Published online 7 December 2006 *Nucleic Acids Research, 2007, Vol. 35, No. 1* ***165–174***
doi:10.1093/nar/gkl1033

Identification of functional, endogenous programmed −1 ribosomal frameshift signals in the genome of *Saccharomyces cerevisiae*

Jonathan L. Jacobs, Ashton T. Belew, Rasa Rakauskaite and Jonathan D. Dinman*

Department of Cell Biology & Molecular Genetics, University of Maryland, 2135 Microbiology Building, College Park, MD 20742, USA

Received August 2, 2006; Revised September 29, 2006; Accepted November 6, 2006

ABSTRACT

In viruses, programmed −1 ribosomal frameshifting (−1 PRF) signals direct the translation of alternative proteins from a single mRNA. Given that many basic regulatory mechanisms were first discovered in viral systems, the current study endeavored to: (i) identify −1 PRF signals in genomic databases, (ii) apply the protocol to the yeast genome and (iii) test selected candidates at the bench. Computational analyses revealed the presence of 10 340 consensus −1 PRF signals in the yeast genome. Of the 6353 yeast ORFs, 1275 contain at least one strong and statistically significant −1 PRF signal. Eight out of nine selected sequences promoted efficient levels of PRF in *vivo*. These findings provide a robust platform for high throughput computational and laboratory studies and demonstrate that functional −1 PRF signals are widespread in the genome of *Saccharomyces cerevisiae*. The data generated by this study have been deposited into a publicly available database called the PRFdb. The presence of stable mRNA pseudoknot structures in these −1 PRF signals, and the observation that the predicted outcomes of nearly all of these genomic frameshift signals would direct ribosomes to premature termination codons, suggest two possible mRNA destabilization pathways through which −1 PRF signals could post-transcriptionally regulate mRNA abundance.

INTRODUCTION

Programmed ribosomal frameshifting (PRF) is a translational recoding phenomenon historically associated with viruses and retrotransposons. A PRF signal stochastically redirects translating ribosomes into a new reading frame (i.e. by +1 or −1 nt) and, in the typical viral context, these signals allow ribosomes to bypass the usual in-frame stop codon and continue synthesis of a C-terminally extended fusion protein. As with most basic molecular mechanisms, although first described in viruses, it is becoming increasingly apparent that PRF is much more widespread and is likely employed by organisms representing every branch in the tree of life [for reviews see (1–4)].

This report focuses on programmed −1 ribosomal frameshifting (−1 PRF) and its use by chromosomally encoded mRNAs of yeast. The most well defined −1 PRF phenomena are directed by an mRNA sequence motif composed of three important elements: a 'slippery site' composed of seven nucleotides where the translational shift in reading frame actually takes place; a short spacer sequence of usually <12 nt and a downstream stimulatory structure (usually a pseudoknot). A 'typical' −1 PRF signal is shown in Figure 1. In eukaryotic viruses, the slippery site has the heptameric motif N NNW WWH (3). Current models posit that aminoacyl- and peptidyl-tRNAs are positioned on this sequence while the ribosome pauses at the downstream secondary structure (5–8). The nature of the slippery sequence enables re-pairing of the non-wobble bases of both the aminoacyl- and peptidyl-tRNAs with the −1 frame codons. While it is generally accepted that mRNA pseudoknots are the most common type of downstream stimulatory structures, other mRNA structures are capable of filling this role as well (9,10). Nonetheless, it is thought that the essential function of the stimulatory structure is to provide an energetic barrier to a translating ribosome.

A growing number of examples now exist of PRF signals in expressed eukaryotic genes (11–16). The existence of these PRF signals in a wide variety of viral and prokaryotic genomes suggests an ancient and possibly universal mechanism for controlling the expression of actively translated mRNAs. There have been several published reports aimed

*To whom correspondence should be addressed. Tel: +1 301 405 0918; Fax: +1 301 314 9489; Email: dinman@umd.edu
Present address:
Jonathan L. Jacobs, Laboratory of Receptor Biology and Gene Expression, National Cancer Institute, Advanced Technology Center, Gaithersburg, MD 20877, USA

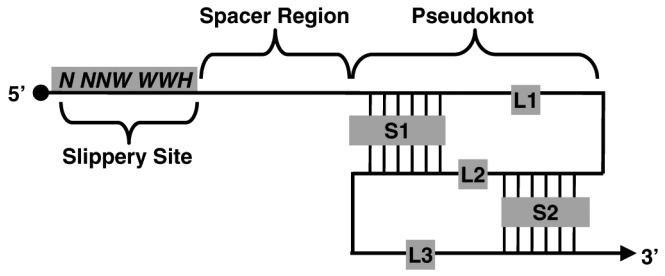

Figure 1. Typical −1 PRF signals consist of a heptameric slippery site that fits the motif N NNW WWH (spaces indicate zero frame codons), a short spacer region of 5–12 nt, and an mRNA pseudoknot with two stem and three loop regions (S1, S2 and L1–L3, respectively). See Materials and Methods for the pseudoknot motif criteria used in this study.

at the *in silico* identification of 'recoding signals' using a wide variety of computational approaches (4,17–21). While the methodologies of each study covered a broad range of bioinformatics techniques, the general goal of each of these [with the exception of (19)] was fundamentally the same. Specifically, the strategy has been to first find out-of-frame ORFs followed by the identification of PRF signals in the overlapping region between them that could act to potentially redirect ribosomes from the upstream ORF into the downstream one, thereby resulting in the translation of a fusion protein. The strength of this approach is that it can identify new classes of cis-acting signals capable of directing efficient PRF. However, this computational strategy is based on the assumption that PRF outcomes should mimic those observed in viral genomes, and thus its weakness is that it cannot identify new functional outcomes of frameshifting.

In contrast, while 'outcome-neutral' approaches using mRNA motifs known to promote efficient PRF cannot identify new frameshift signals, they can enable an expansion of our understanding of functional uses for PRF. In this vein, the first computational search for eukaryotic −1 PRF signals (19) did not focus solely on identifying two overlapping out-of-frame ORFs, but instead aimed to find these motifs throughout the entire CDS of the yeast genome. This early study found some 260 putative −1 PRF signals in the annotated portion of the *Saccharomyces cerevisiae* genome. The shortcomings of the Hammell *et al*. (19) study were its limitation by incomplete annotation of the yeast genome, and relatively insufficient computational resources available at the time (ca. 1995–1998). Thus, in order to achieve a more comprehensive approach, a new set of informatics tools were developed and applied using faster and more robust computational platforms. The results of the current study show that: (i) pattern matching approaches coupled with a predictive method for folding RNA sequences provide a dramatic improvement in the results; (ii) −1 PRF motifs are widespread in the genome of *S.cerevisiae*; (iii) many of the putative signals identified have predicted secondary structures with statistically significant measures of free energy; (iv) putative −1 PRF signals from a variety of *S.cerevisiae* genes promote efficient recoding when tested *in vivo*. Additionally, analysis of the predicted outcomes of −1 PRF events suggests that the vast majority would direct translating ribosomes to premature termination codons, suggesting that PRF could be used to post-transcriptionally regulate gene expression through the nonsense-mediated mRNA decay pathway (NMD).

MATERIALS AND METHODS

Hardware and software

All software was compiled and run on one or more of the following systems: Dell Precision 620, 2× PIII XEON 866 MHz running Mandrake Linux 10.×; Apple Power Macintosh, 2× G4 1.4 GHz PowerPC running OS X Tiger; SGI Cluster 64× MIPS R14K 600 MHz running Irix 6.5; SGI Altix 3000, 64× 1.5 GHz Itanium II running Linux-64. Supercomputing resources were made available courtesy of The National Cancer Institute's Advanced Biomedical Computing Center in Frederick, MD (ABCC, http://www.abcc.ncifcrf.gov). Unless otherwise noted, data mining and analysis was carried out using scripts written in PERL (www.perl.org), each of which are available on request. In all cases, data is stored in a mySQL 4.× relational database (www.mysql.com) referred to as the PRFdb (http://dinmanlab.umd.edu/prfdb).

Pattern matching

RNAMotif (22) was utilized for finding subsequences in the coding regions of *S.cerevisiae* that serve as potential translational frameshift signals. The descriptor of the putative programmed frameshift signal motif was created *ad hoc* from analysis of 56 known viral −1 PRF signals from the RECODE (23) database. The RNAMotif descriptor had the following requirements: (i) define slippery sites as using the IUPAC nomenclature 'N NNW WWH', where: (A) spaces indicate zero-frame codon boundaries, (B) N NN represents any three identical nucleotides, (C) WWW represents AAA or UUU, and (D) H ≠ G; (ii) allow any sequence between 0 and 12 nt in length to serve as the spacer between the slippery site and the pseudoknot; (iii) allow G:U base pairing in pseudoknot stems; (iv) each stem in the pseudoknot must be between 4 and 20 nt in length; (v) the first loop must be between 1 and 3 nt in length; (vi) the second loop is optional; (vii) the third loop must be at least as long as one-half the length of the first stem and no longer than 100 nt.

Genome randomization

The complete coding sequence (CDS) of *S.cerevisiae* was randomized using seven different methods for sequence randomization. Each method of randomization was conducted such that each genome had the same number of open reading frames (ORFs) and identical ORF lengths of the natural *S.cerevisiae* genome. In addition, each random genome was generated in such a way such that stop codons were only present in the terminal 3′ position (i.e. no in-frame termination codons). Beyond these similarities, the seven methods for randomization included: NoBias, randomized ORFs with unbiased nucleotide bias; ntShuffle, nucleotides from each natural ORF are shuffled by triplicate mononucleotide permutations; ntBias, randomized ORFs using the CDS single-nucleotide frequency; cdnShuffle, codons from each natural ORF are shuffled by triplicate monocodon permutation; SilentBias, a silent bias where the codons are randomized in place so as to maintain protein coding sequence; cdnBias,

randomized ORFs using the observed CDS codon usage bias; diNuc, randomized ORFs generated using the observed CDS dinucleotide frequency. A total of 100 randomized replicate genomes were generated for each of the seven methods (700 random genomes total, ~6.5 billion nucleotides). As was done for the natural *S.cerevisiae* genome, RNAMotif was used to search these randomized genomes.

Secondary structure prediction

Pknots (24) was used to predict the minimum free energy (MFE) 'fold' of each motif hit identified by RNAMotif. Each motif hit identified by RNAMotif was folded by pknots and assigned a predicted MFE value (MFE in kcal/mol) and a predicted secondary structure.

Randomization and statistical analysis

Each folded motif hit was randomly shuffled and refolded 100 times using pknots (with pseudoknot folding disabled), producing a distribution of random MFEs specific for each of the motif hits. Distributions of random MFE values using pknots with pseudoknots folding disabled were in general not statistically different from those generated using pknots with this option enabled (data not shown), but had considerably shorter generation time, identical energy parameters, and could be run on the same computing platform. Motif hits were then compared to the resulting distribution and assigned a normalized *z*-score:

$$z_R = \frac{X - \bar{x}}{\sigma}, \qquad \mathbf{1}$$

where X is the predicted MFE value for each sequence, $\bar{x}$ is the estimate of the mean for the distribution of MFE values obtained from 100 randomizations, and σ is the SD of random structure MFE values. The normalized value of z_R (*z*-random) obtained provides an estimate of the statistical significance and uniqueness of the predicted structure for the natural sequence: i.e. is the sequence more or less stable than expected by chance (25–32).

Genetic methods and plasmid construction

Escherichia coli strain *DH5α* was used to amplify plasmids, and *E.coli* transformations were performed using the high efficiency method of Inoue *et al*. (33). YPAD and synthetic complete medium (H^-) were used as described previously (34). Isogenic ResGen yeast strains (Invitrogen, Carlsbad, CA) derived from BY4742 (JD1158: *MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0* and JD1181: *MATα upf3::Kan^R his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) were used for *in vivo* measurement of programmed −1 ribosomal frameshifting. All yeast cells were transformed using the alkali cation method (35). Dual luciferase plasmids pJD375 and pJD376 have been described previously (36).

Computationally identified putative −1 PRF signals derived from *BUB3*, *CTS2*, *EST2*, *FKS1*, *FLR1*, *NUP82*, *PPR1*, *SPR6* and *TBF1* were designed with the appropriate restriction sites on the 5′ and 3′ ends (Supplementary Table 1). Furthermore, naturally occurring termination codons were eliminated from the −1 reading frame by shortening the spacer region between slippery site and the putative downstream stimulatory structure by a 1 nt. PAGE purified oligonucleotides (Integrated DNA Technology, Coralville, IA) corresponding to each −1 PRF signal were annealed and gel purified. pJD375, the 'zero-frame' control dual-luciferase frameshift reporter (DLR) plasmid, was used as a vector backbone and each putative PRF signal was cloned into unique *SalI* and *BamHI* restriction sites located in the multiple cloning site (MCS) between the *Renilla* and firefly luciferase orfs. The resulting new PRF-reporter vectors were verified by DNA sequencing (Macrogen, Seoul, Korea). *In vivo* Dual-Luciferase® Reporter Assays (Promega Corporation, Madison, WI) for programmed −1 ribosomal frameshifting were performed as described previously in yeast strain JD1158 (36). Luminescence readings were obtained using a Turner Designs TD20/20™ Luminometer (Sunnyvale, CA). A minimum of 12 replicate assays were carried out for each candidate −1 PRF signal. Statistical analyses of each luciferase dataset followed an established protocol aimed at identifying outliers, validating the statistical assumptions of sample size and distribution, and for the accurate comparison of multiple bicistronic reporter assays (37).

RESULTS

Pattern matching with RNAMotif

The main differences between this study and the previous work by Hammell *et al*. are (i) the availability of a completely annotated yeast genome, (ii) significantly more powerful computational resources, (iii) application of more sophisticated statistical analyses and (iv) a different parameter was employed for the −1 PRF motif. RNAMotif (22) was exploited, and an appropriate albeit somewhat relaxed, 'descriptor' of known viral −1 PRF signals was developed (see Materials and Methods) by analysis of a database of experimentally confirmed recoding signals (23) as the first step toward computationally identifying putative −1 PRF signals. The results of this pattern matching approach identified 10 340 slippery sites in the 6353 annotated coding sequences (CDS) of the yeast genome, 6016 of which are followed by at least one pseudoknot motif. In total, RNAMotif identified 173 452 sequence windows that matched the specified parameters (many are partly overlapping).

Whole genome randomization

To determine the statistical significance of these results, they were compared to what would be expected by chance. One method of identifying statistically significant motifs in nucleic acid sequences is to repeat the initial motif search using a large set of randomized sequences. The frequency of finding the motif in randomized sequences can provide some insight into the likelihood that a match in a natural sequence occurs by chance. In this report, a conservative approach was applied by randomizing the whole yeast CDS genome using seven different strategies so as to not introduce bias due to choice of any one method. All of the randomized genomes contained the same number of ORFs (rORF) as the natural yeast genome and the same number of total

Table 1. The yeast genome has a significant number of putative programmed −1 ribosomal frameshift signals compared to randomized genomes created using any one of seven different randomization strategies

	RNAMotif	SD	*P*-value
S.cerevisiae	6016	—	—
noBias	3044	64.07	<0.01
nShuffle	4567	70.84	<0.01
nBias	4660	65.89	<0.01
cShuffle	6551	85.13	0.02
sBias	6580	82.13	0.02
cBias	6639	86.52	0.02
dnBias	6774	88.16	0.01

RNAMotif, the number of motif hits using our descriptor of functional −1 PRF signals (22); stdev, standard deviation of for each randomization strategy. Seven methods for randomization were used. These were: NoBias, randomized ORFs with unbiased nucleotide bias; ntShuffle, nucleotides from each natural ORF are shuffled by triplicate mononucleotide permutations; ntBias, randomized ORFs using the CDS single-nucleotide frequency; cdnShuffle, codons from each natural ORF are shuffled by triplicate monocodon permutation; SilentBias, a silent bias where the codons are randomized in place so as to maintain protein coding sequence; cdnBias, randomized ORFs using the observed CDS codon usage bias; diNuc, randomized ORFs generated using the observed CDS dinucleotide frequency.

nucleotides in the CDS sequence space. Furthermore, rORFs with in-frame premature termination codons were discarded and randomly re-generated until full length read-through sequences were obtained. The results (Table 1) show that the actual number of motif hits found is statistically different when compared to any of the seven randomized datasets; suggesting that the prevalence of −1 PRF signals may be under multiple selective pressures.

Each of the randomization types designed to mimic the natural CDS of yeast (cShuffle, sBias, cBias and dnBias; see Materials and Methods), retained the information content of the yeast genome. These randomization strategies generated genomes that harbored more −1 PRF signals than are actually found in the natural genome. This suggests a selective pressure against the random acquisition of functional −1 PRF signals in yeast; i.e. the yeast genome would be expected to have more −1 PRF signals than were actually observed. This is consistent with the notion that −1 PRF signals can lead to aberrant translation and (most likely) dysfunctional proteins. In contrast, randomization strategies that seek to mimic the overall genome-wide or individual CDS nucleotide bias (nBias and nShuffle, respectively) produce random genomes with significantly fewer −1 PRF signals than are actually observed. If there were strong and genome-wide evolutionary pressures against the presence of any −1 PRF signals, then they would be expected to be significantly underrepresented in the yeast genome and statistically indistinguishable from the nShuffle and nBias randomization datasets. However, this is not the case. This set of comparisons suggests that there may be evolutionary pressure for the maintenance of certain classes of existing frameshift signals. In addition, randomized genomes using an unbiased nucleotide frequency (noBias) were generated as a negative control. These random genomes contained far fewer −1 PRF signals than observed for the actual yeast genome and far less than any of the other randomization strategies. In sum, the number of slippery sites followed by at least one pseudoknot motif (6016) present in the actual yeast genome is highly statistically significant ($P \leqslant 0.02$) when compared to the number of expected −1 PRF signals for all of the randomization strategies employed (Table 1). Therefore, although unexpectedly large, this analysis suggests that at least some of the predicted −1 PRF signals have been functionally selected for.

Secondary structure prediction

The next step was to assign additional layers of predictive metrics to the dataset so as to enhance the ability to identify functional −1 PRF signals for empirical testing. The first step was to assign a MFE value to each motif hit identified by RNAMotif. This was not a trivial task since nearly all known −1 PRF signals require an mRNA pseudoknot (7) and pseudoknot prediction represents a well known and computationally difficult problem (38). However, pknots (24), an algorithmic extension of mfold (39) is capable of predicting H-type pseudoknots of the type that are generally found associated with functional −1 PRF signals. Coupled with a set of scripts written in PERL (40), pknots was able to fold every sequence window identified as a potential motif hit by RNAMotif (173 452 sequence windows) in ∼5000 CPU h. (∼5 months) using the computational resources available at the National Cancer Institute's Advanced Biomedical Computing Center in Frederick, MD. Once the initial folding was completed, the dataset was then reduced to a structurally non-redundant dataset of 66 842 structures. The nearly 3-fold reduction in the data was possible due to the huge number of overlapping motif hits initially made by RNAMotif. These analyses provide each non-redundant RNAMotif match with a predicted RNA secondary structure and MFE value. The overall distribution of all MFE values determined by pknots for the most stable predicted secondary structures (lowest kcal/mol) for each structure 3′ from the slippery motif is shown in Supplementary Figure 1A and fits a normal distribution. The distribution of base pair counts for each structure fits an extreme-value distribution and is shown in Supplementary Figure 1B. The feature correlations and summary statistics of these 10 340 predicted structures are shown in Supplementary Tables 2A and B, respectively.

Statistical significance of predicted MFE values

To identify statistically significant motif hits, *z*-scores (z_R) were calculated for each predicted RNA secondary structure folded by pknots (see Materials and Methods). For each candidate signal, the MFE value of the predicted structure was compared to the distribution of MFE values obtained from 100 permutations (mononucleotide shuffles) of the same sequence using an implementation in PERL of a similar algorithm previously described (28). The randomization approach disrupts the nucleotide base order and any potential secondary structure for each input sequence but preserves the exact mononucleotide count of each base within the shuffling window. Significance scores derived from permutation shuffling approaches such as this have previously been successful in finding biologically meaningful RNA structures from primary sequence data both by ourselves (41) and several other research groups (25,26,28,42). Furthermore, it is expected that this measure of significance is sufficient since functional secondary structures in mRNA sequences are considered more stable than random sequence and are under selective

pressure (29,31,43,44). It should be noted, however, that several reports have indicated that this randomization strategy is not accurate for estimating the significance of RNA secondary structures in-general and that a superior method of randomization lies in preserving both mono- and dinucleotide ratios (32,45–47). Nonetheless, for the purposes of this study the randomization strategy employed for the calculation of z_R was adequate. For this dataset, the randomization step was limited to 100 permutations per sequence due to the sheer number of input sequences that required z_R scores. This reasonably estimated a normal distribution of MFE values for each input sequence and a probability plot correlation coefficient (PPCC) goodness-of-fit test (48) was carried out for each distribution to statistically verify each estimation of a normal distribution. A PPCC $\geqslant 0.98$ was found for greater than 99% of all the candidate signals in the database indicating that 100 random shuffles was sufficient for good estimates of z_R (data not shown).

Any $z_R \leqslant -1.65$ indicates a structure that is more stable than expected ($P < 0.05$). The distribution of z_R scores for all candidate PRF signals fits a normal distribution (Supplementary Figure 1C, PPCC = 0.98). A total of 3228 candidate signals (out of 66 842 non-redundant structures) include putative structures that meet or exceed the criteria for significance, having z_R scores in the range of $\boldsymbol{z_R} = [-7.10, -1.65]$. These significant motif hits are distributed among 2025 ORFs. A total of 1203 individual slippery sites in 751 ORFs are found to have more than one significant structure immediately downstream. Each of these statistically significant structures (and the associated 5′ slippery sites) are considered candidate −1 PRF signals (cPRF) open for further investigation.

An interesting finding from this analysis is that statistically significant motif hits do not necessarily have low MFE values; a result that was previously shown to be true for structural RNAs in general (25,28,29). We therefore sought to filter the list of putative −1 PRF signals further by comparing z_R scores and MFE values, which are only weakly correlated features in the database (Correlation = 0.53, Supplementary Table 3A), similar to an approach previously employed (27). Energetically strong candidates with statistically significant predicted secondary structures are considered strong-candidate −1 PRF signals (Figure 2). From this analysis, 1706 strong candidate signals were identified with significant $z_R \leqslant -1.65$ and whose MFE values are in the lowest 25% (MFE $\leqslant$ 17.3 kcal/mol). These strong candidate signals are distributed among 1275 individual ORFs, where 320 ORFs have two or more strong signals.

Selection of candidate signals for empirical analyses

Computationally generated information is best when empirically tested. To this end, nine candidate signals possessing a wide range of feature statistics were selected for empirical testing. First and foremost, −1 PRF signals were selected from genes having scorable phenotypes when under- or over-expressed. Second, eight of the nine candidate signals chosen are predicted to fold into a pseudoknot, the exception being the signal chosen from *FKS1*. Third, not all the selected signals should fully meet the criteria outlined above for strong candidate signals. For example, the two signals from

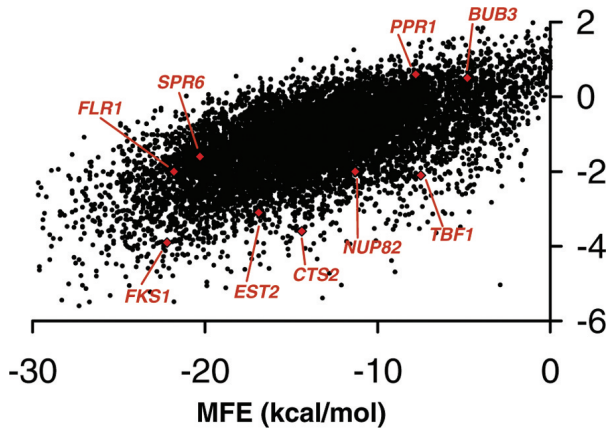

Figure 2. Scatterplot of MFE values (predicted using pknots, (24) versus z_R scores for 10 340 candidate −1 PRF signals demonstrates the weak correlation between these two feature statistics (see Supplementary Table 2B). The red diamonds and associated labels indicate the location and parental gene of nine sequences empirically tested for frameshifting. The hypothetical distributions were created using summary statistics from Supplementary Table 2B.

FLR1 and *SPR6* met all of the criteria for strong-candidate −1 PRF signals having $z_R \leqslant -1.65$ and predicted MFE values in the lowest 25% of all structures in the PRFdb. The signals from *CTS2*, *EST2*, *NUP82* and *TBF1* meet less stringent criteria in that, although they are not in the first quartile of the most stable structures, they nonetheless are considered significant with $z_R \leqslant -1.65$. Candidate signals from *BUB3* and *PPR1* were chosen because they specifically do not meet any of the criteria above. The predicted slippery sites and associated secondary structures are shown in Figure 3. The feature statistics of each candidate signal are summarized in Supplementary Table 3.

Testing for frameshifting

Each of the nine candidate −1 PRF signals were cloned into pJD375, a dual-luciferase frameshift reporter plasmid (DLR), and their abilities to promote −1 PRF was measured after transformation into a wild-type yeast strain (JD1158) as described previously (36). Briefly, the ratio of firefly to *Renilla* luciferase expression promoted by −1 PRF signal containing reporters is normalized to a 'zero-frame' control reporter (pJD375), and these ratios are statistically tested for normalcy, sample size and significance as described previously (37). At least ten replicate experiments were carried out for each reporter. The results demonstrate that every signal containing a predicted mRNA pseudoknot promoted −1 PRF at levels that significantly exceeded non-programmed (or background) frameshifting (Figure 4A and B). In contrast, the sequence derived from *FKS1*, which is not predicted to contain a pseudoknot, did not promote measurable frameshifting. In a broad sense, the experimental data divides the signals into high-, medium- and low-efficiency −1 PRF signals. The signals cloned from *CTS2*, *EST2* and *PPR1* promoted −1 PRF at ∼64, 56 and 43%, respectively (Figure 4A). Although these rates of −1 PRF are extremely high, concurrent studies in our laboratory show levels of −1 PRF as high as 91% in a variant of the SARS-CoV

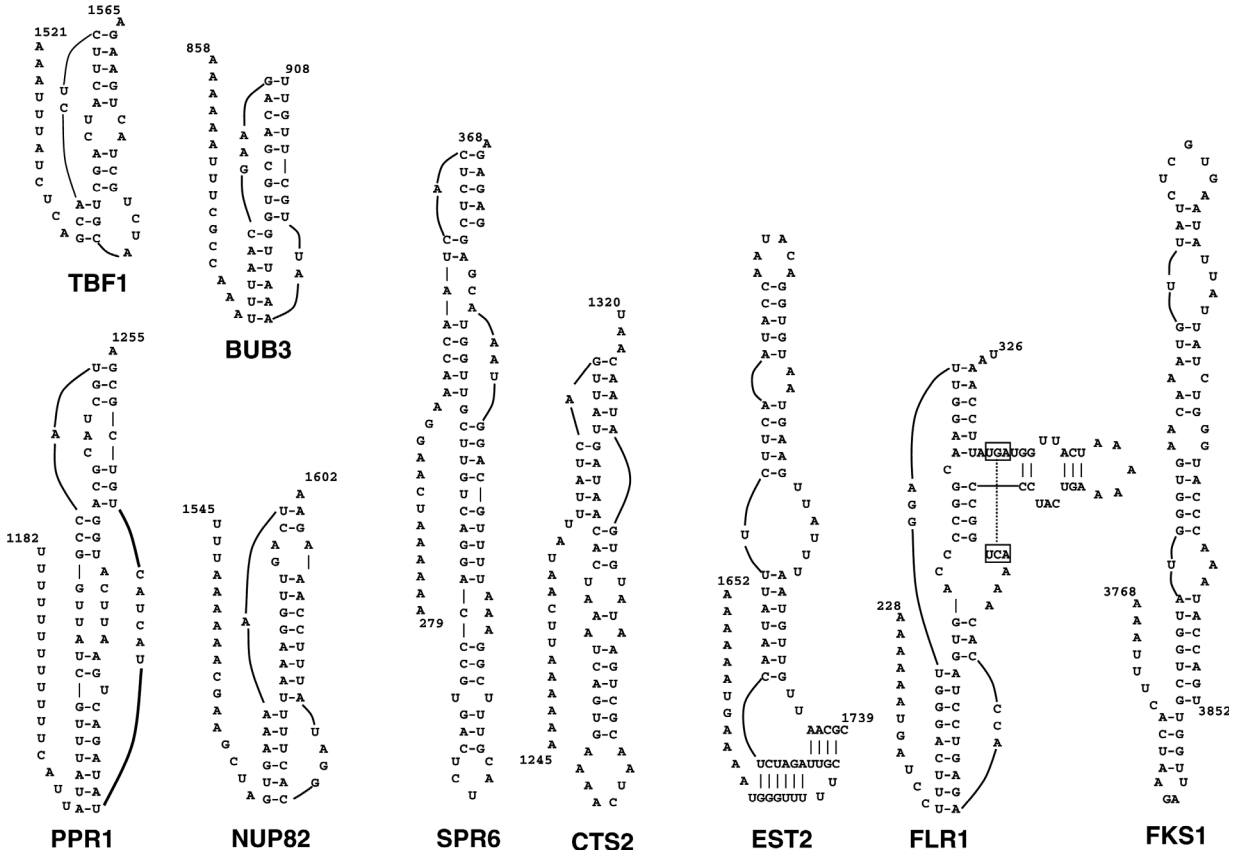

Figure 3. Nine examples of candidate −1 PRF signals chosen to generally represent the diversity of features present in the PRFdb. Gene names are shown with RNA sequence and corresponding CDS nucleotide start and stop locations. The predicted structure is shown for each empirically tested candidate signal. See Supplementary Table 3 for statistical data.

−1 PRF signal (E. P. Plant and J. D. Dinman, unpublished data). The signal cloned from *TBF1* falls in the middle-range (5.2%), while the three remaining functional signals promoted −1 PRF at levels between 0.4–0.9% (Figure 4B). For purposes of comparison, the well-characterized −1 PRF signal from the yeast L-A virus promoted 9.1% frameshifting, while values for out-of-frame controls are <0.02%. It is important to note that in order to measure frameshifting it was necessary to re-engineer the frameshift signals to bypass −1 frame encoded termination codons present downstream of the slippery sites, and that these were all located in the mRNA pseudoknots. Mutating these to sense codons would cause unknown changes to pseudoknot structures, which would uncontrollably affect frameshifting. In contrast, although deleting delete one base in the spacer region should change −1 PRF efficiencies (usually downward, see Ref. (49), such changes should not completely abrogate frameshifting, nor should they create functional −1 PRF signals *de novo*. Thus, although the values presented here cannot be taken as absolute, the important issue is qualitative: these mRNA elements can indeed promote efficient frameshifting.

The PRFdb and data availability

The PRFdb (http://dinmanlab.umd.edu/prfdb) is a publicly available database that stores the results of the bioinformatics data presented in this report. This online resource allows interested readers to search for and analyze candidate −1 PRF signals in the genome of *S.cerevisiae*. Finally, the PRFdb contains a searchable list of strong candidate −1 PRF signals that may warrant further empirical investigations.

DISCUSSION

Programmed ribosomal frameshifting was first identified as a translational phenomenon in the *Rous sarcoma* virus over two decades ago (50). Since then, it has been shown to be a general mechanism of gene regulation utilized by a wide variety of RNA viruses (1–4). Frameshifting has also been demonstrated to be functionally important for the expression of a growing list of prokaryotic (51–53), archaeal (54) and eukaryotic genes (13,14). Thus, it is becoming increasingly apparent that PRF is a fundamental mechanism of post-transcriptional gene regulation and is present in every branch of the tree of life. The need to identify PRF signals in higher organisms has grown in importance as we have become more aware of their prevalence. In response, there have been numerous computational studies aimed at identifying PRF signals (17–21,55–57). Each study has met with varying degrees of success, but empirical testing of predicted PRF signals suggest that there are indeed functional, and previously unannotated, PRF signals in a variety of contexts within

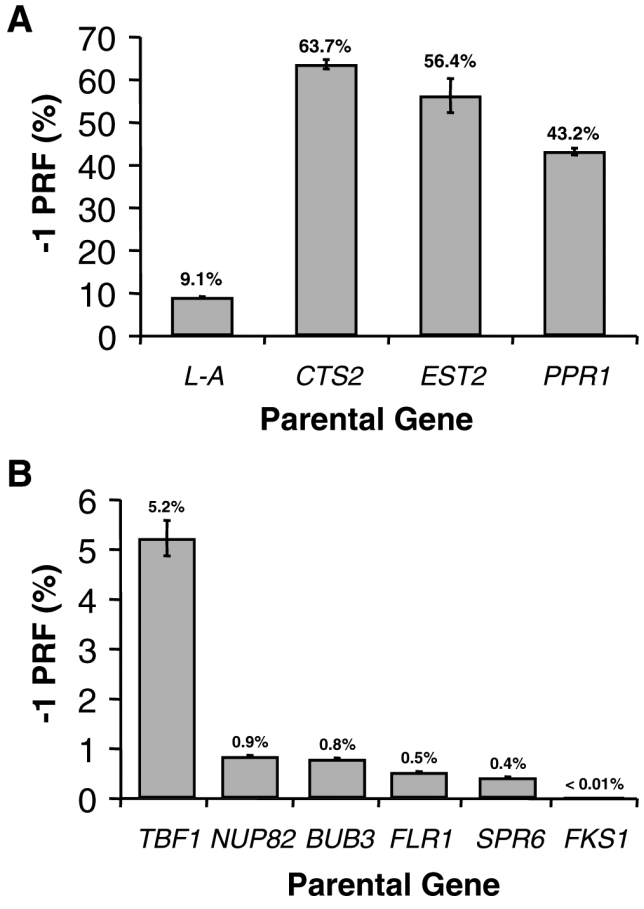

Figure 4. Measurement of −1 PRF efficiency for nine candidate signals. (**A**) High-efficiency frameshifting including the frameshift signal from the endogenous yeast L-A virus. (**B**) Medium- and low-efficiency frameshifting including the sequence from the *FKS1* gene that did not promote −1 PRF above background levels. The parental genes of each candidate signal are indicated with the percentage of −1 PRF efficiency as was measured using a dual-luciferase reporter assay system (36,37).

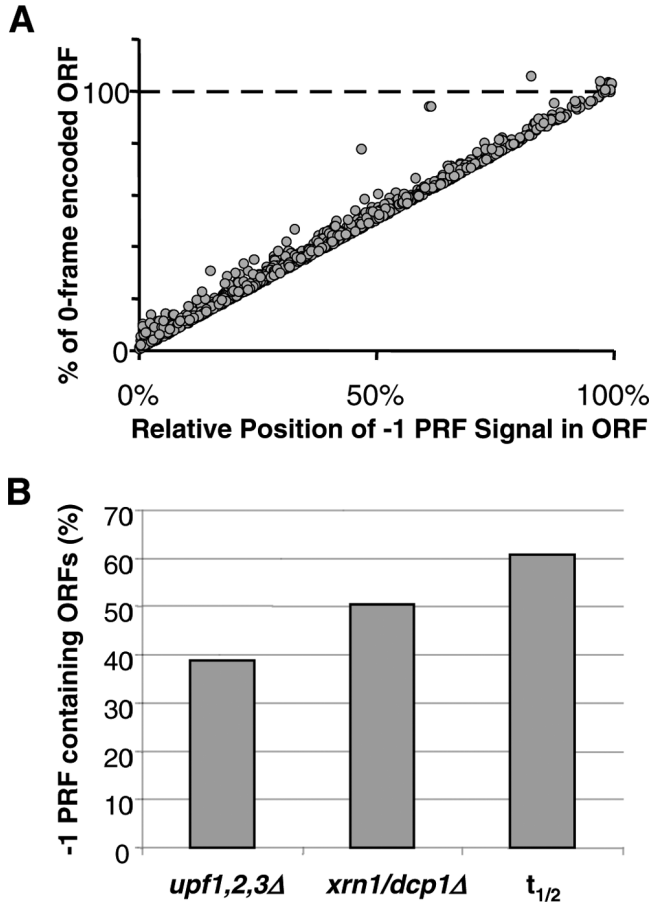

Figure 5. (**A**) The CDS of *S.cerevisiae* is not prone to lengthy out-of-frame translation. The relative positions of candidate −1 PRF signals from the start codon of each ORF compared to the expected overall change in peptide length if a frameshifting event were to occur. (**B**) Fraction of ORFs containing high probability −1 PRF signals represented as mRNAs stabilized in strains deficient in NMD (*upf1*Δ, *upf2*Δ or *upf3*Δ) (60), No-go decay (*xrn1*Δ or *dcp1*Δ) (60), or having half-lives less than the yeast transcriptome average ($t_{1/2}$) (70).

the coding regions of genes derived from higher organisms. With the exception of the earliest study by Hammell *et al*. all of these studies have focused on recoding in the 'viral-context': i.e. they were aimed towards finding PRF signals predicted to direct ribosomes into a new reading frame so as to produce functional alternative C-terminal extensions of the native proteins. The study by Hammell *et al*. was context neutral, focusing instead on searching for mRNA motifs that resembled known viral −1 PRF signals.

The findings presented by this study suggest that PRF signals can function efficiently in a number of different ways. For example, while sequences that are predicted to fold into strong, statistically significant, pseudoknotted mRNA structures serve as efficient stimulators of −1 PRF (e.g. signals from *CTS2* and *EST2*), the presence of multiple overlapping slippery sites can also have an equally strong effect, even if the stimulatory structure is not ideal (e.g. signals from *BUB3*, *PPR1* and *NUP82*). Most importantly, it appears that the presence of a pseudoknot as the 'most stable' structure following a slippery site is critical, providing further support for the 'torsional restraint' model of −1 PRF (8). This is further evidenced by the fact that the very energetically favorable and highly significant structure derived from *FKS1* failed to promote detectable −1 PRF *in vivo*. This last point may be because there is no predicted pseudoknot structure with an MFE value lower than that of the predicted non-pseudoknotted structure immediately following the slippery site of interest in *FKS1* (Figure 3).

The PRFdb was constructed to serve as a repository for all the predicted structures, slippery sites and statistical data in this study. This database is accessible via the Internet at http://dinmanlab.umd.edu/prfdb. Currently, visitors are limited to searching the existing data for putative −1 PRF signals in the yeast genome. It is expected, however, that this service will be expanded in the future to include the genomes of seven additional budding yeast species in addition to *S.cerevisiae*, the human genome, and several other 'model system' genomes.

While the current study revisits the original question posed by Hammell *et al*. (i.e. how often are functional −1 PRF signals present in the yeast genome?), it also asks an important second question: are genome encoded −1 PRF signals capable of promoting −1 PRF, and if so, how might this affect the

expression of the mRNAs encoding them? Analysis of the PRFdb from the perspective of alternative recoding reveals that greater than 99% of the expected outcomes of −1 PRF would result in premature termination (Figure 5A). The prevalence of out-of-frame termination signals is not unexpected since the average distance a ribosome can continue elongation in an alternative reading frame is ∼6 codons in either the +1 or −1 frame for all CDS in yeast (data not shown). Only 10 −1 PRF signals out of 10 340 potential slippery sites are predicted to bypass the normal zero-frame termination codon and encode an alternative C-terminal extension (i.e. −1 PRF in the viral context). However, BLAST analyses (58) revealed that none of these extensions are predicted to encode functional alternative protein domains (data not shown). This suggests that although potential −1 PRF signals are widespread in the yeast genome, they are almost uniformly predicted to direct ribosomes to a premature termination signals. We have previously shown that such PRF events are sufficient to target transcripts to the nonsense-mediated mRNA decay pathway for rapid degradation (59). Thus, we suggest that −1 PRF may be used to post-transcriptionally regulate gene expression of cellular encoded mRNAs. In support of this notion, natural mRNA substrates for NMD that do not contain in-frame PTCs have been discovered in both the yeast and human transcriptomes (60–67), suggesting that regulation of mRNA expression by NMD is broadly conserved and is used to regulate a variety of physiological processes. If this is true, then these mRNAs should be well represented in DNA microarray databases identifying mRNAs that are stabilized when this pathway has been abrogated, e.g. *upf1*Δ *upf2*Δ, or *upf3*Δ (60). Consistent with this hypothesis, the overlap between mRNAs stabilized in these genetic backgrounds and ORFs containing high probability −1 PRF signals is 38.8% (278 of 717 genes) (Figure 5B). Also implicit in this model is that modulation of PRF efficiency could be used as the controlling effector of transcript degradation rates. This model could be applied to a variety of biological examples where the stability of individual mRNAs or whole classes of mRNAs require a flexible stability threshold responsive to environmental cues.

Independent of NMD, It has also been suggested that specific sequences present in coding regions of mRNAs are capable of translationally stalling ribosomes long enough to direct them to be endonucleolytically cleaved and specifically degraded in both prokaryotic and eukaryotic organisms (68,69), a process termed 'No-go decay'. Many of the predicted candidate −1 PRF signals identified in the current work are predicted to be more stable than the mRNA structure used by Doma & Parker (data not shown) and thus would be expected to be similarly capable to stall translating ribosomes. Thus, it is also possible that mRNAs containing these extremely stable mRNA structures identified in the current study may also be targeted for No-go decay. If this is true, then these mRNAs should be stabilized in cells in which this pathway has been abrogated, e.g. *xrn1*Δ or *dcp1*Δ (60). Figure 5B shows the overlap between mRNAs stabilized in these genetic backgrounds and genes containing high probability −1 PRF signals is 50.5% (362 of 717 genes), consistent with this hypothesis. Furthermore, regardless of the mRNA destabilization pathway, −1 PRF signal containing mRNAs should be inherently less stable than those that do not contain this motif. Comparison with the yeast transcriptome half-life microarray database (70) reveals that 60.8% (327 of 538) of the mRNAs containing high probability −1 PRF signals have half-lives less than the transcriptome average (<26 min) (Figure 5B). Conversely, it is possible that the presence of a slippery site just upstream from strong secondary structures may be the specific feature that allows for such mRNAs to evade being subject to this pathway (7). In conclusion, it is reasonable to envision that a general function of PRF signals in the coding regions of eukaryotic mRNAs is to act as post-transcriptional capacitors of gene expression.

AUTHOR CONTRIBUTIONS

J. L. Jacobs and J. D. Dinman conceived and designed experiments. J. L. Jacobs, A. T. Belew and R. Rakauskaite performed experiments and contributed reagents/materials. J. L. Jacobs and A. T. Belew wrote source code, developed the PRFdb and analyzed the computational and empirical data. J. L. Jacobs and J. D. Dinman wrote the paper.

ACCESSION NUMBERS

The primary SDG accession numbers of genes from which the entire CDS or a subssequence of the CDS was used in this study are: *BUB3* (#S000005552), *CTS2* (#S000002779), *EST2* (#S000004310), *FLR1* (#S000000212), *FKS1* (#S000004334), *NUP82* (#S000003597), *PPR1* (#S000004004), *SPR6* (#S000000917), *TBF1* (#S000006049).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

This work was supported by a fellowship to J.L.J. from the National Libraries of Medicine (F37 LM8333), grants to J.D.D. from the National Institutes of Health (R05 GM58859 and R21 GM068123) and the National Science Foundation (MCB-0084559). A.T.B. was partially supported by funds from the National Institutes of Health (T32 AI51967). The authors wish to thank Kristi Muldoon-Jacobs, Ewan Plant and John Russ for advice, suggestions and helpful reviewing of the manuscript. The authors also want to acknowledge the helpful technical support and computing resources from the National Cancer Institute's Advanced Biomedical Computing Center in Frederick, MD. Funding to pay the Open Access publication charges for this article was provided by a grant to J.D.D. from the National Institutes of Health (GM58859).

Conflict of interest statement. None declared.

REFERENCES

1. Baranov,P.V., Gesteland,R.F. and Atkins,J.F. (2002) Recoding: translational bifurcations in gene expression. *Gene*, **286**, 187–201.

2. Cobucci-Ponzano,B., Rossi,M. and Moracci,M. (2005) Recoding in archaea. *Mol. Microbiol.*, **55**, 339–348.
3. Harger,J.W., Meskauskas,A. and Dinman,J.D. (2002) An 'integrated model' of programmed ribosomal frameshifting and post-transcriptional surveillance. *Trends Biochem. Sci.*, **27**, 448–454.
4. Namy,O., Rousset,J.P., Napthine,S. and Brierley,I. (2004) Reprogrammed genetic decoding in cellular gene expression. *Mol. Cell*, **13**, 157–168.
5. Kontos,H., Napthine,S. and Brierley,I. (2001) Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency. *Mol. Cell. Biol.*, **21**, 8657–8670.
6. Lopinski,J.D., Dinman,J.D. and Bruenn,J.A. (2000) Kinetics of ribosomal pausing during programmed −1 translational frameshifting. *Mol. Cell. Biol.*, **20**, 1095–1103.
7. Plant,E.P., Jacobs,K.L.M., Harger,J.W., Meskauskas,A., Jacobs,J.L., Baxter,J.L., Petrov,A.N. and Dinman,J.D. (2003) The 9-angstrom solution: How mRNA pseudoknots promote efficient programmed −1 ribosomal frameshifting. *RNA*, **9**, 168–174.
8. Plant,E.P. and Dinman,J.D. (2005) Torsional restraint: a new twist on frameshifting pseudoknots. *Nucleic Acids Res.*, **33**, 1825–1833.
9. Baril,M., Dulude,D., Steinberg,S.V. and Brakier-Gingras,L. (2003) The frameshift stimulatory signal of human immunodeficiency virus type 1 group O is a pseudoknot. *J. Mol. Biol.*, **331**, 571–583.
10. Kollmus,H., Hentze,M.W. and Hauser,H. (1996) Regulated ribosomal frameshifting by an RNA-protein interaction. *RNA*, **2**, 316–323.
11. Manktelow,E., Shigemoto,K. and Brierley,I. (2005) Characterization of the frameshift signal of Edr, a mammalian example of programmed −1 ribosomal frameshifting. *Nucleic Acids Res.*, **33**, 1553–1563.
12. Matsufuji,S., Matsufuji,T., Wills,N.M., Gesteland,R.F. and Atkins,J.F. (1996) Reading two bases twice: mammalian antizyme frameshifting in yeast. *EMBO J.*, **15**, 1360–1370.
13. Shigemoto,K., Brennan,J., Walls,E., Watson,C.J., Stott,D., Rigby,P.W.J. and Reith,A.D. (2001) Identification and characterisation of a developmentally regulated mammalian gene that utilises −1 programmed ribosomal frameshifting. *Nucleic Acids Res.*, **29**, 4079–4088.
14. Wills,N.M., Moore,B., Hammer,A., Gesteland,R.F. and Atkins,J.F. (2006) A functional −1 ribosomal frameshift signal in the human paraneoplastic Ma3 gene. *J. Biol. Chem.*, **281**, 7082–7088.
15. Lundblad,V. and Morris,D.K. (1997) Programmed translational frameshifting in a gene required for yeast telomere replication. *Curr. Biol.*, **7**, 969–976.
16. Morris,D.R. and Geballe,A.P. (2000) Upstream open reading frames as regulators of mRNA translation. *Mol. Cell. Biol.*, **20**, 8635–8642.
17. Gao,X., Havecker,E.R., Baranov,P.V., Atkins,J.F. and Voytas,D.F. (2003) Translational recoding signals between gag and pol in diverse LTR retrotransposons. *RNA*, **9**, 1422–1430.
18. Gurvich,O.L., Baranov,P.V., Zhou,J., Hammer,A.W., Gesteland,R.F. and Atkins,J.F. (2003) Sequences that direct significant levels of frameshifting are frequent in coding regions of Escherichia coli. *EMBO J.*, **22**, 5941–5950.
19. Hammell,A.B., Taylor,R.L., Peltz,S.W. and Dinman,J.D. (1999) Identification of putative programmed −1 ribosomal frameshift signals in large DNA databases. *Genome Res.*, **9**, 417–427.
20. Moon,S., Byun,Y., Kim,H.J., Jeong,S. and Han,K. (2004) Predicting genes expressed via −1 and +1 frameshifts. *Nucleic Acids Res.*, **32**, 4884–4892.
21. Shah,A.A., Giddings,M.C., Parvaz,J.B., Gesteland,R.F., Atkins,J.F. and Ivanov,I.P. (2002) Computational identification of putative programmed translational frameshift sites. *Bioinformatics*, **18**, 1046–1053.
22. Macke,T.J., Ecker,D.J., Gutell,R.R., Gautheret,D., Case,D.A. and Sampath,R. (2001) RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res.*, **29**, 4724–4735.
23. Baranov,P.V., Gurvich,O.L., Hammer,A.W., Gesteland,R.F. and Atkins,J.F. (2003) Recode 2003. *Nucleic Acids Res.*, **31**, 87–89.
24. Rivas,E. and Eddy,S.R. (1999) A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J. Mol. Biol.*, **285**, 2053–2068.
25. Le,S.Y., Chen,J.H. and Maizel,J.V. (1989) Thermodynamic stability and statistical significance of potential stem–loop structures situated at the frameshift sites of retroviruses. *Nucleic Acids Res.*, **17**, 6143–6152.
26. Le,S.Y., Liu,W.M., Chen,J.H. and Maizel,J.V.,Jr (2001) Local thermodynamic stability scores are well represented by a non-central student's t distribution. *J. Theor. Biol.*, **210**, 411–423.
27. Le,S.Y., Zhang,K. and Maizel,J.V.,Jr (2002) RNA molecules with structure dependent functions are uniquely folded. *Nucleic Acids Res.*, **30**, 3574–3582.
28. Seffens,W. and Digby,D. (1999) mRNAs have greater negative folding free energies than shuffled or codon choice randomized sequences. *Nucleic Acids Res.*, **27**, 1578–1584.
29. Schultes,E.A., Hraber,P.T. and LaBean,T.H. (1999) Estimating the contributions of selection and self-organization in RNA secondary structure. *J. Mol. Evol.*, **49**, 76–83.
30. Tuplin,A., Wood,J., Evans,D.J., Patel,A.H. and Simmonds,P. (2002) Thermodynamic and phylogenetic prediction of RNA secondary structures in the coding region of hepatitis C virus. *RNA*, **8**, 824–841.
31. Chamary,J.V. and Hurst,L.D. (2005) Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biol.*, **6**, R75.
32. Freyhult,E., Gardner,P.P. and Moulton,V. (2005) A comparison of RNA folding measures. *BMC Bioinformatics*, **6**, 241.
33. Inoue,H., Nojima,H. and Okayama,H. (1990) High efficiency transformation of Escherichia coli with plasmids. *Gene*, **96**, 23–28.
34. Dinman,J.D., Icho,T. and Wickner,R.B. (1991) A −1 ribosomal frameshift in a double-stranded RNA virus forms a Gag-pol fusion protein. *Proc. Natl Acad. Sci. USA*, **88**, 174–178.
35. Ito,H., Fukuda,Y., Murata,K. and Kimura,A. (1983) Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.*, **153**, 163–168.
36. Harger,J.W. and Dinman,J.D. (2003) An *in vivo* dual-luciferase assay system for studying translational recoding in the yeast *Saccharomyces cerevisiae*. *RNA*, **9**, 1019–1024.
37. Jacobs,J.L. and Dinman,J.D. (2004) Systematic analysis of bicistronic reporter assay data. *Nucleic Acids Res.*, **32**, e160–e170.
38. Lyngso,R.B. and Pedersen,C.N. (2000) RNA pseudoknot prediction in energy-based models. *J. Comput. Biol.*, **7**, 409–427.
39. Mathews,D.H., Sabina,J., Zuker,M. and Turner,D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.
40. Wall,L., Christiansen,T. and Orwant,J. (2000) *Programming Perl*, 3rd edn. O'Reilly & Associates, Sebastopol, CA.
41. Plant,E.P., Perez-Alvarado,G.C., Jacobs,J.L., Mukhopadhyay,B., Hennig,M. and Dinman,J.D. (2005) A Three-Stemmed mRNA Pseudoknot in the SARS Coronavirus Frameshift Signal. *PLoS. Biol.*, **3**, 1012–1023.
42. Barrette,I., Poisson,G., Gendron,P. and Major,F. (2001) Pseudoknots in prion protein mRNAs confirmed by comparative sequence analysis and pattern searching. *Nucleic Acids Res.*, **29**, 753–758.
43. Katz,L. and Burge,C.B. (2003) Widespread selection for local RNA secondary structure in coding regions of bacterial genes. *Genome Res.*, **13**, 2042–2051.
44. Ringner,M. and Krogh,M. (2005) Folding free energies of 5′-UTRs impact post-transcriptional regulation on a genomic scale in yeast. *PLoS. Comput. Biol.*, **1**, e72.
45. Workman,C. and Krogh,A. (1999) No evidence that mRNAs have lower folding free energies than random sequences with the same dinucleotide distribution. *Nucleic Acids Res.*, **27**, 4816–4822.
46. Rivas,E. and Eddy,S.R. (2000) The language of RNA: a formal grammar that includes pseudoknots. *Bioinformatics*, **16**, 334–340.
47. Clote,P., Ferre,F., Kranakis,E. and Krizanc,D. (2005) Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency. *RNA*, **11**, 578–591.
48. Filliben,J.J. (1975) Probability plot correlation coefficient test for normality. *Technometrics*, **17**, 111–117.
49. Brierley,I., Jenner,A.J. and Inglis,S.C. (1992) Mutational analysis of the 'slippery-sequence' component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.*, **227**, 463–479.
50. Jacks,T. and Varmus,H.E. (1985) Expression of the Rous Sarcoma Virus pol gene by ribosomal frameshifting. *Science*, **230**, 1237–1242.
51. Tsuchihashi,Z. and Kornberg,A. (1990) Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proc. Natl Acad. Sci. USA*, **87**, 2516–2520.

52. Blinkowa,A.L. and Walker,J.R. (1990) Programmed ribosomal frameshifting generates the Escherichia coli DNA polymerase III gamma subunit from within the tau subunit reading frame. *Nucleic Acids Res*., **18**, 1725–1729.
53. Sekine,Y. and Ohtsubo,E. (1989) Frameshifting is required for production of the transposase encoded by insertion sequence 1. *Proc. Natl Acad. Sci. USA*, **86**, 4609–4613.
54. Cobucci-Ponzano,B., Trincone,A., Giordano,A., Rossi,M. and Moracci,M. (2003) Identification of an archaeal alpha-L-fucosidase encoded by an interrupted gene. Production of a functional enzyme by mutations mimicking programmed −1 frameshifting. *J. Biol. Chem*., **278**, 14622–14631.
55. Moon,S., Byun,Y., Kim,H.-J., Jeong,S. and Han,K. (2004) Predicting genes expressed via −1 and +1 frameshifts. *Nucleic Acids Res*., **32**, 4884–4892.
56. Namy,O., Duchateau-Nguyen,G., Hatin,I., Hermann-Le Denmat,S., Termier,M. and Rousset,J.P. (2003) Identification of stop codon readthrough genes in *Saccharomyces cerevisiae*. *Nucleic Acids Res*., **31**, 2289–2296.
57. Harrison,P., Kumar,A., Lan,N., Echols,N., Snyder,M. and Gerstein,M. (2002) A small reservoir of disabled ORFs in the yeast genome and its implications for the dynamics of proteome evolution. *J. Mol. Biol*., **316**, 409–419.
58. Altchul,S.F., Gish,W., Miller,E., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol*., **215**, 403–410.
59. Plant,E.P., Wang,P., Jacobs,J.L. and Dinman,J.D. (2004) A programmed −1 ribosomal frameshift signal can function as a cis-acting mRNA destabilizing element. *Nucleic Acids Res*., **32**, 784–790.
60. He,F., Li,X., Spatrick,P., Castillo,R., Dong,S. and Jacobson,A. (2003) Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5′ to 3′ mRNA decay pathways in yeast. *Mol. Cell*, **12**, 1439–1452.
61. Lelivelt,M.J. and Culbertson,M.R. (1999) Yeast Upf proteins required for RNA surveillance affect global expression of the yeast transcriptome. *Mol. Cell. Biol*., **19**, 6710–6719.
62. Wittmann,J., Hol,E.M. and Jack,H.M. (2006) hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. *Mol. Cell. Biol*., **26**, 1272–1287.
63. Mendell,J.T., Sharifi,N.A., Meyers,J.L., Martinez-Murillo,F. and Dietz,H.C. (2004) Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nature Genet*., **36**, 1073–1078.
64. Pan,Q., Saltzman,A.L., Kim,Y.K., Misquitta,C., Shai,O., Maquat,L.E., Frey,B.J. and Blencowe,B.J. (2006) Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. *Genes Dev*., **20**, 153–158.
65. Taylor,R., Kebaara,B.W., Nazarenus,T., Jones,A., Yamanaka,R., Uhrenholdt,R., Wendler,J.P. and Atkin,A.L. (2005) Gene set coregulated by the *Saccharomyces cerevisiae* nonsense-mediated mRNA decay pathway. *Eukaryot. Cell*, **4**, 2066–2077.
66. Moriarty,P.M., Reddy,C.C. and Maquat,L.E. (1998) Selenium deficiency reduces the abundance of mRNA for Se-dependent glutathione peroxidase 1 by a UGA-dependent mechanism likely to be nonsense codon-mediated decay of cytoplasmic mRNA. *Mol. Cell. Biol*., **18**, 2932–2939.
67. Kim,Y.K., Furic,L., Desgroseillers,L. and Maquat,L.E. (2005) Mammalian Staufen1 recruits Upf1 to specific mRNA 3′UTRs so as to elicit mRNA decay. *Cell*, **120**, 195–208.
68. Doma,M.K. and Parker,R. (2006) Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature*, **440**, 561–564.
69. Sunohara,T., Jojima,K., Yamamoto,Y., Inada,T. and Aiba,H. (2004) Nascent-peptide-mediated ribosome stalling at a stop codon induces mRNA cleavage resulting in nonstop mRNA that is recognized by tmRNA. *RNA*, **10**, 378–386.
70. Arava,Y., Wang,Y.L., Storey,J.D., Liu,C.L., Brown,P.O. and Herschlag,D. (2003) Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA*, **100**, 3889–3894.