, respectively. In *Aspergillus nidulans*, peroxisomal targeting of these enzymes is achieved by the opposite mechanisms: Pgk1 uses alternative splicing while Gadph uses a stop codon readthrough.[111]

Reliable targeting regulation can be achieved by altering the stop codon context to achieve a suitable readthrough rate.[112,113] Such programmed stop codon readthrough or leaky termination was shown to generate dually targeted protein isoforms, for example, cytosolic and peroxisomal forms of NAD-dependent lactate dehydrogenase B (LDHB) and NAD-dependent malate dehydrogenase 1 (MDH1).[114]

By reconstructing the evolutionary history of a protein family that uses frameshifting, Yanagida et al. showed how a phenotypic mutation was later replaced by a genetic solution.[6] In this intriguing example, frameshifting controls the localization of NADP-dependent isocitrate dehydrogenases (IDP) in the yeast *Eremothecium gossypii*. The gene contains a cryptic peroxisomal signal peptide that is only translated after a frameshift shortly before the stop codon. Yeast species that underwent whole-genome duplication, such as *S. cerevisiae*, possess two genes, one codes for the cytosolic form, IDP2, and the other for the peroxisomal form, IDP3, marked for localization via a C-terminal signal peptide (Figure 5b). However, species without genome duplication only possess one IDP gene, likely corresponding to the cytosolic IDP2. Since the 3′UTR of IDP and the cryptic peroxisomal signal is conserved across those species, the frameshift mediated mechanism is likely conserved. Thus a single gene allows for an intermediate phenotype between a single cytosolic gene and two gene copies, preceding gene duplication and the genetic solution to a phenotypic mutation.[6]

4 | PHENOTYPIC MUTATIONS IN DISEASE

4.1 | Phenotypic mutations as therapeutics

There is not much known about the endogenous phenotypic consequences of phenotypic mutations since there is no error-free cell or organism to compare to. We have, however, experimental evidence that an increased error rate reduces growth rate in bacteria. In fact, inducing extensive translational errors is an effective antibiotic treatment.[115] Aminoglycoside antibiotics work by binding to the bacterial ribosomal decoding center, which leads to misincorporation of near-cognate aminoacyl-tRNAs and induces translation error clusters.[115] At high concentration they inhibit protein synthesis and induce death. Aminoglycosides have only very low affinity to eukaryotic ribosomes. This low affinity could still be exploited to induce stop codon readthrough for genetic diseases caused by a premature stop codon (or nonsense mutation). Increasing stop codon readthrough was suggested as a therapy for approximately 40 genetic diseases caused by premature termination due to nonsense mutation.[38,116]

Recently, a gene, named Shiftless after its function, was discovered to inhibit −1 PRF[117] and was suggested as a broad-spectrum inhibitor of viruses that use PRF, such as HIV-1. Later it was shown to reduce viral replication of Flaviviruses such as Zika[118] and coronaviruses, among others.[119]

4.2 | Phenotypic mutations and diseases

Increased transcriptional infidelity contributes to several human diseases, such as cancer.[120] Errors in translation underlie the pathogenesis of many neurodegenerative disorders[121] such as Alzheimer's[122] and Huntington's diseases.[123,124]

Three mechanisms have been proposed to account for the production of disease-specific protein variants: (a) ribosomal frameshifting,[122,125] (b) alternative initiation sites, such as repeat-associated non-AUG (RAN) translation,[126] and (c) near-cognate start codon initiated translation.[127] Here we list some examples for each mechanism.

Several diseases are linked to repeat expansions, for example, CAG expansion, including Huntington disease, spinal bulbar muscular atrophy, and spinocerebellar ataxia types 1, 2, 3, 6, 7, and 17.

In spinocerebellar ataxia type 3 (SCA3), ataxin 3 contains CAG repeats whose in-frame translation produces an extended polyglutamine stretch; however, polyalanine-containing proteins are also generated in patient neurons.[125]

In Huntington's disease (HD), the most common CAG repeat disease, the repeats produce polyglutamine when in frame, and polyserine and polyalanine in the +1 and −1 frames, which have been detected in autopsy samples of HD patients, and in a transgenic mouse model

of HD.[123,124] The translation of expanded CAG repeats leads to a depletion of charged glutaminyl-transfer RNA that pairs exclusively to the CAG codon. This results in translational frameshifting in cell culture.[128]

Polyalanine has also been detected in tissues of patients with another CAG repeat disease, spinocerebellar ataxia type 8 (SCA8) associated with ataxin 8 (ATXN8).[126] Rather than frameshifting, in this case polyalanine seems to be produced by RAN translation.[126]

In myotonic dystrophy type 1 (DM1), CAG expansion transcripts result in the accumulation of polyglutamine expansion proteins.[126] In general, CAG expansion constructs express homopolymeric polyglutamine, polyalanine, and polyserine proteins in the absence of an ATG start codon.

Abnormal disease-specific repeat proteins that are synthesized from both sense and antisense transcripts through RAN translation have been detected in tissues from patients with GGGGCC repeat expansion in the first intron of C9ORF72 that causes both amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD),[129,130] as well as in patients with CCTG repeat expansion in the first intron of ZNF9 in myotonic dystrophy type 2 (DM2).[131]

Lastly, an example for near-cognate initiation is the FMR1 protein that is associated with Fragile X-associated tremor/ataxia syndrome (FXTAS). This neurodegenerative disease is caused by a limited expansion of CGG repeats in the 5′ UTR of FMR1. The CGG repeats are translated through initiation by an ACG codon into polyglycine protein, that is toxic in mice.[127]

The above examples show how repeat associated frameshifting and translation initiation can produce alternative protein forms that are toxic and contribute to diseases. Thus, alternative protein forms produced by translational errors rather than the canonical proteins may drive the pathogenesis of many diseases.

5 | CONCLUSIONS AND FUTURE DIRECTIONS

Phenotypic mutations are orders of magnitude more frequent than genetic ones (Tables 1 and 2), yet their evolutionary impact is less understood. Due to their transient and stochastic nature, proteome- and transcriptome-wide experimental characterizations of phenotypic mutations only recently became feasible by the advance of mass spectrometry[36] and RNA-Seq technics.[20,23,25,132]

In the future, an innovative technique could obtain single amino acid sequence information for millions of molecules in parallel.[133,134] Proteins are cleaved, and the resulting peptides are fluorescently labeled and immobilized on a glass surface. Then, the peptides are imaged by microscopy to detect changes in fluorescence after Edman degradation. This method has the potential to process complex proteomic samples with the sensitivity required to detect translational errors at the single-molecule level.

There is large variability of protein synthesis fidelity across organisms (Tables 1 and 2). The observed error rates are the result of balancing deleterious effects and the cost of higher fidelity protein synthesis (Figure 3a). The deleterious effect of phenotypic mutations can be mitigated by increasing expression levels or accumulation of stabilizing genetic mutations (Figure 3b). Thus, phenotypic mutations play a role in shaping protein properties such as expression levels, stability, and tolerance to genetic mutations. Experiments on TEM-1 as a model protein indicated an immediate effect of phenotypic mutations on protein dose and stability (Figure 4).[83,84]

Phenotypic mutations may act as bridging intermediates,[81,89–91] or look-ahead mutations that allow otherwise deleterious intermediates to survive in the population.[90] Rockah-Shmuel et al. demonstrated experimentally how deleterious frameshifts could be bypassed via phenotypic mutations.[91] Yanagida et al. discovered a gene family where the phenotypic mutation preceded the genetic solution and paved the way to evolutionary adaptation via gene duplication (Figure 5b).[6] There are many examples in all kingdoms of life where programmed frameshifts and stop codon readthroughs are used as a regulatory mechanism or lead to altered protein function, for example, a change in protein localization (Figure 5a).[6,111]

It is difficult to distinguish adaptive, that is, programmed and nonadaptive molecular errors.[135] For example, recently, the CCR5 frameshifting was found to be an artifact.[58] While, ribosome profiling identified numerous stop codon readthrough events in *S. cerevisiae*,[136] *D. melanogaster*[137,138] and mammals.[139] There is a negative correlation with the expression level and readthrough rates, and read-through motifs are avoided in highly expressed genes suggesting that most stop codon readthrough events are nonadaptive molecular errors.[135] Programmed errors have likely high rate of occurrence, that is, they are frequent; and are likely conserved among multiple species and less tolerant to loss of function mutations (since they are under functional constraints). Candidates for programmed frameshifts and stop codon readthrough ultimately have to be experimentally verified, for example by measuring the functional and/or fitness effects of altering the read-through rate.

Phenotypic mutations have often been linked to fitness decrease and diseases. It has been shown how transcriptional infidelity causes cancer heterogeneity.[120]

Furthermore, translational inaccuracy, often due to repeats in protein sequences, contributes to several neurodegenerative disorders.[121–124] Inducing phenotypic mutations is used as therapeutic. For example, aminoglycosides inhibit ribosome translocation inducing translational errors. Since they have low affinity to eukaryotic ribosomes, aminoglycosides are often used as an antibiotic. Also, their propensity to induce translation errors can be utilized to alleviate the symptoms of human genetic diseases, for example, by inducing stop codon readthrough for diseases caused by nonsense mutations.

It is unknown how long the effect of phenotypic mutations lasts. Proteins harboring phenotypic mutations can be passed to the next generation as protein half-life in many cases exceeds cell cycle time. Phenotypic mutations can also affect new generations by triggering transcription network loops. For example, an increased error rate during transcription modulates switching the lac operon from the uninduced state to the induced after cell division.[140]

Phenotypic mutations might be incorporated directly into the genome via reverse transcription.[30] Reverse transcription is mainly used by retroviruses and occasionally by cellular life.[141] This mechanism was shown in cancer, where intron-less versions of human pseudogenes were identified in cancerous genomes that were most likely the result of reverse transcription in somatic cells.[142] Reverse transcription could serve as a potential evolvability mechanism by which transcriptional errors could become genetic. However, this hypothesis has not been tested yet. Thus, future studies are needed to answer the unmet questions of whether phenotypic mutations can be replaced with genetic ones.

Although phenotypic mutations are not individually subjected to inheritance, as genetic mutations are, the examples highlighted in this review suggest that they, nevertheless, play a critical role in evolution. Understanding the protein diversity and phenotypic diversity caused by phenotypic mutations will advance our understanding of protein evolution and will have implications on human health and diseases.

AUTHOR CONTRIBUTIONS

Maria Luisa Romero Romero: Conceptualization (supporting); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Cedric Landerer:** Conceptualization (supporting); visualization (equal); writing – original draft (supporting); writing – review and editing (equal). **Jonas Poehls:** Visualization (equal); writing – original draft (supporting); writing – review and editing (equal). **Agnes Toth-Petroczy:** Conceptualization (lead); supervision (lead); writing – original draft (equal); writing – review and editing (lead).

ORCID

Maria Luisa Romero Romero https://orcid.org/0000-0003-3397-6758
Cedric Landerer https://orcid.org/0000-0002-2122-6506
Jonas Poehls https://orcid.org/0000-0002-6137-2794
Agnes Toth-Petroczy https://orcid.org/0000-0002-0333-604X

REFERENCES

1. Woese CR. On the evolution of the genetic code. Proc Natl Acad Sci U S A. 1965;54:1546–1552.
2. Kunkel TA, Bebenek K. DNA replication fidelity. Annu Rev Biochem. 2000;69:497–529.
3. Milo R, Jorgensen P, Springer M, Weber G. Bionumbers. The database to useful biological numbers; 2007. bionumbers.hms.harvard.edu.
4. Drummond DA. How infidelity creates a sticky situation. Mol Cell. 2012;48:663–664.
5. Tawfik DS. Messy biology and the origins of evolutionary innovations. Nat Chem Biol. 2010;6:692–696.
6. Yanagida H, Gispan A, Kadouri N, et al. The evolutionary potential of phenotypic mutations. PLoS Genet. 2015;11:e1005445.
7. Bradley CC, Gordon AJE, Halliday JA, Herman C. Transcription fidelity: new paradigms in epigenetic inheritance, genome instability and disease. DNA Repair. 2019;81:102652.
8. Saxowsky TT, Meadows KL, Klungland A, Doetsch PW. 8-Oxoguanine-mediated transcriptional mutagenesis causes Ras activation in mammalian cells. Proc Natl Acad Sci U S A. 2008;105:18877–18882.
9. Traverse CC, Ochman H. Conserved rates and patterns of transcription errors across bacterial growth states and lifestyles. Proc Natl Acad Sci U S A. 2016;113:3311–3316.
10. Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. Cell. 2005;123:1025–1036.
11. Springgate CF, Loeb LA. On the fidelity of transcription by *Escherichia coli* ribonucleic acid polymerase. J Mol Biol. 1975;97:577–591.
12. Trautner TA, Swartz MN, Kornberg A. Enzymatic synthesis of deoxyribonucleic acid. X. Influence of bromouracil substitutions on replication. Proc Natl Acad Sci U S A. 1962;48:449–455.

13. Rosenberger RF, Foskett G. An estimate of the frequency of in vivo transcriptional errors at a nonsense codon in *Escherichia coli*. Mol Gen Genet. 1981;183:561–563.

14. Rosenberger RF, Hilton J. The frequency of transcriptional and translational errors at nonsense codons in the lacZ gene of *Escherichia coli*. Mol Gen Genet. 1983;191:207–212.

15. Blank A, Gallant JA, Burgess RR, Loeb LA. An RNA polymerase mutant with reduced accuracy of chain elongation. Biochemistry. 1986;25:5920–5928.

16. Bubunenko MG, Court CB, Rattray AJ, et al. A Cre transcription fidelity reporter identifies GreA as a major RNA proofreading factor in *Escherichia coli*. Genetics. 2017;206:179–187.

17. Nesser NK, Peterson DO, Hawley DK. RNA polymerase II subunit Rpb9 is important for transcriptional fidelity in vivo. Proc Natl Acad Sci U S A. 2006;103:3268–3273.

18. Shaw RJ, Bonawitz ND, Reines D. Use of an in vivo reporter assay to test for transcriptional and translational fidelity in yeast. J Biol Chem. 2002;277:24420–24426.

19. Strathern JN, Jin DJ, Court DL, Kashlev M. Isolation and characterization of transcription fidelity mutants. Biochim Biophys Acta. 2012;1819:694–699.

20. Gout J-F, Kelley Thomas W, Smith Z, Okamoto K, Lynch M. Large-scale detection of in vivo transcription errors. Proc Natl Acad Sci U S A. 2013;110:18584–18589.

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

22. Acevedo A, Brodsky L, Andino R. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. Nature. 2014;505:686–690.

23. Lou DI, Hussmann JA, McBee RM, et al. High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. Proc Natl Acad Sci U S A. 2013;110:19872–19877.

24. Li W, Lynch M. Universally high transcript error rates in bacteria. Elife. 2020;9:e54898.

25. Gout J-F, Li W, Fritsch C, et al. The landscape of transcription errors in eukaryotic cells. Sci Adv. 2017;3:e1701484.

26. Imashimizu M, Oshima T, Lubkowska L, Kashlev M. Direct assessment of transcription fidelity by high-resolution RNA sequencing. Nucleic Acids Res. 2013;41:9090–9104.

27. James K, Gamba P, Cockell SJ, Zenkin N. Misincorporation by RNA polymerase is a major source of transcription pausing in vivo. Nucleic Acids Res. 2017;45:1105–1113.

28. Lynch M. Evolution of the mutation rate. Trends Genet. 2010; 26:345–352.

29. Zheng J, Guo N, Wagner A. Mistranslation reduces mutation load in evolving proteins through negative epistasis with DNA mutations. Mol Biol Evol. 2021;38:4792–4804.

30. Yona AH, Frumkin I, Pilpel Y. A relay race on the evolutionary adaptation spectrum. Cell. 2015;163:549–559.

31. Brierley I, Dos Ramos FJ. Programmed ribosomal frameshifting in HIV-1 and the SARS–CoV. Virus Res. 2006;119:29–42.

32. Atkins JF, Loughran G, Bhatt PR, Firth AE, Baranov PV. Ribosomal frameshifting and transcriptional slippage: from genetic steganography and cryptography to adventitious use. Nucleic Acids Res. 2016;44:7007–7078.

33. Jungreis I, Chan CS, Waterhouse RM, Fields G, Lin MF, Kellis M. Evolutionary dynamics of abundant stop codon readthrough. Mol Biol Evol. 2016;33:3108–3132.

34. Loftfield RB, Vanderjagt D. The frequency of errors in protein biosynthesis. Biochem J. 1972;128:1353–1356.

35. Kramer EB, Farabaugh PJ. The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. RNA. 2007;13:87–96.

36. Mordret E, Dahan O, Asraf O, et al. Systematic detection of amino acid substitutions in proteomes reveals mechanistic basis of ribosome errors and selection for translation fidelity. Mol Cell. 2019;75:427–441.e5.

37. Drummond DA, Wilke CO. The evolutionary consequences of erroneous protein synthesis. Nat Rev Genet. 2009;10: 715–724.

38. Bidou L, Rousset J-P, Namy O. Translational errors: from yeast to new therapeutic targets. FEMS Yeast Res. 2010;10: 1070–1082.

39. Yan S, Wen JD, Bustamante C, Tinoco I. Ribosome excursions during mRNA translocation mediate broad branching of frameshift pathways. Cell. 2015;160:870–881.

40. Caliskan N, Peske F, Rodnina MV. Changed in translation: mRNA recoding by −1 programmed ribosomal frameshifting. Trends Biochem Sci. 2015;40:265–274.

41. Chen J, Petrov A, Johansson M, Tsai A, O'leary SE, Puglisi JD. Dynamic pathways of −1 translational frameshifting. Nature. 2014;512:328–332.

42. Kurland CG. Translational accuracy and the fitness of bacteria. Annu Rev Genet. 1992;26:29–50.

43. Atkins JF, Elseviers D, Gorini L. Low activity of β-galactosidase in frameshift mutants of *Escherichia coli*. Proc Natl Acad Sci U S A. 1972;69:1192–1195.

44. Meyerovich M, Mamou G, Ben-Yehuda S. Visualizing high error levels during gene expression in living bacterial cells. Proc Natl Acad Sci U S A. 2010;107:11543–11548.

45. Kelly JA, Olson AN, Neupane K, et al. Structural and functional conservation of the programmed −1 ribosomal frameshift signal of SARS coronavirus 2 (SARS-CoV-2). J Biol Chem. 2020;295:10741–10748.

46. Tsuchihashi Z, Kornberg A. Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. Proc Natl Acad Sci U S A. 1990;87:2516–2520.

47. 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.

48. Shigemoto K, Brennan J, Walls E, et al. Identification and characterisation of a developmentally regulated mammalian gene that utilises −1 programmed ribosomal frameshifting. Nucleic Acids Res. 2001;29:4079–4088.

49. Farabaugh PJ. Programmed translational frameshifting. Annu Rev Genet. 1996;30:507–528.

50. Gurvich OL, Baranov PV, Zhou J, Hammer AW, Gesteland RF, Atkins JF. Sequences that direct significant levels of frameshifting are frequent in coding regions of *Escherichia coli*. EMBO J. 2003;22:5941–5950.

51. Jacobs JL, Belew AT, Rakauskaite R, Dinman JD. Identification of functional, endogenous programmed −1 ribosomal frameshift signals in the genome of Saccharomyces cerevisiae. Nucleic Acids Research. 2007;35:165–174.

52. Clark MB, Jänicke M, Gottesbühren U, et al. Mammalian gene PEG10 expresses two reading frames by high efficiency

−1 frameshifting in embryonic-associated tissues. J Biol Chem. 2007;282:37359–37369.

53. Wills NM, Moore B, Hammer A, Gesteland RF, Atkins JF. A functional −1 ribosomal frameshift signal in the human paraneoplastic Ma3 gene. J Biol Chem. 2006;281:7082–7088.

54. Dinan AM, Atkins JF, Firth AE. ASXL gain-of-function truncation mutants: defective and dysregulated forms of a natural ribosomal frameshifting product? Biol Direct. 2017;12:24.

55. Matsufuji S, Matsufuji T, Miyazaki Y, et al. Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. Cell. 1995;80:51–60.

56. Ivanov IP, Gesteland RF, Atkins JF. A second mammalian antizyme: conservation of programmed ribosomal frameshifting. Genomics. 1998;52:119–129.

57. Ivanov IP, Atkins JF. Ribosomal frameshifting in decoding antizyme mRNAs from yeast and protists to humans: close to 300 cases reveal remarkable diversity despite underlying conservation. Nucleic Acids Res. 2007;35:1842–1858.

58. Khan YA, Loughran G, Steckelberg A-L, et al. Evaluating ribosomal frameshifting in CCR5 mRNA decoding. Nature. 2022;604:E16–E23.

59. Lobanov AV, Heaphy SM, Turanov AA, et al. Position-dependent termination and widespread obligatory frameshifting in Euplotes translation. Nat Struct Mol Biol. 2017;24: 61–68.

60. Wang R, Zhang Z, Du J, Fu Y, Liang A. Large-scale mass spectrometry-based analysis of Euplotes octocarinatus supports the high frequency of +1 programmed ribosomal frameshift. Sci Rep. 2016;6:1–8.

61. Figureau A, Pouzet M. Genetic code and optimal resistance to the effects of mutations. Orig Life. 1984;14:579–588.

62. Maeshiro T, Kimura M. The role of robustness and change-ability on the origin and evolution of genetic codes. Proc Natl Acad Sci U S A. 1998;95:5088–5093.

63. Koonin EV, Novozhilov AS. Origin and evolution of the genetic code: the universal enigma. IUBMB Life. 2009;61: 99–111.

64. Wichmann S, Ardern Z. Optimality in the standard genetic code is robust with respect to comparison code sets. Biosystems. 2019;185:104023.

65. Itzkovitz S, Alon U. The genetic code is nearly optimal for allowing additional information within protein-coding sequences. Genome Res. 2007;17:405–412.

66. 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.

67. Dellus-Gur E, Toth-Petroczy A, Elias M, Tawfik DS. What makes a protein fold amenable to functional innovation? Fold polarity and stability trade-offs. J Mol Biol. 2013;425:2609–2621.

68. Tóth-Petróczy Á, Tawfik DS. The robustness and innovability of protein folds. Curr Opin Struct Biol. 2014;26:131–138.

69. Zheng J, Guo N, Wagner A. Selection enhances protein evolvability by increasing mutational robustness and foldability. Science. 2020;370:eabb5962.

70. Jackson C, Toth-Petroczy A, Kolodny R, et al. Adventures on the routes of protein evolution-in memoriam Dan Salah Tawfik (1955–2021). J Mol Biol. 2022;434:167462.

71. Eyre-Walker A, Keightley PD, Smith NGC, Gaffney D. Quantifying the slightly deleterious mutation model of molecular evolution. Mol Biol Evol. 2002;19:2142–2149.

72. Rockah-Shmuel L, Tóth-Petróczy Á, Tawfik DS. Systematic mapping of protein mutational space by prolonged drift reveals the deleterious effects of seemingly neutral mutations. PLoS Comput Biol. 2015;11:e1004421.

73. Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. Am J Hum Genet. 2007;80: 727–739.

74. Soskine M, Tawfik DS. Mutational effects and the evolution of new protein functions. Nat Rev Genet. 2010;11: 572–582.

75. Bucciantini M, Giannoni E, Chiti F, et al. Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. Nature. 2002;416:507–511.

76. Bloom JD, Silberg JJ, Wilke CO, Drummond DA, Adami C, Arnold FH. Thermodynamic prediction of protein neutrality. Proc Natl Acad Sci U S A. 2005;102:606–611.

77. 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 U S A. 2005; 102:1697–1701.

78. Nangle LA, Motta CM, Schimmel P. Global effects of mistranslation from an editing defect in mammalian cells. Chem Biol. 2006;13:1091–1100.

79. Lee JW, Beebe K, Nangle LA, et al. Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. Nature. 2006;443:50–55.

80. Bürger R, Willensdorfer M, Nowak MA. Why are phenotypic mutation rates much higher than genotypic mutation rates? Genetics. 2006;172:197–206.

81. Rajon E, Masel J. Evolution of molecular error rates and the consequences for evolvability. Proc Natl Acad Sci U S A. 2011; 108:1082–1087.

82. Meer KM, Nelson PG, Xiong K, Masel J. High transcriptional error rates vary as a function of gene expression level. Genome Biol Evol. 2020;12:3754–3761.

83. Goldsmith M, Tawfik DS. Potential role of phenotypic mutations in the evolution of protein expression and stability. Proc Natl Acad Sci U S A. 2009;106:6197–6202.

84. Bratulic S, Gerber F, Wagner A. Mistranslation drives the evolution of robustness in TEM-1 β-lactamase. Proc Natl Acad Sci U S A. 2015;112:12758–12763.

85. Kalapis D, Bezerra AR, Farkas Z, et al. Evolution of robustness to protein mistranslation by accelerated protein turnover. PLoS Biol. 2015;13:e1002291.

86. Zhang H, Lyu Z, Fan Y, et al. Metabolic stress promotes stop-codon readthrough and phenotypic heterogeneity. Proc Natl Acad Sci U S A. 2020;117:22167–22172.

87. Nelson P, Masel J. Evolutionary capacitance emerges spontaneously during adaptation to environmental changes. Cell Rep. 2018;25:249–258.

88. Rajon E, Masel J. Compensatory evolution and the origins of innovations. Genetics. 2013;193:1209–1220.

89. Masel J. Cryptic genetic variation is enriched for potential adaptations. Genetics. 2006;172:1985–1991.

90. Whitehead DJ, Wilke CO, Vernazobres D, Bornberg-Bauer E. The look-ahead effect of phenotypic mutations. Biol Direct. 2008;3:18.

91. Rockah-Shmuel L, Tóth-Petróczy Á, Sela A, Wurtzel O, Sorek R, Tawfik DS. Correlated occurrence and bypass of frame-shifting insertion-deletions (InDels) to give functional proteins. PLoS Genet. 2013;9:e1003882.

92. Bogarad LD, Deem MW. A hierarchical approach to protein molecular evolution. Proc Natl Acad Sci U S A. 1999;96:2591–2595.

93. Biba D, Klink G, Bazykin G. Pairs of mutually compensatory frameshifting mutations contribute to protein evolution. Mol Biol Evol. 2022;39(3):mscac031.

94. Hung M, Patel P, Davis S, Green SR. Importance of ribosomal frameshifting for human immunodeficiency virus Type 1 particle assembly and replication. J Virol. 1998;72:4819–4824.

95. Kelly JA, Woodside MT, Dinman JD. Programmed −1 ribosomal frameshifting in coronaviruses: a therapeutic target. Virology. 2021;554:75–82.

96. Cobucci-Ponzano B, Trincone A, Giordano A, Rossi M, Moracci M. Identification of an archaeal alpha-L-fucosidase encoded by an interrupted gene. Production of a functional enzyme by mutations mimicking programmed −1 frameshifting. J Biol Chem. 2003;278:14622–14631.

97. Wagner LA, Weiss RB, Driscoll R, Dunn DS, Gesteland RF. Transcriptional slippage occurs during elongation at runs of adenine or thymine in Escherichia coli. Nucleic Acids Res. 1990;18:3529–3535.

98. Tamas I, Wernegreen JJ, Nystedt B, et al. Endosymbiont gene functions impaired and rescued by polymerase infidelity at poly(A) tracts. Proc Natl Acad Sci U S A. 2008;105:14934–14939.

99. Cobucci-Ponzano B, Guzzini L, Benelli D, et al. Functional characterization and high-throughput proteomic analysis of interrupted genes in the archaeon Sulfolobus solfataricus. J Proteome Res. 2010;9:2496–2507.

100. Wernegreen JJ, Kauppinen SN, Degnan PH. Slip into something more functional: selection maintains ancient frameshifts in homopolymeric sequences. Mol Biol Evol. 2010;27:833–839.

101. Craigen WJ, Caskey CT. Expression of peptide chain release factor 2 requires high-efficiency frameshift. Nature. 1986;322:273–275.

102. Baranov PV, Gesteland RF, Atkins JF. Release factor 2 frameshifting sites in different bacteria. EMBO Rep. 2002;3:373–377.

103. Gesteland RF, Weiss RB, Atkins JF. Recoding: reprogrammed genetic decoding. Science. 1992;257:1640–1641.

104. Ivanov IP, Gesteland RF, Atkins JF. Evolutionary specialization of recoding: frameshifting in the expression of S. cerevisiae antizyme mRNA is via an atypical antizyme shift site but is still +1. RNA. 2006;12:332–337.

105. Ivanov IP, Gesteland RF, Atkins JF. Survey and summary: antizyme expression: a subversion of triplet decoding, which is remarkably conserved by evolution, is a sensor for an autoregulatory circuit. Nucleic Acids Res. 2000;28:3185-3196.

106. Ivanov IP, Simin K, Letsou A, Atkins JF, Gesteland RF. The Drosophila gene for antizyme requires ribosomal frameshifting for expression and contains an intronic gene for snRNP Sm D3 on the opposite strand. Mol Cell Biol. 1998;18:1553–1561.

107. Morris DK, Lundblad V. Programmed translational frameshifting in a gene required for yeast telomere replication. Curr Biol. 1997;7:969–976.

108. Asakura T, Sasaki T, Nagano F, et al. Isolation and characterization of a novel actin filament-binding protein from Saccharomyces cerevisiae. Oncogene. 1998;16:121–130.

109. Dohrmann PR, Correa R, Frisch RL, Rosenberg SM, McHenry CS. The DNA polymerase III holoenzyme contains γ and is not a trimeric polymerase. Nucleic Acids Res. 2016;44:1285–1297.

110. Meydan S, Klepacki D, Karthikeyan S, et al. Programmed ribosomal frameshifting generates a copper transporter and a copper chaperone from the same gene. Mol Cell. 2017;65:207–219.

111. Freitag J, Ast J, Bölker M. Cryptic peroxisomal targeting via alternative splicing and stop codon read-through in fungi. Nature. 2012;485:522–525.

112. Wangen JR, Green R. Stop codon context influences genome-wide stimulation of termination codon readthrough by aminoglycosides. Elife. 2020;9:e52611.

113. Anzalone AV, Zairis S, Lin AJ, Rabadan R, Cornish VW. Interrogation of eukaryotic stop codon readthrough signals by in vitro RNA selection. Biochemistry. 2019;58:1167–1178.

114. Stiebler AC, Freitag J, Schink KO, et al. Ribosomal readthrough at a short UGA stop codon context triggers dual localization of metabolic enzymes in fungi and animals. PLoS Genet. 2014;10:e1004685.

115. Wohlgemuth I, Garofalo R, Samatova E, et al. Translation error clusters induced by aminoglycoside antibiotics. Nat Commun. 2021;12:1830.

116. Keeling KM, Xue X, Gunn G, Bedwell DM. Therapeutics based on stop codon readthrough. Annu Rev Genomics Hum Genet. 2014;15:371–394.

117. Wang X, Xuan Y, Han Y, et al. Regulation of HIV-1 gag-pol expression by shiftless, an inhibitor of programmed −1 ribosomal frameshifting. Cell. 2019;176:625–635.e14.

118. Hanners NW, Mar KB, Boys IN, et al. Shiftless inhibits flavivirus replication in vitro and is neuroprotective in a mouse model of Zika virus pathogenesis. Proc Natl Acad Sci U S A. 2021;118:e2111266118.

119. Napthine S, Hill CH, Nugent HCM, Brierley I. Modulation of Viral programmed ribosomal frameshifting and stop codon readthrough by the host restriction factor shiftless. Viruses. 2021;13:1230.

120. Brulliard M, Lorphelin D, Collignon O, et al. Nonrandom variations in human cancer ESTs indicate that mRNA heterogeneity increases during carcinogenesis. Proc Natl Acad Sci U S A. 2007;104:7522–7527.

121. Gao FB, Richter JD, Cleveland DW. Rethinking unconventional translation in neurodegeneration. Cell. 2017;171:994–1000.

122. van Leeuwen FW, Burbach JP, Hol EM. Mutations in RNA: a first example of molecular misreading in Alzheimer's disease. Trends Neurosci. 1998;21:331–335.

123. Davies JE, Rubinsztein DC. Polyalanine and polyserine frameshift products in Huntington's disease. J Med Genet. 2006;43:893–896.

124. Bañez-Coronel M, Ayhan F, Tarabochia AD, et al. RAN translation in Huntington disease. Neuron. 2015;88:667–677.

125. Gaspar C, Jannatipour M, Dion P, et al. CAG tract of MJD-1 may be prone to frameshifts causing polyalanine accumulation. Hum Mol Genet. 2000;9:1957–1966.

126. Zu T, Gibbens B, Doty NS, et al. Non-ATG-initiated translation directed by microsatellite expansions. Proc Natl Acad Sci U S A. 2011;108:260–265.

127. Sellier C, Buijsen RAM, He F, et al. Translation of expanded CGG repeats into FMRpolyG is pathogenic and May contribute to fragile X tremor ataxia syndrome. Neuron. 2017;93: 331–347.

128. Girstmair H, Saffert P, Rode S, et al. Depletion of cognate charged transfer RNA causes translational frameshifting within the expanded CAG stretch in huntingtin. Cell Rep. 2013;3:148–159.

129. Ash PEA, Bieniek KF, Gendron TF, et al. Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. Neuron. 2013;77: 639–646.

130. Mori K, Weng S-M, Arzberger T, et al. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science. 2013;339:1335–1338.

131. Zu T, Cleary JD, Liu Y, et al. RAN translation regulated by muscleblind proteins in myotonic dystrophy type 2. Neuron. 2017;95:1292–1305.e5.

132. Reid-Bayliss KS, Loeb LA. Accurate RNA consensus sequencing for high-fidelity detection of transcriptional mutagenesis-induced epimutations. Proc Natl Acad Sci U S A. 2017;114: 9415–9420.

133. Swaminathan J, Boulgakov AA, Hernandez ET, et al. Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. Nat Biotechnol. 2018;36: 1076–1082.

134. Marx V. A dream of single-cell proteomics. Nat Methods. 2019;16:809–812.

135. Li C, Zhang J. Stop-codon read-through arises largely from molecular errors and is generally nonadaptive. PLoS Genet. 2019;15:1–16.

136. Artieri CG, Fraser HB. Evolution at two levels of gene expression in yeast. Genome Res. 2014;24:411–421.

137. Jungreis I, Lin MF, Spokony R, et al. Evidence of abundant stop codon readthrough in Drosophila and other metazoa. Genome Res. 2011;21:2096–2113.

138. Dunn JG, Foo CK, Belletier NG, Gavis ER, Weissman JS. Ribosome profiling reveals pervasive and regulated stop codon readthrough in Drosophila melanogaster. Elife. 2013;2: e01179.

139. Loughran G, Chou MY, Ivanov IP, et al. Evidence of efficient stop codon readthrough in four mammalian genes. Nucleic Acids Res. 2014;42:8928–8938.

140. Gordon AJE, Halliday JA, Blankschien MD, Burns PA, Yatagai F, Herman C. Transcriptional infidelity promotes heritable phenotypic change in a bistable gene network. PLoS Biol. 2009;7:e44.

141. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nat Rev Genet. 2009;10:691–703.

142. Cooke SL, Shlien A, Marshall J, et al. Processed pseudogenes acquired somatically during cancer development. Nat Commun. 2014;5:3644.

143. Sung W, Ackerman MS, Dillon MM, et al. Evolution of the insertion-deletion mutation rate across the tree of life. G3. 2016;6(8):2583–2591.

144. Curran JF, Yarus M. Base substitutions in the tRNA anticodon arm do not degrade the accuracy of reading frame maintenance. Proc Natl Acad Sci U S A. 1986;83:6538–6542.

145. Jørgensen F, Kurland CG. Processivity errors of gene expression in *Escherichia coli*. J Mol Biol. 1990;215:511–521.

146. Arava Y, Boas FE, Brown PO, Herschlag D. Dissecting eukaryotic translation and its control by ribosome density mapping. Nucleic Acids Res. 2005;33:2421–2432.

147. Mejlhede N, Atkins JF, Neuhard J. Ribosomal −1 frameshifting during decoding of Bacillus subtilis cdd occurs at the sequence CGA AAG. J Bacteriol. 1999;181:2930–2937.

148. Clare JJ, Belcourt M, Farabaugh PJ. Efficient translational frameshifting occurs within a conserved sequence of the overlap between the two genes of a yeast Ty1 transposon. Proc Natl Acad Sci U S A. 1988;85:6816–6820.

149. Malagnac F, Fabret C, Prigent M, Rousset J-P, Namy O, Silar P. Rab-GDI complex dissociation factor expressed through translational frameshifting in filamentous ascomycetes. PLoS One. 2013;8:e73772.

150. Belew AT, Meskauskas A, Musalgaonkar S, et al. Ribosomal frameshifting in the CCR5 mRNA is regulated by miRNAs and the NMD pathway. Nature. 2014;512:265–269.