PAPER ID: PMC3500957

TITLE:
On programmed ribosomal frameshifting: the alternative proteomes

ABSTRACT:
Frameshifting results from two main mechanisms: genomic insertions or deletions (indels) or programmed ribosomal frameshifting. Whereas indels can disrupt normal protein function, programmed ribosomal frameshifting can result in dual-coding genes, each of which can produce multiple functional products. Here, I summarize technical advances that have made it possible to identify programmed ribosomal frameshifting events in a systematic way. The results of these studies suggest that such frameshifting occurs in all genomes, and I will discuss methods that could help characterize the resulting alternative proteomes.

Introduction:
Frameshifting is a process whereby the ribosome is guided toward a triplet nucleotide that is either shifted one nucleotide position upstream (+1 frameshift) or one nucleotide position downstream (−1 frameshift). Such frameshifting occurs in all known organisms, from E. coli to mammals (Namy et al., 2004; Dinman, 2012a).
There are two main mechanisms that produce out of frame peptides: changes in the genome sequence that result in insertions or deletions (indels) and programmed ribosomal frameshifting as a consequence of the ribosome either slipping back one nucleotide (−1 frameshifting) or skipping one nucleotide (+1 frameshifting) (Figure 1). Indels generally produce non-functional proteins and are associated with either spontaneous mutations across the genome or somatic genomic instability, for instance, as a consequence of tumour progression. By contrast, programmed ribosomal frameshifting can result in dual-coding genes that produce alternative functional proteins, which form an integral part of the organism's physiology.
Programmed ribosomal frameshifting has been historically associated with viruses and retrotransposons. Retroviruses require frameshifting for replication and infection (Maia et al., 1996; Brierley and Dos Ramos, 2006; Dulude et al., 2006). For example, the HIV1 polyprotein gag-pol requires efficient −1 frameshifting for expression of the individual gag and pol gene products. This form of frameshifting usually depends on a combination of a "slippery sequence," a spacer sequence of 1–15 nucleotides, and a stem-loop secondary RNA structure such as a pseudoknot (Figure 1) (Namy et al., 2006). The slippery sequence is generally of the type X XXY YYZ, where X denotes any nucleotide, Y denotes A or U, and Z is A, U, or C. Pseudoknots are secondary RNA substructures that contain two or more stem-loop motifs with intercalated stems. The pseudoknot or stem-loop structure in the mRNA is thought to result in pausing of the ribosome, resulting in eventual frameshifting (Namy et al., 2006). Structural evidence for this mechanism comes from the crystal structure of the mouse mammary tumor virus (MMTV) pseudoknot, which has an unpaired adenine that acts as a hinge to mediate frameshifting (Chen et al., 1996). Generally, there is a correlation between the mechanical strength of an mRNA pseudoknot and its frameshifting efficiency (Hansen et al., 2007): the stronger the pseudoknot the higher the frameshifting efficiency, although very strong pseudoknots can cause a road block that limits translation downstream (Tholstrup et al., 2011).
Other examples of mammalian genes that utilize −1 frameshifting are the mouse embryonic carcinoma differentiation regulated (EDR) gene and its human ortholog PEG10. A slippery sequence of G GGA AAC, in combination with a pseudoknot, mediates highly efficient −1 frameshifting, similar to viral frameshifting motifs (Clark et al., 2007). Recently, a programmed ribosomal −1 frameshift has been identified in the adenomatous polyposis coli (APC) mRNA in Caenorhabditis elegans that is mediated by a slippery sequence A AAA AAA or A AAA AAC (Baranov et al., 2011). The functional relevance of this frameshift is uncertain.
Although the slippery sequence and pseudoknot are the most common motifs for frameshifting identified thus far, there are alternative mechanisms that may result in the production of out-of-frame proteins. Alternative splicing may contribute to frameshifting (Hiller et al., 2005), as can codon bias. For instance, rare tRNA codons can favor −1 and +1 frameshifting (Gurvich et al., 2005; Laine et al., 2008), and rare arginine codons prime mitochondrial sequences for frameshifting (Temperley et al., 2010). Moreover, CAG repeats are prone to frameshifting, which results in poly-alanine proteins that may contribute to the pathogenesis of neurodegenerative

diseases (Toulouse et al., 2005). The use of the peptidyltransferase inhibitor anisomycin reduces −1 frameshifting in these cases and reduces the toxicity associated with the expanded triplet repeats. Importantly, out-of-frame proteins (compared to the standard ORFeome annotation) can also result from alternative AUG or CUG start sites (Ingolia et al., 2011), thereby considerably increasing the size of an alternative "frameshifted" proteome. Main sources of out-of-frame peptides and proteins are shown in Figure 2.

An interesting example of programmed ribosomal +1 frameshifting is that of the ornithine decarboxylase ODC) gene that produces the antizyme from frameshifting of the mRNA sequence (Bekaert et al., 2008). ODC catalyzes the production of polyamines, such as putrescine, spermidine, and spermine from ornithine through decarboxylation. ODC activity is terminated by the antizyme (Murakami et al., 1992)—providing an elegant mechanism for shutting down the activity of an enzyme by producing an out-of-frame antizyme from the same mRNA. This frameshifting is tightly regulated and can be enhanced by treatment of cells with polyamines (Nilsson et al., 1997). The ODC antizyme mechanism is highly conserved throughout all eukaryotes (Ivanov et al., 2000b). The frameshifting requires the UGA stop codon and a 3′ stem loop that forms a RNA pseudoknot. These RNA hallmarks are still the standard way to identify other +1 frameshifted proteins.

It has been proposed that frameshifting is a common mechanism to increase protein-coding capacity of small genomes such as those of viruses and mitochondria. In agreement with this proposal, frameshifting is common in mitochondrial genes, and genome size seems to correlate with the abundance of frameshifting (Seligmann, 2010). In fact, some organisms display a high complexity in frameshifting: In the dinoflagellate Perkinsus marinus, the mitochondrial gene that encodes cytochrome c oxidase subunit 1 can shift up to 10 times within the same mRNA sequence in order to produce the correct gene product (Masuda et al., 2010). Clearly, such complex frameshifting requires efficient regulatory control.

Regulation of frameshifting:

There is increasing evidence that ribosomal frameshifting is a regulated event. Several enzymatic and non-enzymatic mechanisms have been proposed that result in an enhancement of frameshifting. For instance, the production of antizyme by +1 frameshifting is enhanced by the end-products of ODC—spermine, putrescine, and spermidine (Ivanov et al., 2000a). How polyamines regulate +1 frameshifting is not well understood, but one hypothesis is that polyamine binding to RNA may enable read-through of the termination codon. Similar to polyamines, amino-glycosides such as gentamicin allow read-through of stop-codons (Martin et al., 1989; Malik et al., 2010).

It is not clear at present whether there are normal regulator proteins that enhance frameshifting efficiencies. Because −1 frameshifting is essential for retroviral gene expression, it has been proposed that chemical interference with frameshifting would be a good anti-viral strategy. Recently, a genome-wide screen to identify regulators of HIV-1 frameshifting has identified eRF1 as an essential host gene required for −1 frameshifting (Kobayashi et al., 2010), suggesting that eRF1 may be a good therapeutic target in AIDS (Brakier-Gingras et al., 2012). In addition, it has been proposed that HIV frameshifting can be modulated by protein kinase R, as well as by factors that modulate translation efficiency such as rapamycin (Gendron et al., 2008).

Interestingly, annexin A2 (ANXA2) can bind the pseudoknot structure of avian coronavirus infectious bronchitis virus (IBV) and reduce −1 frameshifting (Kwak et al., 2011). As a consequence, ANXA2 has been suggested as a more general antiviral regulator in eukaryotic cells. Other potential anti-viral agents could either specifically bind to the frameshift signal (such as antisense oligonucleotides, non-coding RNAs, or frameshift signal binding compounds) or interfere with peptidyltransferase activity (e.g., anisomycin or sparsomycin) or eEF2 activity (e.g., sordarin) (Dinman, 2012b).

For certain diseases, however, it may be benefitial to enhance frameshifting. For instance, in monogenetic diseases such as cystic fibrosis or Duchenne's muscular dystrophy, where frameshift mutations result in premature translation termination, the deliberate induction of frameshifting may overcome the problem by skipping the affected sites (Aurino and Nigro, 2006). It has been noted that aminoglycosides such as gentamicin can enhance frameshifting and stop codon read-through. Gentamicin-induced read-through of stop codons has been evaluated as a treatment option for Duchenne's muscular dystrophy (Malik et al., 2010). While the authors conclude from this phase I clinical trial that gentamicin may not be a good treatment option, they note that other read-through agents may have benefits. A phase I/II clinical trial using a morpholino oligomer (AVI-4658) to correct a frameshift mutation in the dystrophin gene has been

completed with the conclusion that AVI-4658 was well tolerated and had significant benefit in patients with Duchenne's muscular dystrophy (Cirak et al., 2011).

Dual coding:
Genome-wide analysis of the yeast genome (Jacobs et al., 2007) and other genomes (Hammell et al., 1999) suggests that frameshifting is more common than previously thought. However, it has been proposed that frameshifting may predominantly serve to modulate RNA levels rather than to produce frameshifted proteins (Plant et al., 2004). Evolutionary studies argue that the generation of out-of-frame proteins has been minimized by codon optimization that results in non-functional small peptides rather than functional proteins (Bollenbach et al., 2007). However, it is becoming increasingly recognized that programmed ribosomal frameshifting can result in peptides or proteins with physiological functions (Dinman, 2012a).
One question is whether frameshifting compromises the function of the original frame. It has been proposed that dual coding limits the evolutionary flexibility of the underlying nucleotide sequence (Firth and Brown, 2006; Rancurel et al., 2009). Thus, one possibility may be that frameshifting occurs predominantly in highly conserved, essential genes. However, it has been argued that once a frameshift event is released from selective pressure, as occurs in gene duplication, it can evolve to produce a beneficial functional protein (Raes and Van de Peer, 2005). Indeed, in some cases, novel genes seem to have emerged by frameshifting of a pre-existing coding sequence (Ohno, 1984; Ranz et al., 2003). Moreover, regulated frameshifting can allow the same gene to produce alternative beneficial proteins. In fact, evolutionary studies indicate a high abundance of frameshift events in human and mouse genomes that may be linked to an increased usage of the opal TGA stop codon (Okamura et al., 2006).
It has been proposed that at least 1% of the human genome consists of dual coding regions (Michel et al., 2012) and that the number of out-of-frame peptides or proteins may be even higher than that. Other studies have suggested that ~10% of the genome contains −1 frameshift signals (Belew et al., 2011). I would argue that the number of frameshifted peptides or proteins is somewhere in the range of 1–10% of the genome. This is a very significant fraction of the genome, thus suggesting that out-of-frame peptides are an inherent part of animal physiology and part of evolutionary selection processes. Accordingly, it can be anticipated that dual coding is regulated and a common mechanism for producing additional gene products. The main questions are:
What is the sequence or structural motif for dual coding?What is the identity of all dual-coding genes and out-of-frame proteins in the genome?How is dual coding regulated?

To answer these questions, we need to look at commonly used methods to identify frameshift events and out-of-frame peptides and proteins.

Computational prediction of frameshifting events ::: Methods to identify out-of frame peptides:
Several databases have been created to help predict frameshift sequences in the genome. Such predictions are based on primary mRNA sequence stretches or on secondary hairpins or pseudo-knots within the mRNA sequence. For instance, Hammell et al. searched the genome for slippery sequences and pseudoknot structures (Hammell et al., 1999). This approach identified over 200 putative programmed ribosomal frameshifting events. For identification of pseudoknots and slippery sequences in the genome databases such as RECODE, KnotInFrame, PRFdb, and FSdB are very useful (Table 1).
RECODE provides information about programmed frameshifting, read-through and bypassing, based on published results in the literature (Baranov et al., 2001; Bekaert et al., 2010). The database utilizes information from ~1500 known frameshifted gene products. A majority of the data on frameshifting comes from two frameshifted proteins, RF2 and antizyme, and is constantly updated using the respective prediction tools [ARFA (Bekaert et al., 2006) and OAF (Bekaert et al., 2008), respectively]. PseudoViewer is used for visualization of pseudoknot structures. (http://recode.ucc.ie)
KnotInFrame is a program for predicting sites of −1 frameshifting based on the formation of RNA pseudoknots (Theis et al., 2008). The authors have developed a specialized RNA-folding program called pknotsRG-fs that compares the minimal free energy of an enforced pseudoknot structure to that of a freely folded structure such as that given by RNAfold. The spacer region after the common slippery sequence X XXY YYZ is between 1 and 12 nt long. (http://bibiserv.techfak.uni-bielefeld.de/knotinframe)

PRFdb (http://prfdb/umd.edu/) is limited to −1 frameshifting in eukaryotes only. Again, the prediction if based on the presence of a heptameric slippery sequence in combination with a pseudoknot (Jacobs et al., 2007). The slippery sequence is modeled with a 1–8 nt spacer, and the pseudoknot is identified using RNAMotif. The pseudoknot is then further confirmed with other secondary RNA-structure-prediction tools, including Pknots (Rivas and Eddy, 1999), Nupack (Dirks and Pierce, 2004), and HotKnots (Ren et al., 2005).

The Frameshift Signal Database (FSDB) is a compilation of all known frameshift motifs, plus some reported (predicted) frameshift sequences (Moon et al., 2007). Based on commonalities between these sequences, the associated FSFinder allows mining of the genome for potential frameshift sequences. (http://wilab.inha.ac.kr/fsdb) At present, the database contains a total of 63 experimental and 190 predicted sequences from viruses, prokaryotes and eukaryotes. FSFinder uses a combination of slippery sequences and pseudoknot or stem-loop prediction. The heptameric slippery sequences for +1 and −1 frameshifting are listed in Table 2. It is noteworthy that FSDB includes deviations from the standard X XXY YYZ slippery sequence for prediction of −1 frameshifting, whereas +1 frameshift sequences often contain stop codons.

Although most prediction databases use similar principles (slippery sequence plus pseudoknot), they differ in prediction of frameshifting events. This may be due to variable thresholds applied or different RNA folding algorithms. In order to identify expressed proteins, one has to consider the length of the predicted out-of-frame protein. Most frameshifted proteins initiated by predicted slippery sites will terminate within 5–10 codons, thus producing truncated or non-functional peptides or proteins. Alternative approaches that take out-of-frame protein length into account might be beneficial. For instance, MLOGD (http://Guinevere.otago.ac.nz/aef/MLOGD) is a program for detection of overlapping coding sequences based on sequence alignments and analysis of mutation patterns (Firth and Brown, 2006). The limitations of this approach are that less conserved coding sequences will not be identified, and -2 frame overlaps can be identified as false positives.

In some organisms, including mammals, codon usage has evolved to minimize frameshifting. Thus, it can be expected that certain codons may favour frameshifting. FSCAN is a program to identify +1 frameshift sequences based on codon usage in E. coli (Liao et al., 2009). FScan searches 16 nt sequences and calculates a score for aa-tRNA competition between the zero and +1 frame. Accordingly, a stop codon, or a rare codon in the zero frame can be a predictor of +1 frameshifting. Shah et al. suggested that selective pressure would lead to an under-representation of frameshift sites in protein-coding sequences relative to an organism's codon bias (Shah et al., 2002). They predicted the sequences CUU AGG C and CUU AGU U, which mediate +1 frameshifting of ABP140 and EST3, respectively, to be highly under-represented and predictive for frameshifting in S. cerevisiae.

Once frameshifting events have been identified, it is important to characterize the gene products. One could apply the filters mentioned above, such as the 50-nt rule, although it is quite possible that short peptides are stable and endogenously expressed (Kondo et al., 2010; Ingolia et al., 2011). Ultimately, all predictions of frameshifted peptides need to be validated with experimental methods.

Low-throughput experimental methods ::: Methods to identify out-of frame peptides:
Experimental methods include generation of frameshift-specific antibodies, reporter constructs, biophysical methods, and single molecule measurements. The most commonly used frameshift reporter is a tandem luciferase construct where the two luciferases with different substrate specificities are separated by a stretch of nucleotides of a length that shifts the downstream luciferase either one nucleotide up or down (Grentzmann et al., 1998). The downstream luciferase will be expressed only if the nucleotide stretch can mediate frameshifting, while the upstream luciferase serves as an expression control. Alternative reporter genes include the use of fluorescent proteins (Cardno et al., 2009), although the dynamic range of luciferases is generally much higher. This reporter can be used experimentally to confirm predicted frameshift events. One hypothesis is that the −1 frameshifting efficiency correlates with the mechanical force required for pulling the RNA pseudoknot apart. With biophysical single-molecule methods that measure these forces using optical tweezers, Hansen et al. have confirmed that unfolding of a IBV-based pseudoknot required ~500 kJ/mol, compared to the theoretically determined 292 kJ/mol (Hansen et al., 2007). Chen et al. have determined that a 100% confidence in −1 frameshifting is reached by an unfolding force of ~57 pN (Chen et al., 2009). Such single-molecule biophysical approaches may help to identify potential frameshift sequences, but as they require

immobilization of the RNA, they will be technically challenging to implement on a genome-wide scale.

An alternative is the use of single molecule Foerster Resonance Energy Transfer (smFRET) (Aitken and Puglisi, 2010). FRET is a useful technique to measure proximity of biomolecules. A donor fluorophore attached to a molecule can transfer photons to an acceptor fluorophore on a different molecule when both molecules are close together (generally less than 10 nm). Aitken and Puglisi have used this technique to label individual tRNA molecules with donor and acceptor dyes that result in energy transfer when in close proximity and correct orientation. This technique enables to monitor relative tRNA positions and movement of tRNAs on ribosomes at a millisecond scale. It has been used to identify ribosomal translocation events on fluorescent ribosomes on immobilized RNA sequences and predicted the slipperiness of various RNA sequences. Again, this will be technically difficult to implement on a genome-wide scale.

All of the above methods will record mostly frameshifting events, but they will not validate the expression of an out-of-frame protein. It is worth noting that frameshifting will most likely result in chimeric sequences composed of a stretch of zero frame peptides linked to out-of-frame peptides, which can be a small or large part of the overall protein, depending on where the frameshift site is. Alternatively, frameshifting could lead to truncation of the original protein, where a very small fraction—if any—of the overall protein is out-of-frame (Figure 4). Those proteins that have a sufficient predicted length could then be validated by raising specific antibodies against that sequence. In order to raise an antibody, one has to know the precise sequence of the frameshifted protein. Therefore, this approach is useful for confirming a known out-of-frame protein, but it is not amenable to large-scale genome-wide screening. In addition, the detection of an endogenously expressed out-of-frame protein does not necessarily indicate a functional relevance for this protein.

In order to validate a physiological role for the frameshifted gene product, the best method is to genetically ablate the expression of the frameshifted protein while preserving the in-frame sequence. This can be accomplished by gene targeting with a gene sequence that harbours multiple wobble base pair mutations or mutating the frameshift motif, thus altering the RNA sequence but not the in-frame protein sequence. This is technically challenging and may not always be feasible. For instance, an OAZ3 knockout mouse model has been generated, where both the zero frame and the +1 frame has been deleted (Tokuhiro et al., 2009). To my knowledge, no specific gene targeting of a frameshift protein has been done so far.

Genomic methods ::: Methods to identify out-of frame peptides:

Genomic methods such as RNA sequencing (RNA-seq) have emerged as powerful tools to profile RNA content in cells. RNA-seq is based on high-throughput sequencing of a cDNA library generated from cellular RNA. In its standard form, RNA-seq enables the identification of indels, but will fail to identify post-transcriptional phenotypes and will therefore fail to identify programmed ribosomal frameshifting or dual coded genes. Nevertheless, ribosome profiling methods such as Ribo-seq have been developed that allow the identification of active translation, based on sequencing of cDNA libraries generated from ribosome-protected fragments (Ingolia et al., 2009). In this case, mRNA bound to ribosomes is first cross-linked and then isolated using, e.g., sucrose gradient density centrifugation. Next, a nuclease digestion step results in removal of mRNA sequence that is not bound ("protected") to ribosomes. The protected RNA bound to ribosomes is then reverse transcribed into cDNA and sequenced. Therefore, the precise position of a ribosome can be matched to the site of active translation. An adaptation of this method using the drug harringtonine to cause ribosome accumulation at initiation codons has allowed the identification of translation start sites and confirmed that many proteins are initiated at non-AUG or alternative AUG sites (Ingolia et al., 2011). In this study, 44% of detected AUG start sites were unannotated, and a large fraction of these encoded out-of-frame peptides. Further uses of ribosome profiling have confirmed that the identification of frameshifts is possible using genomic technologies (Michel et al., 2012). Although this approach is unbiased in the sense that it does not pre-filter genetic regions, one problem is the non-uniformity of ribosome-protected fragment reads. For instance, the preparation of cDNA libraries generated from ribosome-protected fragments can result in over- or under-representation of sequence reads. This can be overcome by a computational approach that measures the cumulative subcodon proportion difference of ribosome-protected fragments relative to local subcodon positions. The authors have therefore combined experimental data from Ribo-Seq with a computational approach in order to identify novel frameshifted protein sequences. Using this approach, several new frameshifted protein sequences were identified, most of which were dual-encoded. The authors estimate that more

than 1% of the genome may consist of dual-coding regions, and this is likely an underestimation. Further improvements to the method, including deeper sequencing to get better coverage of ribosome-protected fragments, will help to identify these genes.

Phenotypic screening ::: Methods to identify out-of frame peptides:
Most genome-wide screening approaches such as RNAi-based knockdown methods will fail to identify dual-coding regions, as both gene products will be deleted. Further, most commonly used cDNA libraries such as the ORFeome are designed to avoid expression of frameshifted peptides —for instance, by deletion of the 5′ and 3′-UTR. It may, however, be possible that certain phenotypes in cDNA screening are exerted by out-of-frame proteins. In order to fish for phenotypic effects exerted by such peptides, cDNA libraries can be designed in a way that genomic fragments are inserted downstream of an AUG start codon in +1 or −1 frame so that all genomic fragments are deliberately frameshifted. A subsequent functional phenotypic screen can then identify phenotypes associated with expression of such deliberate frameshift fragments. However, there are limitations in such an approach, e.g., where to place the frameshift. Also, deliberately expressed out-of-frame proteins may have little physiological relevance.
An alternative may be to use agents such as gentamicin that enhance programmed frameshifting. In that case, cells expressing a cDNA library would be treated with a frameshift inducer, and the occurrence of differential phenotypes plus/minus frameshift inducer would be recorded. It is imperative that cDNA libraries with 3′-UTR regions are used in such an approach to facilitate the out-of-frame peptide expression after stop codon read-through.
A similar approach can be used to identify endogenous regulators of frameshifting. One could use the dual luciferase reporter construct (Grentzmann et al., 1998) with a known frameshift motif and screen siRNA or cDNA libraries to identify genes that enhance or inhibit the frameshifting efficiency of this reporter. The existence of non-coding RNAs that modulate frameshifting suggests that frameshifting is regulated by endogenous gene products.
The dual luciferase reporter mentioned above can also be used to probe the sequence space for optimal frameshifting motifs (Rakauskaite et al., 2011). Recently, an adaptation of this approach has been developed using fluorophores that are amenable to high-throughput screening applications (Cardno et al., 2009). An in vivo adaptation for yeast cells has been proposed for use in high-throughput screening experiments (Harger and Dinman, 2003). For instance, a random nucleotide sequence can be inserted between the in-frame and the out-of-frame luciferase to determine which sequence will result in high levels of out-of-frame expression. The main advantages of such a system are the broad linear range of the assay, the internal mRNA expression control (luciferase 1), the possibility to normalize relative frameshift expression, and the ease of use. One problem with the approach is that the secondary structure is affected by the length of the sequence inserted in the dual luciferase reporter. Thus, the context of the frameshift motifs needs to be taken into account. Further, the high number of potential nucleotide combinations may outweigh the capacity of even highly automated processes. Even though this will identify putative sequences with high potential for frameshifting, it is still not clear whether the out-of-frame peptides or proteins are stable and functional. Their presence in open reading frames may hint that a frameshift is buried within the gene. Subsequent low-throughput experimental methods (see above) need to be designed to confirm expression of the frameshifted protein.

Proteomics ::: Methods to identify out-of frame peptides:
Proteomic approaches have enabled the identification of expressed peptides under physiological conditions. In mass spectrometry (MS), peptides of endogenous proteins are first detected as a mass per ion. In the most common approach, recorded masses are matched to all potential masses in the respective database by search algorithms like SEQUEST, MASCOT, Andromeda, or PEAKS. In a next step, sequence information generated in tandem MS (MS/MS) experiments is employed to identify potential hits within the shortlisted peptide variants. Commonly searched databases by MASCOT are SwissProt, NCBInr, and EMBL EST. These databases integrate cDNA and EST sequences and generally do not contain out-of-frame peptides. However, as MS database search algorithms compute the entire sequence space of potential peptide matches to identify peptides within analyzed samples, it is in principle possible to identify out-of-frame peptides.
An interesting approach for the identification of frame-shift peptides is de novo sequence analysis by MS. This is based solely on the analysis of MS and MS/MS spectra, without amino acid

sequence information from databases. However, identification rates are typically lower than in classical database-based proteomics experiments.

A major difficulty is to assign frame-shifted peptides to a particular gene. Decoding of the underlying nucleotide sequence is sometimes problematic, as, for instance, isoleucine and leucine have the same exact mass. In some cases, the identified peptides do not match the database for various reasons, including errors in gene/protein annotation and post-translational modifications in the peptides that are not accounted for. Usually, unmatched peptides are not reported in publications and are disregarded from further analysis. It is possible that some of these unmatched peptides correspond to an alternative reading frame. In order to match these peptides, one would need to generate a database derived from the +1/−1 frames of the ORFeome, similar to what Okamura et al. have done (Okamura et al., 2006), and make these accessible to the proteomic community. One problem is the "breakpoint" of frameshifting, which would generate a peptide that is partially composed of the original frame and partially of the frameshifted peptide (see Figure 4). Another problem in this approach may be the paucity of frameshifted peptides, as they are commonly expressed at lower levels (see above) and may not be easily detectable by MS approaches. However, MS is probably the most powerful method to identify out-of-frame peptides to confirm endogenous expression.

An alternative is to study peptides presented by cell-surface, class I MHC proteins. They present peptides derived from intracellular proteins to enable immune tolerance and immune surveillance. MHC-presented out-of-frame peptides were discovered in the early 1990's (Shastri et al., 1995) and have enabled the unbiased identification of endogenous frameshifting in mammalian cells long before the technical advances of MS. However, it is technically challenging to use this approach as a systematic tool for the identification of genomic out-of-frame peptides.

Outlook:

In which cellular processes should we expect to see a high abundance of frameshifted proteins? In principle, such proteins might be involved in any cellular process, but may be correlated with certain cellular pathways. For instance, amino acid starvation can induce frameshifting in bacteria (Barak et al., 1996). One hypothesis is that this is due to a short supply of amino-acylated tRNA as a consequence of amino acid limitations. This is supported by the observation that antizyme expression (the +1 frame) is maintained under conditions of amino acid starvation, while expression of ODC (the 0 frame) is reduced in rat intestinal epithelial cells (Ray et al., 2012). Moreover, treatment with mTOR inhibitors such as rapamycin also reduce 0 frame expression while maintaining antizyme expression. One could hypothesize that reduced fidelity in translation may enhance frameshifting. On the other hand, both translation inhibition and amino acid starvation are conditions that increase autophagy in eukaryotic cells, raising the possibility that the production of out-of-frame proteins may be functionally coupled to the regulation of autophagy.

Conflict of interest statement ::: Conclusion:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.