

# Chapter 14

## Sequences Promoting Recoding Are Singular Genomic Elements

Pavel V. Baranov and Olga Gurvich

**Abstract** The distribution of sequences which induce non-standard decoding, especially of shift-prone sequences, is very unusual. On one hand, since they can disrupt standard genetic readout, they are avoided within the coding regions of most genes. On the other hand, they play important regulatory roles for the expression of those genes where they do occur. As a result, they are preserved among homologs and exhibit deep phylogenetic conservation. The combination of these two constraints results in a characteristic distribution of recoding sequences across genomes: they are highly conserved at specific locations while they are very rare in other locations. We term such sequences singular genomic elements to signify their rare occurrence and biological importance.

### Contents

14.1	Singular Genomic Elements . . . . .	302
14.2	Sequences Promoting Ribosomal Frameshifting as Singular Genomic Elements . .	303
14.2.1	+1 Frameshifting Cassette in Bacterial Release Factor 2 mRNA . . . . .	303
14.2.2	-1 Frameshifting Cassette in Coronavirus Polyprotein-Encoding Gene . .	308
14.3	Cars and Ribosomes, Fast and Furious: Role of mRNA in the Accuracy of Translation	309
14.4	Strategies for Searching Recoding Cases as Singular Elements . . . . .	311
14.5	Possible Functions of Products Generated by Low-Level Aberrant Translation . .	315
14.6	Conclusions . . . . .	316
	References . . . . .	317

In this chapter we describe the distribution of several known shift-prone patterns and stimulatory signals and how they relate to the concept of singular genomic elements. We also discuss how their characteristic distribution can be utilized for identification of novel recoded genes and describe studies where such strategies have been employed.

---

P.V. Baranov (✉)  
Biochemistry Department, University College Cork, Ireland  
e-mail: brave.oval.pan@gmail.com

## 14.1 Singular Genomic Elements

A characteristic property of all biological systems is diversity and specialization of their component parts that play distinctive functional roles. The tendency for specialization and uniqueness is profound on the genomic level, as the existence of identical multiple copies of the same gene (unless they relate to mobile elements) is rare. Functional specialization of gene products demands similarly specific regulation of their biosynthesis and processing. Such specificity can be achieved through a combinatorial effect of several regulatory mechanisms acting on different levels of gene expression – from initiation of transcription to posttranslational modifications, where similar regulatory sequences occur in groups of functionally related genes. Specificity is gained through differential combination of these sequences which could be idiosyncratic for a particular gene. However, it is attractive to imagine a simpler scheme where a unique regulatory element would be responsible for the regulation of a specific gene. Such a sequence could respond to changes in particular cellular conditions associated with expression of the regulated gene and so provide feedback control. Indeed such regulatory elements are known and they are characteristically distributed across genomes. Their occurrence at random locations in a single genome is avoided while their occurrence at specific genomic locations across several species is preserved. Such a distribution is easy to explain. Suppose we have a genomic feature  $F$  that specifically regulates expression of a gene  $G$ . The feature  $F$  then should be avoided in all locations where it may have an effect on expression of genes other than  $G$ . On the other hand, since association of the feature  $F$  with the gene  $G$  is beneficial for the organism, such association would likely be preserved during speciation and therefore it will occur in orthologs of the gene  $G$ . In other words these regulatory sequence elements are avoided and hence underrepresented across a single genome or across particular types of sequences (e.g., those coding for proteins) in a single genome. However, they are present in orthologous genes from multiple related organisms. Here we introduce the term *singular genomic element* to denote such elements. There are a number of different biologically active nucleotide sequences that exhibit properties of singular genomic elements. Examples are unique sites of restriction, sites encoding unique protease cleavage sites, cases of transcriptional slippage discussed in Chapter 19, or even miRNA targets (Farh et al., 2005). Nonetheless, perhaps, the most striking type of known singular genomic elements is sequences promoting recoding events. Such elements interfere with standard genetic decoding and increase the chances of erroneous translation. Thus, their occurrence in the protein coding sequence of most genes is detrimental. At the same time they do play important roles in those genes that utilize non-standard decoding in their expression and consequently undergo purifying selection during evolution in their corresponding locations. In this chapter we discuss different examples of sequences implicated in recoding events and their distribution across different regions of single genomes and among orthologous genes. We also discuss how searches for singular genomic elements could be used as a strategy for identification of new cases of recoding events and novel genes that are expressed via recoding mechanisms.

## 14.2 Sequences Promoting Ribosomal Frameshifting as Singular Genomic Elements

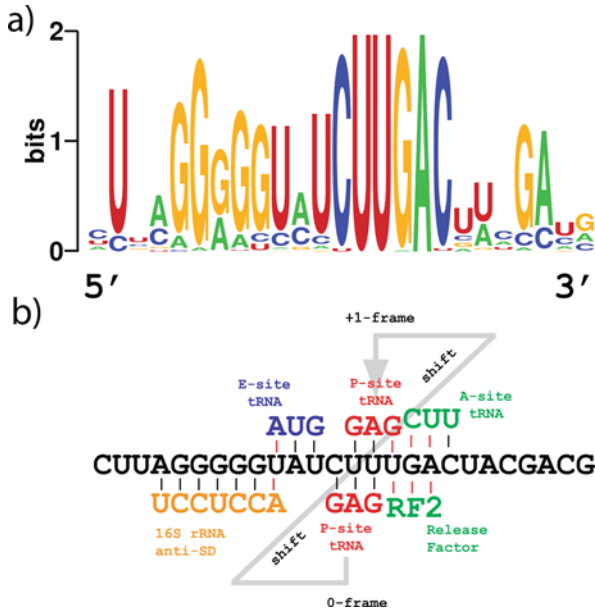
### 14.2.1 +1 Frameshifting Cassette in Bacterial Release Factor 2 mRNA

The *Escherichia coli* gene *prfB* encodes release factor 2 (RF2) and was among the first discovered chromosomal genes requiring programmed ribosomal frameshifting for their expression (Craigen et al., 1985; Craigen and Caskey, 1986). In bacteria, two class-I release factors are responsible for recognition of codons specifying termination of translation, RF1 and RF2 (reviewed Kisselev and Buckingham, 2000). These factors are semi-specific, they both recognize UAA stop codons. UAG is recognized exclusively by RF1, while UGA recognition is specific to RF2 (Scolnick et al., 1968; Capecchi and Klein, 1970). In *E. coli* and most (~87%) other bacteria, RF2 is encoded in two overlapping ORFs (Kisselev and Buckingham, 2000; Bekaert et al., 2006). While the main portion of RF2 protein is encoded in the second long ORF (see Fig. 14.1), this ORF does not have its own translation initiation site. Initiation of translation takes place at the start of first short ORF whereas the second ORF can be translated only if elongating ribosomes shift reading frame in the +1 direction at the end of the first ORF. The nucleotide sequence and its conservation across RF2 genes from multiple bacterial species are illustrated in Fig. 14.2. The shift cassette consists of a number of modular elements that are responsible for the stimulatory effects on frameshifting. The ribosomal frameshifting itself takes place at a CUU codon followed by U, i.e., CUU\_U (where the underlined space separates codons). When the CUU codon is located at the P-site, tRNA<sup>Leu</sup> repositions relative to mRNA by shifting 1 nucleotide toward the 3' -end (+1 frameshift) so that its anticodon forms base pairs with the overlapping UUU codon. Consequently the new frame now corresponds to C\_UUU (Curran, 1993). As can be seen from Fig. 14.2, this sequence is nearly universally conserved with the exception of the first nucleotide C, where in some bacteria it is U. In those bacteria, the repositioning tRNA is likely to be tRNA<sup>Phe</sup> which shifts from one Phe codon to another so that transition of the reading frame occurs from UUU\_U to U\_UUU.

The second modular element that exhibits similarly astonishing conservation is the stop codon that overlaps the frameshift site (Fig. 14.2). The stop codon is nearly

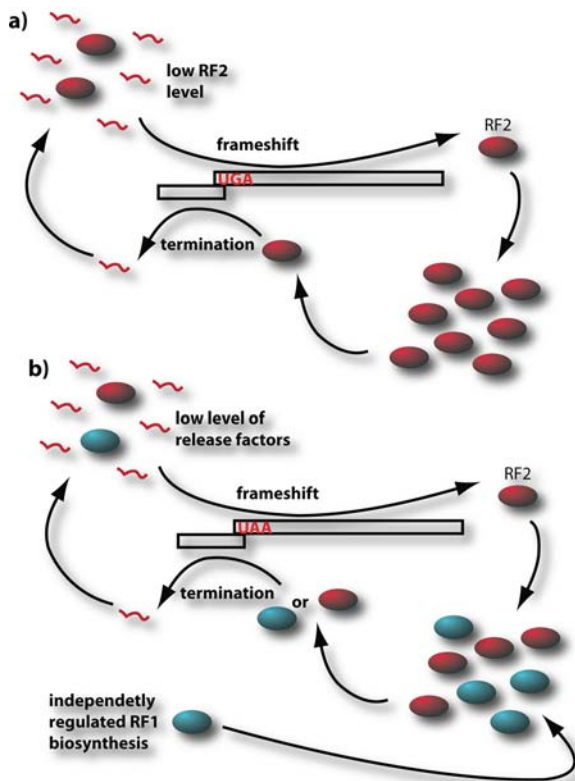


**Fig. 14.1** FSfinder2 (Moon et al., 2007) plot of ORF organization in the *E. coli* gene encoding release factor 2 (RF2). The coding sequence is highlighted in pale yellow and consists of a first short ORF in the “0” frame and a long second ORF in the “+1” frame



**Fig. 14.2** (a) Diagram of RF2 frameshift site conservation, the height of symbols indicates conservation of nucleotides, while their weight shows the relative frequency of nucleotides at corresponding positions. The diagram was built using WebLogo3 (Crooks et al., 2004), sequences of RF2 frameshift sites were obtained using ARFA (Bekaert et al., 2006). (b) Sequence of *E. coli* K12 RF2 frameshift site aligned to the diagram above. Interactions with different ligands which play roles in the ribosomal frameshifting are indicated with vertical strokes, red strokes correspond to competing interactions

always UGA (with few exceptions where it is UAA). This stop codon is the key element responsible for sensitivity of frameshifting efficiency to the cellular concentration of RF2. When ribosomes approach the end of the first ORF and the stop codon occupies the ribosomal A-site, either of two major events occur: termination of translation or +1 slippage of P-site tRNA which directs translation to the longer ORF. These two events are in competition, so that increasing termination efficiency results in decreasing frameshifting efficiency and vice versa. As termination efficiency is directly influenced by the concentration of release factors, frameshifting efficiency also depends on the concentration of release factors. Since UGA is not recognized by RF1, frameshifting efficiency is solely dependent on the concentration of RF2. This mechanism creates an elegant regulatory feedback loop, as illustrated in Fig. 14.3a, where the level of RF2 biosynthesis depends on the cellular concentration of RF2. With those cases where the stop codon at the frameshift site is UAA, it is likely that frameshifting senses the cumulative concentration of both factors, as illustrated in Fig. 14.3b. Such indiscriminate sensing of the concentration of both release factors could be beneficial as well, since in this case



**Fig. 14.3** Regulatory feedback provided for RF2 biosynthesis by the frameshifting mechanism. (a) The first ORF has a UGA stop codon. The regulation is autonomous and the level of RF2 biosynthesis depends on its own concentration. (b) The first ORF has a UAA stop codon. The level of RF2 biosynthesis depends on the concentration of both release factors, RF1 and RF2. RF1 and RF2 likely compensate for each other

low cellular concentration of RF1 may be compensated for by increased synthesis of RF2. Perhaps this is particularly beneficial for those bacteria where UGA and UAA codons are more frequently used than UAG codons. However, whether there is indeed a correlation between occurrence of UAA in the RF2 frameshifting cassette and differential utilization of stop codons in the corresponding bacterial genomes has not been investigated.

All other stimulatory elements in the RF2 frameshifting cassette are not responsible for the sensitivity of frameshifting to release factor concentration. However, they are responsible for elevation of the absolute level of frameshifting efficiency, which in their absence would be insignificant even at low concentrations of release factors. The element whose role in the frameshifting mechanism is relatively easy to understand is the identity of the nucleotide 3' adjacent to the stop codon. Unlike all sense codons that are recognized by RNA molecules via complementary interactions, stop codons are recognized by protein molecules. The recent analysis of

crystal structure of the ribosome complex with RF2 reveals details of RF2 interactions with the UGA stop codon in mRNA (Weixlbaumer et al., 2008). Unfortunately, the crystal structure does not provide information on interactions of RF2 with the mRNA region downstream of the stop codon which seems to interact with release factors as evident from earlier cross-linking studies (Poole et al., 1998). While these interactions do not play a role in stop codon discrimination, they do affect termination efficiency. Since frameshifting efficiency negatively correlates with termination efficiency, it is not surprising that the weakest termination context has been selected in the RF2 frameshift site during its evolution (Major et al., 1996). It can be seen in Fig. 14.2 that the 3' nucleotide adjacent to the stop codon is nearly always C, which has been shown to be the most inefficient context codon for termination in eubacterial organisms (Mottagui-Tabar and Isaksson, 1998; Pavlov et al., 1998).

Another important stimulatory element in the RF2 frameshifting cassette is the internal Shine–Dalgarno (SD) sequence located upstream of the shift site (Weiss et al., 1987, 1988; Curran and Yarus, 1988). Normally SD sequences are used for the initiation of translation in bacteria and are located upstream of initiator codons (Shine and Dalgarno, 1975). The increase in local concentration of initiating ribosomes around initiator sites is achieved through interactions between the SD and the corresponding complementary region of 16S rRNA, termed anti-Shine–Dalgarno (anti-SD). The internal SD 5' of the frameshifting site could serve the same purpose, and initiation of translation at the UUG codon (which is a part of the frameshifting site) has been demonstrated (Baranov et al., 2002), although no potential functional role for this internal initiation event has been implicated. It could be that this is simply an unintentional side effect caused by sequence constraints of the RF2 frameshifting cassette. Irrespective of internal initiation, the main role of the internal SD is clearly to target elongating ribosomes. One particular important aspect of the SD stimulatory effect on frameshifting efficiency is the location of the SD relative to the frameshift site (Weiss et al., 1987). The length of the spacer between the SD sequence and the P-site tRNA during the frameshift is shorter than the distance between the SD and initiator codons (Ma et al., 2002). It is reasonable to assume that the distance between an SD and an initiator codon is optimal for the relaxed conformation of the ribosomal RNA during the initiation. If so, the shorter distance between the internal SD and the shift site should create tension in the ribosomal RNA between the anti-SD and the decoding center of the ribosome. Such tension likely acts in a manner of a compressed spring, whose relaxation is achieved by a progressive movement of tRNA with the decoding center of the ribosome toward the 3'-end of mRNA. This movement would explain the stimulatory effect of an SD on +1 frameshifting. Accordingly it is known that an internal SD stimulates frameshifting in the opposite direction when the spacer is longer than the optimal for initiation, in which case RNA likely acts as a stretched spring that alleviates tRNA movement toward the 5'-end of mRNA (Atkins et al., 2001). The conservation of the SD sequence and its location is illustrated in Fig. 14.2. Since base pairing between the SD and rRNA does not have to be perfect to cause the effect, there is a certain degree of flexibility in the RF2 frameshift stimulatory SD sequences; hence, its conservation is less profound than that of the shift site and the stop codon.

While the size of the spacer separating the shift site from the internal SD sequence is crucially important for its stimulatory effect, the identity of the spacer is not inconsequential either (Baranov et al., 2002). During frameshifting the spacer corresponds to the codon located in the ribosomal E-site. It has been suggested that there is a competition between the anti-SD and E-site tRNA for interactions with the corresponding part of mRNA. This interference of the SD with normal occupation of the E-site codon by the E-site tRNA affects fidelity of the ribosome (Baranov et al., 2002; Marquez et al., 2004; Sanders and Curran, 2007). Consequently, as the affinity of different tRNAs for the E-site fluctuates (Lill and Wintermeyer, 1987), it is not surprising that the identity of the spacer affects frameshifting efficiency.

Analysis of the distribution of sequences similar to the RF2 frameshifting cassette in bacterial genomes in terms of its “singularity” is meaningless, due to its size and complexity. If we represent the RF2 frameshift cassette as some kind of a roughly estimated consensus sequence such as GRGGNNYTT-Stop-C, the probability of its appearance in random sequences of the same length is  $1/16,384$ . Since we are interested only in those stop codons that are really used for the termination of translation, then the probability of such a sequence in a genome similar to *E. coli* (~4,000 genes) will be about 0.2 and the probability of two such sequences in such a genome will be only ~0.05. Even if a deviation of a single nucleotide in the above consensus sequence is allowed, the probability of two random occurrences of such sequences in a genome of a size similar to that of *E. coli* would be less than  $1/2$ . In other words, the fact that the above consensus sequence does not occur at the end of any other *E. coli* gene does not indicate evolutionary selection against such sequences. As for the individual modular stimulatory signals constituting the RF2 frameshifting cassette, they are insufficient to trigger ribosomal frameshifting with comparable efficiency and hence they are relatively frequent in the genomes. Nonetheless, some tendency for their avoidance can be illustrated using the following simple and perhaps somewhat naïve measures. For example, while C nucleotides constitute a 0.25 fraction of the *E. coli* K12 genome (NC\_000913), the fraction of C nucleotides adjacent to the 3'-end of *E. coli* stop codons is 0.17, and 0.14 for those adjacent to UGA, whereas the portion of Cs after any UGA trinucleotide in the *E. coli* genome ( $N_{TGAC}/N_{TGAN}$  ratio) is 0.22. This seeming underrepresentation of Cs after stop codons and UGA in particular is, of course, due to its weakening effect on termination of translation. A similar tendency could be sensed for the usage of a codon upstream of stop codons. For example, the proportion of UUU codons among all Phe codons in the *E. coli* K12 genome is 0.66. But the proportion of UUU codons among Phe codons that are located upstream of stop codons is 0.47 and only 0.24 upstream of UGA codons. For CUU similar calculations give the less profound corresponding values of 0.16, 0.17, and 0.13. There is no avoidance of SD-like sequences at the end of *E. coli* genes compared to other locations within mRNA coding sequences. On the contrary, analysis of a larger number of bacterial genomes suggests that SD sequences are even over-represented at the end of coding sequences, perhaps due to translational coupling where such SD sequences are used for the initiation of downstream genes (PVB, unpublished).

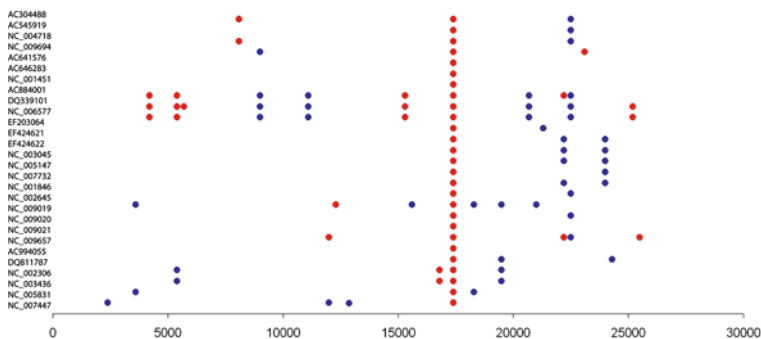
Summarizing, the entire RF2 frameshifting cassette constitutes a relatively large and complex constrained sequence pattern whose random occurrence in small genomes, such as the one in *E. coli*, has a low probability. Smaller and simpler components of the frameshift cassette are relatively ineffective in triggering efficient non-standard translation events; nonetheless they probably can increase the chance of errors and thus some level of selection against such sequences can be detected. In the following section we deal with the analysis of relatively short sequences, so their random occurrence is considerably more likely. Despite their shortness, however, they are sufficient to trigger efficient non-standard translational events.

### ***14.2.2 –1 Frameshifting Cassette in Coronavirus Polyprotein-Encoding Gene***

The coronaviral gene encoding the ORF1AB polyprotein consists of two overlapping ORFs and the synthesis of the full length protein product requires programmed ribosomal –1 frameshifting (Brierley et al., 1989). The frameshift cassette consists of the slippery heptamer sequence U\_UU.U\_AA.C (underlined spaces indicate separation of codons in the initial phase and dots separate codons in the frame after the shift). The frameshifting is stimulated by RNA structures downstream of the slippery sequence. There is a degree of variation among the stimulatory structures. In some viruses the structure is formed by two distant stem loops forming complementary interactions between their apical loops (kissing stem-loop structures) (Herold and Siddell, 1993). In others, it is a classical H-type pseudoknot with variable features, for example, in SARS-CoV there is an important RNA stem-loop structure located within the second loop of the pseudoknot (Baranov et al., 2005; Plant et al., 2005; Su et al., 2005).

Although the presence of a structure is evident in all known coronaviruses and is likely essential to support functional frameshifting efficiency, even in its absence frameshifting is detectable at levels greatly exceeding the average background frequency of frameshift errors (Brierley et al., 1991). The distribution of U\_UU.A\_AA.C sequences within a 27-way alignment of selected coronaviruses is shown in Fig. 14.4. Apparently there is no strong selection against such sequences in coronaviral genomes. Based on combinatorial codon usage analysis of these representative coronaviral genomes U UU.A AA.C patterns are expected to occur about two times per ORF1AB gene. Indeed the real number of patterns corresponds to this expectation value and varies from 1 to 6 per gene (Fig. 14.4). Nonetheless, the overall distribution clearly illustrates the behavior typical of singular genetic elements where U UU.A AA.C is present in all genomes in a particular location, while other occurrences are distributed in a more random manner. For comparison Fig. 14.4 also shows the distribution of the same nucleotide patterns, but in different reading phases. It is clear that their distribution is less ordered. The existence of U\_UU.A\_AA.C patterns in locations other than the frameshift site can be explained either by neighboring nucleotide context disfavoring ribosomal frameshifting or by the possibility that such low frameshifting levels (in the absence of a stimulator) at a few locations can be tolerated by viruses.





**Fig. 14.4** Distribution of UUUAAC patterns across multiple alignments of coronavirus *orfAB*. Red spots correspond to the patterns in the shift-prone phase U\_UUA\_AAC, blue spots correspond to the same pattern in other reading phases. The sequences for the alignment were extracted from the CoVDB (Huang et al., 2008). Genbank accession numbers are given within the figure. Initially, the nucleotide sequences were translated and aligned with ClustalW and then the alignment obtained was back-translated and processed with custom-designed Perl scripts

### 14.3 Cars and Ribosomes, Fast and Furious: Role of mRNA in the Accuracy of Translation

One striking difference between erroneous frameshifting and programmed frameshifting lies in their efficiencies. The translational apparatus is able to decode mRNA with remarkable accuracy; misincorporation of an amino acid due to recognition of incorrect tRNAs occurs with frequencies in the range of  $10^{-3}$ – $10^{-5}$  depending on the exact type of error. These estimates come from a number of studies in *E. coli*, reviewed in Parker (1989). This high accuracy for amino acid incorporation is observed despite the fact that not all such errors are necessarily harmful, since substitution of a single amino acid in a protein does not necessarily lead to its inactivation. The extent of tolerance to misincorporation errors is best illustrated by *Candida albicans* where CUG codons are decoded as both Leu and Ser due to ambiguous aminoacylation of the corresponding tRNA (Moura et al., 2007). In contrast, errors in processivity, such as frameshift errors, pose a greater danger during translation since they result in alterations not of just a single amino acid but of the entire sequence following such an error. It is reasonable to expect that the decoding apparatus should be able to prevent such errors with even greater accuracy. Indeed, it has been estimated that background levels of frameshifting errors fluctuate in the range of  $10^{-5}$ – $10^{-7}$  (Kurland, 1979; Parker, 1989). At the 2007 ribosomal meeting in Cape Cod, Mons Ehrenberg summarized his talk with the following statement: “Ribosomes are very fast and very accurate and this is the summary of my talk.” It would be hard and perhaps juvenile to argue with such a statement as it would be hard to argue with commercials advertising modern cars saying that they are fast and safe. Cars are, but the traffic is not, at least not always. The safety and speed of traffic depends not only on cars but also on road conditions. By analogy we can describe mRNAs as the roads for the ribosomal traffic. We will argue that the

observed accuracy of translation relies not only on the properties of the ribosome but also on mRNA sequence. Under certain circumstances mRNA can force translating ribosomes to alter their behavior so that translation can no longer be considered accurate.

Frameshifting occurs with strikingly high efficiencies at certain recoding sites exceeding background levels by  $10^6$  and under certain conditions could be even more efficient than standard triplet translation. Of course, such efficiency is frequently achieved by an ensemble of complex stimulatory signals that have evolved to increase frameshifting efficiency at a local site. This was described above for RF2 mRNA frameshifting and is also evident from many other examples throughout this book. However, even relatively simple sequences such as the heptameric C.UU\_A.GG\_C in yeast transposon Ty1 cause frameshifting with efficiency comparable to that of standard translation at the same site without additional stimulators (Belcourt and Farabaugh, 1990). Other simple short sequences are also shift-prone and can lead to frameshifting events of lower efficiency, but still much greater than the background levels. Evidently the accuracy of translation in terms of reading frame maintenance is highly dependent on mRNA nucleotide context. Why is there such dependence and why do ribosomes not translate all sequences with a similar accuracy?

A plausible explanation may lie in the fact that the ribosomes as we know them have evolved to achieve the global optimum compromise between speed and fidelity of translation (Kurland et al., 1996). It is possible to increase fidelity of the ribosome by introducing certain mutations leading to hyper-accurate ribosomes. However such improved accuracy has a cost, the speed of translation is reduced, and this diminishes the potential benefit from the higher accuracy of translation. Hyper-accurate ribosomes are usually streptomycin dependent, as addition of streptomycin presumably increases the speed of translation by decreasing its accuracy. Can the ribosome be modified further to increase the speed of translation without losing accuracy? The potential for further improvement lies in mRNA sequences and the set of tRNAs used to decode them. The solution is alteration of the codon bias and the set of unequally distributed tRNAs. To understand how this could help improve both accuracy and speed consider a simple model. The probability of incorporation of a particular tRNA<sub>*k*</sub> at a particular codon<sub>*k*</sub> competing with a set of *N* tRNAs can be represented as

$$\frac{a_k T_k}{\sum_{i=1}^N a_i T_i}$$

where *a* is the tRNA affinity toward codon<sub>*k*</sub> in the ribosomal A-site and *T* is its local concentration. An increase of tRNA<sub>*k*</sub> concentration will increase the probability of its incorporation at codon<sub>*k*</sub> as will a decrease in concentration of other tRNAs, even though their affinities (that are partially determined by the ability of the ribosome to discriminate between them) remain the same. If all codons were distributed equally in mRNA sequences, there would be no benefit from such a manipulation. However, a global positive effect can be achieved if there is codon usage bias, with some

codons being abundant and others being rare. In this case, corresponding manipulation of the set of tRNAs will lead to improved accuracy and speed of global translation. But this would not come without a cost: decreased accuracy and speed of translation of rare codons. Of course the above scenario is a simplification compared to the real situation since the affinity of the tRNAs to their codons is also context dependent. Further, for frameshifting errors, the probability of its occurrence depends also on the specific combination of codons in the ribosome and the probability of rearrangement of tRNAs in the ribosome relative to mRNA (Baranov et al., 2004; Liao et al., 2008). Consequently a biased occurrence of combinations of codons is also evident (Fedorov et al., 2002; Moura et al., 2007). These simple considerations illustrate the concept of how biases in codon usage and their combination can be used for the benefit of global translation accuracy and speed. This, of course, does not mean that such biases exist purely to increase the efficiency of global translation. There are other, perhaps even more important contributing factors, such as GC content, biases in the usage of amino acids, mutational bias (Bernardi and Bernardi, 1986; Wan et al., 2004). In fact it has been possible to predict codon usage biases for a hundred microbial organisms purely based on a combination of GC content and nucleotide mutational bias, obtained from the analysis of intergenic regions (Chen et al., 2004). However, irrespective of the evolutionary reasons underlying the existence of codon bias, there is a relationship between codon bias and relative tRNA abundance (Ikemura, 1981). It is clear that the translational apparatus, for at least the set of tRNAs used for mRNA decoding, has adapted to these biases, since higher codon biases associate with conserved and highly expressed genes (Stoletzki and Eyre-Walker, 2007). Such adaptation results in decreased accuracy of translation of mRNAs that do not show bias, as is evident from highly erroneous expression of heterologous translation (Kurland and Gallant, 1996), e.g., during synthesis of human proteins in bacterial species whose translational apparatus has not been modified specifically for such purposes (Gustafsson et al., 2004). Moreover, accuracy of translation depends not only on simple codon bias but also on a bias among co-occurring codons in mRNA. This fact has been recently utilized to design a synthetic poliovirus whose genome was modified to encode native capsid protein with CDS consisting of underrepresented codon pairs (Coleman et al., 2008). Such virus triggers host immune response, but reduced translation rates alter virus viability, suggesting an elegant method for immunization.

This explains why certain relatively simple sequences can be particularly prone to frameshift errors and why they are rare in most coding regions. However, the situation is not always so simple as we will see in the following sections.

## 14.4 Strategies for Searching Recoding Cases as Singular Elements

A number of studies have attempted to search for new cases of programmed frameshifting based on the assumption that the sequences that promote ribosomal frameshifting should behave like singular genomic elements and as such be avoided

in the coding regions unless the triggered frameshifting is positively selected for. The simplest idea is to search for further occurrences of sequences, of the type known to be utilized for programmed ribosomal frameshifting, throughout the coding regions of completed genomes. Although this approach limits the search to motifs already known to trigger frameshifting and will not increase our knowledge of frameshift-prone sequences, it could reveal novel cases of utilization of these sequences for gene expression purposes. To analyze the frequency of occurrence of sequences capable of stimulating  $-1$  frameshifting in *Saccharomyces cerevisiae*, Jacobs et al. (2007) searched for viral consensus slippery sites X<sub>XX</sub>.Y<sub>YY</sub>.Z, where XXX represents any three identical nucleotides, YYY represents AAA or UUU, Z ≠ G. With this approach they identified 10,340 slippery sites in the 6,353 annotated coding sequences of the yeast genome, 6,016 of which are followed by at least one pseudoknot motif. According to statistical analyses employed by the authors these signals are underrepresented in the *S. cerevisiae* genome. Of the 6,353 yeast ORFs, 1,275 contain at least one strong and statistically significant  $-1$  frameshift signal [in a recent study Theis et al. (2008) have argued that in some cases there are alternative structures that are more stable than the predicted pseudoknots]. Eight out of nine sequences, selected for experimental verification using artificial genetic constructs, supported efficient levels of frameshifting in vivo. The authors hypothesized that many other frameshift candidates found in their study could lead to significant levels of frameshifting. If frameshifting indeed takes place at those locations, in the vast majority of cases it would result in production of truncated and most likely dysfunctional products. The authors hypothesized that the role of frameshifting could be regulatory (see the following section). It is unclear how beneficial such a regulation might be for the cells and no data on phylogenetic conservation of these sequences have been provided.

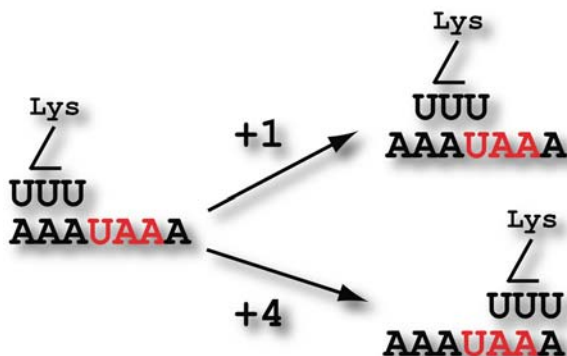
In a different work (Gurvich et al., 2003), the *E. coli* K12 genome was searched for occurrences of the very well-known prokaryotic slippery sequence A<sub>AA</sub>.A<sub>AA</sub>.G. Frameshifting at A<sub>AA</sub>.A<sub>AA</sub>.G is utilized for expression of the  $\gamma$  subunit of DNA polymerase III, while the  $\tau$  subunit is expressed by standard translation from the same gene (*dnaX*) (Blinkowa and Walker, 1990; Flower and McHenry, 1990; Tsuchihashi and Kornberg, 1990). Frameshifting at this sequence is also utilized by a number of insertion sequence elements in *E. coli* (Hu et al., 1996; Baranov et al., 2006). Seventy instances of this sequence have been found in 68 *E. coli* genes. Twelve genes have been chosen for experimental analysis and all of them have been shown to support  $-1$  frameshifting at levels above background. The authors used comparative phylogenetic analysis to address potential utilization of any of those sequences for gene expression purposes. Apart from the *dnaX* gene, six IS2-like elements and the *ydaY* gene of unknown function, utilize A<sub>AA</sub>.A<sub>AA</sub>.G for gene expression. Although the number of occurrences is quite high, according to the statistical analysis this sequence is underrepresented in coding regions, and thus does behave as a singular element. The distribution of three other known shift-prone sequences in *E. coli* K12, CCC<sub>UGA</sub> (Gurvich et al., 2003), AGG<sub>AGG</sub>, and AGA<sub>AGA</sub> (Gurvich et al., 2005), was also examined. All three sequences trigger  $+1$  frameshifting in *E. coli*. Frameshifting at C<sub>CC</sub>.U<sub>GA</sub> occurs

through near-cognate recognition of the CCC codon by tRNA<sup>Pro</sup><sub>5'U\*GG3'</sub> (where U\* designates the cmo<sup>5</sup>U34 modification) (O'Connor, 2002). Because of suboptimal base pairing with the CCC codon, this tRNA is prone to shift into the +1 frame to re-pair to mRNA at the cognate CCU codon. As with RF2 mRNA frameshifting, that on C.CC\_U.GA is in direct competition with termination mediated by RF2 and its efficiency is increased due to slow decoding of the termination codon. Although not known to be utilized for gene expression in *E. coli*, frameshifting at C.CC\_U.GA is employed for expression of antizyme genes in some eukaryotes (Ivanov and Atkins, 2007) and for expression of the *tsh* gene of *Listeria monocytogenes* phage PSA (Zimmer et al., 2003). Nineteen genes in *E. coli* K12 end with C.CC\_TGA and in half of them frameshifting occurs at above 1% (Gurvich et al., 2003).

Frameshifting on A.GG\_A.GG and A.GA\_A.GA is due to limited abundance of the cognate arginine tRNA<sup>Arg</sup><sub>3'UCC5'</sub> and tRNA<sup>Arg</sup><sub>3'UCU\*5'</sub> (where U\* is 5-methylaminomethyl-2-thiouridine), respectively. Due to sequestration of the sparse tRNA by the first of the tandem codons, its availability for the second codon is drastically reduced. When the second codon occupies the A-site of the translating ribosome the longer-than-usual time for arrival of the cognate tRNA increases the chance for dissociation of the peptidyl-tRNA which may re-pair to mRNA in the overlapping +1 frame (or potentially -1 frame as has been shown for an A.GA\_A.GA tandem by Lainé et al. (2008)). Frameshifting to the new frame is greatly favored by availability of the tRNA cognate to the new codon in the +1 frame. The A.GG\_A.GG and A.GA\_A.GA tandems were originally reported to trigger up to 50% frameshifting (Spanjaard and van Duin, 1988; Spanjaard et al., 1990). Although such high levels of frameshifting are likely due to overexpression of the mRNAs containing these sequences (Gurvich et al., 2005) and due to the use of streptomycin-resistant strains, in which ribosomes translate the mRNA more slowly making them prone to +1 frameshifting at the rare codons (Sipley and Goldman, 1993). Nevertheless, even at the lowest possible expression level of the transgene, frameshifting at A.GA\_A.GA (and likely A.GG\_A.GG) occurs at about 1% level (Gurvich et al., 2005). All three frameshift-prone sequences C.CC\_U.GA, A.GG\_A.GG, and A.GA\_A.GA are not underrepresented in *E. coli* and in fact C.CC\_U.GA is significantly overrepresented. However, none of these sequences including A\_AA.A\_AA.G, occur in the subset of highly expressed genes in *E. coli* (Karlin et al., 2001). This means that although not significantly underrepresented in coding regions, overall these sequences are selected against in highly expressed ORFs and in the way they behave as singular elements in highly expressed genes. In contrast to the Jacobs et al. study, Gurvich et al. suggested that the occurrence of these frameshift candidates in protein coding regions does not have a functional role, since they do not exhibit phylogenetic conservation. Gurvich et al. argued that frameshifting above background level in lowly expressed genes could easily be tolerated by cells, since only a few aberrant protein molecules would be produced as a result of frameshifting. Therefore, the presence of shift-prone sequences in certain locations can be explained not by their beneficial effects but by the lack of strong selection against such sequences. Future studies are expected to resolve the contrasting interpretations.

The most general *ab initio* study related to singular elements supporting frameshifting was performed by Shah et al. (2002) where the distribution of all heptamers occurring in coding regions of the yeast *S. cerevisiae* genome was analyzed. A fraction of the least abundant and the most underrepresented heptamers have been tested for their ability to trigger ribosomal frameshifting. All sequences tested stimulated ribosomal frameshifting at above background levels with some of them promoting highly efficient frameshifting. Notably, the heptamer sequences C.UU\_A.GU\_U and C.UU\_A.GG\_C used to trigger programmed ribosomal frameshifting for expression of EST3 (Morris and Lundblad, 1997; Taliaferro and Farabaugh, 2007) and ABP140 (Asakura et al., 1998), respectively, are ranked among the least represented in coding regions of *S. cerevisiae*. While this approach appeared to have good predictability for sequences supporting +1 frameshifting in yeast, it failed in predicting sequences that would stimulate -1 frameshifting. The authors suggested this could be because the sequences utilized for -1 programmed frameshifting in yeast do not stimulate frameshifting at sufficiently high efficiency without additional *cis*-acting elements.

Frameshift-prone sequences do not necessarily exhibit properties of singular elements. In certain organisms frameshifting could be highly abundant. This seems to be the case in the ciliate *Euplotes*. To date there are eight different types of genes identified in *Euplotes* that utilize +1 frameshifting for their expression (Klobutcher, 2005). Only about 90 genes have been sequenced in *Euplotes* and the current estimate is that about 10% of the *Euplotes* genes require frameshifting for expression. Interestingly, three of these genes require multiple frameshift events for expression. Of these eight genes, five encode enzymes, one encodes a protein associated with the RNA component of telomerase, and two have unknown function. None of these genes is expected to be highly expressed, even though the subset of sequenced genes in *Euplotes* is biased toward highly expressed genes. All genes share the same sequence A.AA\_U.AA\_A (A.AA\_U.AG\_A for one gene) within the overlap of the upstream and downstream ORFs. Thus, it is likely that frameshifting takes place at these sequences and its mechanism is likely the same for all genes. The frameshift propensity of A.AA\_U.AA\_A heptanucleotide in *Euplotes* entails inefficient translation termination at the UAA stop codon and slippage of the tRNA<sup>Lys</sup> from the AAA codon to AAU. Ineffective translation termination at the UAA codon in *Euplotes* is proposed to be linked to UGA stop codon reassignment to cysteine (Klobutcher and Farabaugh, 2002; vallabhaneni et al., 2009). Such reassignment is complemented by changes in eukaryotic release factor one (eRF1), so that it no longer recognizes UGA codons. However, such changes may have rendered eRF1 less potent in recognition of UAA and UAG stop codons as well. If so, translation termination in *Euplotes* might be generally slow and inefficient, consequently favoring a competing process of +1 frameshifting. As a result it might be that *Euplotes* is tolerant of arising frameshift mutations that would be compensated by +1 frameshifting on A.AA\_U.AR\_A. Conservation of the 3' A adjacent to the stop codon is believed to weaken the stop codon as a signal for termination and the conservation of AAA 5' of the stop codon is explained by an unknown special feature of the corresponding tRNA<sup>Lys</sup> that makes it shift prone compared to other



**Fig. 14.5** Tentative alternative mechanisms of frameshifting in the ciliate *Euplotes*. The stop codon in the frameshift site is shown in red. tRNA<sup>Lys</sup> could be repositioned in two alternative ways, by a +1 shift (*above*) or a +4 shift (*below*)

tRNAs (Klobutcher and Farabaugh, 2002). Otherwise it is unclear why frameshift sequences (such as U.UU\_U.AA\_A) with other slippage-prone codons 5' of the stop codon have not been found. An alternative explanation for why frameshifting does not occur at other X.XX\_U.AA\_A (where X≠A) would be that ribosomes shift +4 (or bypass 1 nucleotide), see Fig. 14.5. Slow decoding of the UAA could facilitate repositioning of the P-site tRNA<sup>Lys</sup> 4 nucleotides downstream to re-pair to mRNA at the AAA codon in the +1 frame. For the A.AA\_U.AG\_A sequence, repositioning of the tRNA<sup>Lys</sup>, which likely has the anticodon xm<sup>5</sup>s<sup>2</sup>UUU (Björk et al., 2007), would re-pair to mRNA at the AGA codon, which has only slightly lower thermodynamic stability than cognate pairing (see Fig. 14.5). Though such a shift mechanism would make frameshifting at A.AG\_U.AA\_A equally plausible, no such sequences have been identified as potential frameshift sites. Direct sequencing or mass spectrometry is essential to decipher the exact mechanism, since +1 frameshifting would yield two lysines corresponding to this site, while +4 shift would result in incorporation of a single lysine. Mass spectroscopic analysis has been carried out only in the analysis of *p45*-encoded telomerase component of *Euplotes*; however, no peptides matching the ORF junction has been detected (Aigner et al., 2000).

## 14.5 Possible Functions of Products Generated by Low-Level Aberrant Translation

As has been shown by several studies described above, shift-prone sequences, although somewhat underrepresented throughout the genome and absent in highly expressed genes, are frequent in coding sequences. In a few distinct cases specific functional consequences of frameshifting can be envisioned. However, such cases are rare and in general the frameshifting on frameshift-prone sequences will result in premature termination and production of a nonfunctional peptide that

gets degraded. Most likely such frameshift events occur without any specific functional role and constitute minor faults of the translation process. Nevertheless, some general impact of such erroneous frameshifting on regulation of different cellular processes has been proposed. Some authors suggest that erroneous frameshifting can posttranscriptionally regulate mRNA stability, since encountering a premature termination codon by translating ribosome would trigger mRNA degradation through nonsense-mediated decay (NMD) pathway (Jacobs et al., 2007). However, the growing evidence suggests that in higher eukaryotes NMD can be triggered only during the first, so-called pioneer round of translation [review in Chang et al. (2007)]. If frameshifting occurs at a level of about 1%, then an mRNA containing such a frameshift site would be degraded through the NMD pathway only in 1% of the cases. On the other hand, in *S. cerevisiae* where NMD is inefficient and can be triggered after a number of translations of the PTC-containing mRNA, some downregulation of the mRNAs containing frameshift sites is feasible.

A consequence of erroneous frameshifting is production of an aberrant peptide. In some cases, when frameshifting occurs near the end of the coding region, the peptide synthesized might retain its function and could be utilized along with the products of standard translation (Mejlhede et al., 1999). In all other cases it is generally assumed that nonfunctional peptides get degraded. However, the exact fate is indeed unknown. Peptides produced by erroneous frameshifting can be potentially utilized as cryptic epitopes in the immune system. Two such cases have been described in the literature to date. One was identified in a patient with Reuter's syndrome. There, a transframe peptide produced via frameshifting from the IL-10 gene served as cryptic epitope to activate cytotoxic T cells (Saulquin et al., 2002). Intriguingly, the authors speculated that the frameshifting in the IL-10 could be of pathophysiological relevance since the preliminary data suggested recognition of the same epitope in another rheumatoid arthritis patient. Another example was identified in the herpes simplex virus (HSV) *tk* gene which encodes thymidine kinase (TK). Thymidine kinase is crucial for reactivation of the virus from a latent phase and is a target for antiviral therapy with the drug acyclovir. An acyclovir-resistant mutant has the insertion of a single *G* nucleotide in a run of 7 *G*'s in *tk* gene, resulting in a run of 8 *G*'s (Horsburgh et al., 1996). This frameshift mutation results in synthesis of nonfunctional TK and the mutant is resistant to acyclovir, which has to be phosphorylated by TK and subsequently by host kinases to an active form that interferes with viral replication (Elion, 1982). However, low levels of functional TK that are crucial for viral propagation are synthesized via ribosomal frameshifting on the run of 8 *G* (Griffiths et al., 2006; Besecker et al., 2007). In the wild-type *tk* gene the run of 7 *G* also causes about 1% frameshifting and the truncated peptide serves as a cryptic epitope and can trigger an immune response (Zook et al., 2006).

## 14.6 Conclusions

As we demonstrated in this chapter, sequences responsible for highly efficient alterations of standard genetic readout are sometimes underrepresented in protein coding regions of genomes. When such sequences play crucial roles for gene expression,



e.g., required for the biosynthesis of functional gene products, they exhibit deep phylogenetic conservation. Such sequences can be classified as singular genetic elements. Yet, there are a substantial number of sequences prone to low-level aberrant translational events and their underrepresentation in coding sequences is less pronounced. Even though the negative impact of such sequences in gene expression is less critical and their genomic locations are not strictly conserved, the subsequent non-canonical translational events have important functional implications, such as fine-tuning of expression levels during posttranscriptional regulation or production of epitopes for an immune response.

**Acknowledgments** P.V.B. thanks Science Foundation Ireland for Support.

## References

- Aigner S, Lingner J, Goodrich KJ, Grosshans TA, Shevchenko A, Mann M, Cech TR (2000). *Euplotes* telomerase contains a La motif protein produced by apparent translational frameshifting. *EMBO J* 19:6230–6239
- Asakura T, Sasaki T, Nagano F, Satoh A, Obaishi H, Nishioka H, Imamura H, Hotta K, Tanaka K, Nakanishi H, Takai, Y (1998) Isolation and characterization of a novel actin filament-binding protein from *Saccharomyces cerevisiae*. *Oncogene* 16:121–130
- Atkins JF, Baranov PV, Fayet O, Herr AJ, Howard MT, Ivanov IP, Matsufuji S, Miller WA, Moore B, Prere MF, Wills NM, Zhou J, Gesteland RF (2001) Overriding standard decoding: implications of recoding for ribosome function and enrichment of gene expression. *Cold Spr Harb Symp Quant Biol* 66:217–232
- Baranov PV, Fayet O, Hendrix RW, Atkins JF (2006) Recoding in bacteriophages and bacterial IS elements. *Trends Genet* 22:174–181
- Baranov PV, Gesteland RF, Atkins JF (2002) Release factor 2 frameshifting sites in different bacteria. *EMBO Rep* 3:373–377
- Baranov PV, Gesteland RF, Atkins JF (2004) P-site tRNA is a crucial initiator of ribosomal frameshifting. *RNA* 10:221–230
- Baranov PV, Henderson CM, Anderson CB, Gesteland RF, Atkins JF, Howard MT (2005) Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology* 332:498–510
- Bekaert M, Atkins JF, Baranov PV (2006) ARFA: A program for annotating bacterial release factor genes, including prediction of programmed ribosomal frameshifting. *Bioinformatics* 22:2463–2465
- Belcourt MF, Farabaugh PJ (1990) Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. *Cell* 62:339–352
- Bernardi G, Bernardi G (1986) Compositional constraints and genome evolution. *J Mol Evol* 24:1–11
- Besecker MI, Furness CL, Coen DM, Griffiths A (2007) Expression of extremely low levels of thymidine kinase from an acyclovir-resistant herpes simplex virus mutant supports reactivation from latently infected mouse trigeminal ganglia. *J Virol* 81:8356–60
- Björk GR, Huang B, Persson OP, Bystrom AS (2007) A conserved modified wobble nucleoside (mcm5s2U) in lysyl-tRNA is required for viability in yeast. *RNA* 13:1245–1255
- Blinkowa AL, Walker JR (1990) Programmed ribosomal frameshifting generates the *Escherichia coli* DNA polymerase III gamma subunit from within the tau subunit reading frame. *Nucl Acids Res* 18:1725–1729
- Brierley I, Digard, P, Inglis SC (1989) Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* 57:537–547
- Brierley I, Rolley NJ, Jenner AJ, Inglis SC (1991) Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *J Mol Biol* 220:889–902

- Capecchi MR, Klein HA (1970) Release factors mediating termination of complete proteins. *Nature* 226:1029–1033
- Chang YF, Imam JS, Wilkinson MF (2007) The nonsense-mediated decay RNA surveillance pathway. *Ann Rev Biochem* 76:51–74
- Chen SL, Lee W, Hottes AK, Shapiro L, McAdams HH (2004) Codon usage between genomes is constrained by genome-wide mutational processes. *Proc Natl Acad Sci USA* 101:3480–3485
- Coleman JR, Papamichail D, Skiena S, Fitcher B, Wimmer E, Mueller S (2008) Virus attenuation by genome-scale changes in codon pair bias. *Science* 320:1784–1787
- Craig WJ, Caskey CT (1986) Expression of peptide chain release factor 2 requires high-efficiency frameshift. *Nature* 322:273–275
- Craig WJ, Cook RG, Tate WP, Caskey CT (1985) Bacterial peptide chain release factors: conserved primary structure and possible frameshift regulation of release factor 2. *Proc Natl Acad Sci USA* 82:3616–3620
- Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190
- Curran JF (1993) Analysis of effects of tRNA: message stability on frameshift frequency at the *Escherichia coli* RF2 programmed frameshift site. *Nucleic Acids Res* 21:1837–1843
- Curran JF, Yarus M (1988) Use of tRNA suppressors to probe regulation of *Escherichia coli* release factor 2. *J Mol Biol* 203:75–83
- Elion GB (1982) Mechanism of action and selectivity of acyclovir. *Am J Med* 73:7–13
- Farh KK, Grimson A, Jan C, Lewis BP, Johnston WK, Lim LP, Burge CB, Bartel DP (2005) The widespread impact of mammalian Micro RNAs on mRNA repression and evolution. *Science* 310:1817–1821
- Fedorov A, Saxonov S, Gilbert W (2002) Regularities of context-dependent codon bias in eukaryotic genes. *Nucleic Acids Res* 30:1192–1197
- Flower AM, McHenry CS (1990) The gamma subunit of DNA polymerase III holoenzyme of *Escherichia coli* is produced by ribosomal frameshifting. *Proc Natl Acad Sci USA* 87:3713–3717
- Griffiths A, Link MA, Furness CL, Coen DM (2006) Low-level expression and reversion both contribute to reactivation of herpes simplex virus drug-resistant mutants with mutations on homopolymeric sequences in thymidine kinase. *J Virol* 80:6568–6574
- Gurvich OL, Baranov PV, Gesteland RF, Atkins JF (2005) Expression levels influence ribosomal frameshifting at the tandem rare arginine codons AGG\_AGG and AGA\_AGA in *Escherichia coli*. *J Bacteriol* 187:4023–4032
- Gurvich OL, Baranov PV, Zhou J, Hammer AW, Gesteland RF, Atkins JF (2003) Sequences that direct significant levels of frameshifting are frequent in coding regions of *Escherichia coli*. *EMBO J* 22:5941–5950
- Gustafsson C, Govindarajan S, Minshull J (2004) Codon bias and heterologous protein expression. *Trends Biotechnol* 22:346–353
- Herold J, Siddell SG (1993) An ‘elaborated’ pseudoknot is required for high frequency frameshifting during translation of HCV 229E polymerase mRNA. *Nucl Acids Res* 21:5838–5842
- Horsburgh BC, Kollmus H, Hauser H, Coen DM (1996) Translational recoding induced by G-rich mRNA sequences that form unusual structures. *Cell* 86:949–959
- Hu ST, Lee LC, Lei GS (1996) Detection of an IS2-encoded 46-kilodalton protein capable of binding terminal repeats of IS2. *J Bacteriol* 178:5652–5659
- Huang Y, Lau SK, Woo PC, Yuen KY (2008) CoVDB: a comprehensive database for comparative analysis of coronavirus genes and genomes. *Nucl Acids Res* 36:D504–511
- Ikemura T (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J Mol Biol* 146:1–21
- Ivanov IP, Atkins JF (2007) 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* 35:1842–1858

- Jacobs JL, Belew AT, Rakauskaite R, Dinman JD (2007) Identification of functional, endogenous programmed  $-1$  ribosomal frameshift signals in the genome of *Saccharomyces cerevisiae*. *Nucl Acids Res* 35:165–174
- Karlin S, Mrazek J, Campbell A, Kaiser D (2001) Characterizations of highly expressed genes of four fast-growing bacteria. *J Bacteriol* 183:5025–5040
- Kisselev LL, Buckingham RH (2000) Translational termination comes of age. *Trends Biochem Sci* 25:561–566
- Klobutcher LA (2005) Sequencing of random *Euplotes crassus* macronuclear genes supports a high frequency of  $+1$  translational frameshifting. *Eukaryotic Cell* 4:2098–2105
- Klobutcher LA, Farabaugh PJ (2002) Shifty ciliates: frequent programmed translational frameshifting in euplotids. *Cell* 111:763–766
- Kurland C (1979) Reading frame errors on ribosomes. In: Celis J, Smith JD (eds) *Nonsense mutations and tRNA suppressors*, Academic Press, London, pp 97–108
- Kurland C, Gallant J (1996) Errors of heterologous protein expression. *Curr Opin Biotechnol* 7:489–493
- Kurland CG, Hughes D, Ehrenberg M (1996) Limitations of translational accuracy. In *Escherichia coli and Salmonella typhimurium: Cellular and molecular biology*, ASM Press, Washington, DC, pp 979–1004
- Lainé S, Thouard A, Komar AA, Rossignol JM (2008) Ribosome can resume the translation in both  $+1$  or  $-1$  frames after encountering an AGA cluster in *Escherichia coli*. *Gene* 412:95–101
- Liao PY, Gupta P, Petrov AN, Dinman JD, Lee KH (2008) A new kinetic model reveals the synergistic effect of E-, P- and A-sites on  $+1$  ribosomal frameshifting. *Nucl Acids Res* 36:2619–2629
- Lill R, Wintermeyer W (1987) Destabilization of codon-anticodon interaction in the ribosomal exit site. *J Mol Biol* 196:137–148
- Ma J, Campbell A, Karlin S (2002) Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. *J Bacteriol* 184:5733–5745
- Major LL, Poole ES, Dalphin ME, Mannering SA, Tate WP (1996) Is the in-frame termination signal of the *Escherichia coli* release factor-2 frameshift site weakened by a particularly poor context? *Nucl Acids Res* 24:2673–2678
- Marquez V, Wilson DN, Tate WP, Triana-Alonso, F, Nierhaus KH (2004) Maintaining the ribosomal reading frame: the influence of the E site during translational regulation of release factor 2. *Cell* 118:45–55
- Mejlhede N, Atkins JF, Neuhard, J (1999) Ribosomal  $-1$  frameshifting during decoding of *Bacillus subtilis* *cdd* occurs at the sequence CGA AAG. *J Bacteriol* 181:2930–2937
- Moon S, Byun Y, Han K (2007) FSDB: a frameshift signal database. *Computat Biol Chem* 31:298–302
- Morris DK, Lundblad, V (1997) Programmed translational frameshifting in a gene required for yeast telomere replication. *Curr Biol* 7:969–976
- Mottagui-Tabar, S, Isaksson LA (1998) The influence of the 5' codon context on translation termination in *Bacillus subtilis* and *Escherichia coli* is similar but different from *Salmonella typhimurium*. *Gene* 212:189–196
- Moura G, Pinheiro M, Arrais J, Gomes AC, Carreto L, Freitas A, Oliveira JL, Santos MA (2007) Large scale comparative codon-pair context analysis unveils general rules that fine-tune evolution of mRNA primary structure. *PLoS ONE* 2:e847
- O'Connor M (2002) Imbalance of tRNA(Pro) isoacceptors induces  $+1$  frameshifting at near-cognate codons. *Nucl Acids Res* 30:759–765
- Parker J (1989) Errors and alternatives in reading the universal genetic code. *Microbiol Rev* 53:273–298
- Pavlov MY, Freistoffer DV, Dinbas V, MacDougall J, Buckingham RH, Ehrenberg, M (1998) A direct estimation of the context effect on the efficiency of termination. *J Mol Biol* 284:579–590

- Plant EP, Perez-Alvarado GC, Jacobs JL, Mukhopadhyay B, Hennig, M, Dinman JD (2005) A three-stemmed mRNA pseudoknot in the SARS coronavirus frameshift signal. *PLoS Biol* 3:e172
- Poole ES, Major LL, Mannering SA, Tate WP (1998) Translational termination in *Escherichia coli*: three bases following the stop codon crosslink to release factor 2 and affect the decoding efficiency of UGA-containing signals. *Nucl Acids Res* 26:954–960
- Sanders CL, Curran JF (2007) Genetic analysis of the E site during RF2 programmed frameshifting. *RNA* 13:1483–1491
- Saulquin X, Scotet E, Trautmann L, Peyrat MA, Halary F, Bonneville, M, Houssaint, E (2002) +1 Frameshifting as a novel mechanism to generate a cryptic cytotoxic T lymphocyte epitope derived from human interleukin 10. *J Exper Med* 195:353–358
- Scolnick E, Tompkins R, Caskey, T, Nirenberg, M (1968) Release factors differing in specificity for terminator codons. *Proc Natl Acad Sci USA* 61:768–774
- Shah AA, Giddings MC, Parvaz JB, Gesteland RF, Atkins JF, Ivanov IP (2002) Computational identification of putative programmed translational frameshift sites. *Bioinformatics* 18: 1046–1053
- Shine J, Dalgarno L (1975) Determinant of cistron specificity in bacterial ribosomes. *Nature* 254:34–38
- Sipley J, Goldman E (1993) Increased ribosomal accuracy increases a programmed translational frameshift in *Escherichia coli*. *Proc Natl Acad Sci USA* 90:2315–2319
- Spanjaard RA, Chen K, Walker JR, van Duin J (1990) Frameshift suppression at tandem AGA and AGG codons by cloned tRNA genes: assigning a codon to argU tRNA and T4 tRNA(Arg). *Nucl Acids Res* 18:5031–5036
- Spanjaard RA, van Duin J (1988) Translation of the sequence AGG-AGG yields 50% ribosomal frameshift. *Proc. Natl Acad Sci USA* 85:7967–7971
- Stoletzki N, Eyre-Walker A (2007) Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. *Mol Biol Evol* 24:374–381
- Su MC, Chang CT, Chu CH, Tsai CH, Chang KY (2005) An atypical RNA pseudoknot stimulator and an upstream attenuation signal for –1 ribosomal frameshifting of SARS coronavirus. *Nucl Acids Res* 33:4265–4275
- Taliaferro D, Farabaugh PJ (2007) An mRNA sequence derived from the yeast EST3 gene stimulates programmed +1 translational frameshifting. *RNA* 13:606–613
- Theis C, Reeder J, Giegerich R (2008) KnotInFrame: prediction of –1 ribosomal frameshift events. *Nucl Acids Res* 36:6013–6020
- Tsuchihashi Z, Kornberg A (1990) Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proc Natl Acad Sci USA* 87:2516–2520
- Vallabhaneni H, Fan-Minogue H, Bedwell DM, Farabaugh PJ (2009) Connection between stop codon reassignment and frequent use of shifty stop frameshifting. *RNA* 15:889–897
- Wan XF, Xu D, Kleinhofs, A, Zhou, J (2004) Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. *BMC Evol Biol* 4:19
- Weiss RB, Dunn DM, Atkins JF, Gesteland RF (1987) Slippery runs, shifty stops, backward steps, and forward hops: –2, –1, +1, +2, +5, and +6 ribosomal frameshifting. *Cold Spr Harb Symp Quant Biol* 52:687–693
- Weiss RB, Dunn DM, Dahlberg AE, Atkins JF, Gesteland RF (1988) Reading frame switch caused by base-pair formation between the 3' end of 16S rRNA and the mRNA during elongation of protein synthesis in *Escherichia coli*. *EMBO J* 7:1503–1507
- Weixlbaumer A, Jin H, Neubauer C, Voorhees RM, Petry S, Kelley AC, Ramakrishnan V (2008) Insights into translational termination from the structure of RF2 bound to the ribosome. *Science* 322:953–956
- Zimmer M, Sattelberger E, Inman RB, Calendar, R, Loessner MJ (2003) Genome and proteome of *Listeria monocytogenes* phage PSA: an unusual case for programmed +1 translational frameshifting in structural protein synthesis. *Mol Microbiol* 50:303–317
- Zook MB, Howard MT, Sinnathamby G, Atkins JF, Eisenlohr LC (2006) Epitopes derived by incidental translational frameshifting give rise to a protective CTL response. *J Immunol* 176:6928–6934