



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

An Extended Signal Involved in Eukaryotic -1 Frameshifting Operates through Modification of the E Site tRNA

Michaël Bekaert and Jean-Pierre Rousset*
Institut de Génétique et Microbiologie
Centre National de la Recherche Scientifique-Unité
Mixte de Recherche 8621
Bâtiment 400
Université Paris-Sud
91405 Orsay Cedex
France

Summary

By using a sensitive search program based on hidden Markov models (HMM), we identified 74 viruses carrying frameshift sites among 1500 fully sequenced virus genomes. These viruses are clustered in specific families or genera. Sequence analysis of the frameshift sites identified here, along with previously characterized sites, identified a strong bias toward the two nucleotides 5' of the shifty heptamer signal. Functional analysis in the yeast *Saccharomyces cerevisiae* demonstrated that high frameshifting efficiency is correlated with the presence of a Ψ 39 modification in the tRNA present in the E site of the ribosome at the time of frameshifting. These results demonstrate that an extended signal is involved in eukaryotic frameshifting and suggest additional interactions between tRNAs and the ribosome during decoding.

Introduction

The universal rules of genetic translation have long been known. However, the present situation is more complex. Firstly, the genetic code is not universal. In several organelles and protists, all genes are decoded by a variant code in which a stop codon reads as a sense codon. Moreover, the genetic code can be locally extended by alteration of standard rules in an individually specific manner for given mRNAs. Such extensions of the genetic code are termed reprogrammed genetic decoding or recoding events. Recoding is determined by particular sequences that force the ribosome to escape standard translation. In vivo, recoding is often in competition with standard decoding and permits the synthesis of an elongated polypeptide. Only a defined proportion of the ribosomes translating a given recoded mRNA actually use reprogrammed genetic decoding. Recoding extends the possibilities of increased diversity in gene expression or regulation.

During programmed -1 frameshifting, ribosomes switch to an alternative frame at a specific shift site, and classical triplet decoding follows. Efficient -1 frameshifting necessitates specific signals on the mRNA. Basically, the model proposed by Jacks and Varmus identified two components of frameshift signals: a slippery heptamer sequence, X₂XY₂YZ (the frame of the initiator AUG is indicated by a space), and a downstream struc-

tural element, often a pseudoknot or a stem-loop (Jacks et al., 1988). The secondary structure induces a ribosomal pause at the slippery site, and the heptamer sequence allows the slippage of ribosome bound A and P site tRNAs by one nucleotide in the 5' direction. The pause increases the probability of ribosomal movement in the 5' direction, although pausing seems necessary, but not sufficient, for stimulation (Kontos et al., 2001). More recently, a new sequence element—the spacer region located between the heptamer and the secondary structure—has been shown to modulate frameshift efficiency both in prokaryotes and eukaryotes (Bekaert et al., 2003; Bertrand et al., 2002).

Most -1 frameshift events have been reported in viruses and transposons, where they serve in synthesis of replicase activities. This mode of expression allows both a very precise control of the Gag-Pol/Gag ratio and a means of incorporating the enzymes necessary for the replication cycle into the viral particle. Enhancement or reduction of the efficiency of this mechanism can influence virus viability (Dinman and Wickner, 1992; Shehu-Xhilaga et al., 2001). Very few cellular genes with -1 frameshifting are presently known: the *dnaX* gene of *Escherichia coli* (Tsuchihashi and Kornberg, 1990), the *cdd* gene of *Bacillus subtilis* (Mejlhede et al., 1999), and the *Edr* gene of mouse (Shigemoto et al., 2001). Even though the identification of novel genes regulated by -1 frameshifting constitutes one of the postgenomic challenges, there is presently no general method to do so (Namy et al., 2004). We recently began a computational approach to model frameshifting sites. The rationale was to observe more elements of the sequence in order to obtain a more precise description of characterized frameshift signals, identify pertinent characteristics, and develop appropriate algorithms. This approach previously allowed us to identify a strong bias in the spacer sequence of eukaryotic viral frameshift signals. This bias was shown to be functionally relevant for the frameshifting mechanism (Bekaert et al., 2003). However, only few frameshift signals have been functionally characterized, although several other sites are suspected to possess such signals from sequence analysis. The original goal of this study was to identify new -1 frameshift sites in viruses, so as to enrich our collection of sites in the modeling process. We first quantified frameshifting efficiency directed by 20 viruses carrying a putative or characterized frameshift site. From these results, we were able to develop a sensitive search based on HMM. This algorithm enabled us to detect -1 frameshift sites in 74 viruses among the available, fully sequenced viruses. Sequence alignment of these virus sites identified a strong composition bias just upstream of the shifty heptanucleotide site. We demonstrate that the sequence of this region actually affects frameshift efficiency in *Saccharomyces cerevisiae*. These results point to the possible involvement of pseudouridine at position 39 of the tRNA present in the E site of the host cell ribosome in modulating frameshifting efficiency. By using *PUS3* yeast mutants lacking Ψ 38/39 modifications, we show that frameshifting efficiency is modu-

*Correspondence: jean-pierre.rousset@igmors.u-psud.fr

lated by the modification status of the E site tRNA. Overall, our results propose an extended model for -1 frameshift sites.

Results and Discussion

Characterization of Viral -1 Frameshift Sites

The RECODE database resource (<http://recode.genetics.utah.edu>, Baranov et al., 2001) describes 35 viruses suspected or demonstrated to carry a frameshift site. Few -1 frameshift signals are fully documented. The genomes of these 35 viruses are entirely sequenced, and only five sites are precisely characterized, including the structure of the stimulatory pseudoknot (BWYV, HIV-1, MMTV, PEMV-1, and SRV-1). For the remaining sites, 13 have been analyzed by extensive directed mutagenesis coupled with quantification of frameshifting efficiency, but the others are only partially characterized. Most of the sites are therefore putative; i.e., they carry the typical heptamer and secondary structure but have never been proven to be functional.

Initially, we functionally characterized a larger number of viruses containing a putative frameshifting site. To explore the widest viral diversity possible (order, family, and genus), we deduced a neighbor-joining tree from the 35 viruses, based on the multiple alignment of the polymerase protein sequence (Figure 1). From this tree, we selected a subset of 20 viruses representative of the global viral diversity. To assay the frameshift competence of each putative site, we cloned the entire viral -1 frameshift region of the 20 representative viruses in a dual-reporter vector and estimated *in vivo* frameshifting efficiency in yeast (see Experimental Procedures). Frameshift sites from different eukaryotic species have been shown to function in yeast (Bekaert et al., 2003; Stahl et al., 1995). The existence of a functional frameshift signal was demonstrated for all candidates (Table 1) except the *Sugarcane yellow leaf virus* (ScYLV). It is unlikely that the low level of expression of ScYLV is due to the use of a heterologous host cell, because ten frameshift sites from other plant viruses are functional in our assay. This site might be nonfunctional or carry polymorphic variations. The -1 frameshifting frequencies varied between 8% and 31%, compatible with those previously obtained with *in vitro* or *in vivo* assays (e.g., EAV 15%–20%, den Boon et al. [1991]; BWYV 5%, Kim et al. [1999]; HCV 20%–30%, Herold and Siddell [1993]; MMTV 20%, Chamorro et al. [1992]; and RSV 4%, Marczinke et al. [1998]). Although our assay does not provide exact frameshift rates in natural host cells (except for L-A and LB-C viruses, which are natural in *S. cerevisiae*), these results strongly suggest that most of the frameshift sites identified purely by sequence analysis are in fact functional.

We then aligned the newly characterized sites with sites already identified. Strikingly, we observed an important bias not only at the slippery heptamer but also in the spacer region and just upstream of the heptamer. The upstream bias was never before observed, and its detailed analysis is presented below.

Sensitive Search of Viral Frameshift Sites

We established a HMM profile of efficient viral -1 frameshift signals with the alignment of the slippery re-

gions from the 20 viruses that we had functionally characterized (see Experimental Procedures) and used it to search the GenBank viral genome database (release 02/10/2004). 285 motifs were identified and subsequently manually inspected to eliminate false positives. We checked (1) that the first nucleotide of the heptamer is in frame with the AUG of the upstream coding region, (2) for a protein motif associated to the upstream and downstream coding regions, and (3) for the presence of a potential secondary structure downstream of the heptamer. By this procedure, we identified 74 frameshift sites in viral genomes. Most false positives exhibited no secondary structure after the shifty site and were found in the large and highly complex *herpesvirus*, *papillomavirus*, and *nucleopolyhedrovirus* genomes. We consider this assessment to be accurate, because it depends not only on *in silico* methods but also on the biological assay of the HMM learning set. It is noteworthy that this method is very efficient, even though we did not take into account the stimulatory secondary structures when we defined the profile. RNA folding algorithms are time consuming and cannot be restrained to a defined window in the vicinity of the heptamer. Moreover, the theoretical evaluation of thermodynamic stability of secondary structures is not accurate for pseudoknots (Walter et al., 1994).

With the HMM profile based on only 20 sites, we were able to find all known frameshifting viruses and 39 that are new or uncharacterized (Table 2). The list of the viruses with the position of their frameshift signals is in Supplemental Table S1 available online at <http://www.molecule.org/cgi/content/full/17/1/61/DC1/>. Ten putative frameshift sites were never previously annotated and are associated to an upstream and a downstream coding region. 12 frameshifting structures were already annotated as such in the RECODE database, and eight were only annotated in the sequence field of GenBank or in relation to a publication that did not mention any evidence of frameshift. For the remaining 44 sequences, a site was suspected, but it was not precisely localized between two coding regions. For those, we were able to propose a precise position for the frameshift event and in some cases, a more accurate annotation. For example, for the *Ovine astrovirus* (ssRNA⁺, *Astroviridae* family), a putative -1 frameshift event sequence was previously reported but without data on the position of the frameshift site (Jonassen et al., 1998). By adding the newly identified virus to the initial set, we established an enhanced profile HMM of viral -1 frameshift sites (see Experimental Procedures), available as Supplemental Data.

Among 82 viral families, only seven are involved in -1 frameshifting: *Astroviridae*, *Arteriviridae*, *Coronaviridae*, *Luteoviridae*, *Retroviridae*, *Tombusviridae*, and *Totiviridae*. Within each family, only a few subfamilies/genera were capable of -1 frameshifting (see Supplemental Table S1 for details). However, in this latter case, all members of the genus submitted to HMM analyses appear capable of -1 frameshifting: they carry not only the HMM profile but also secondary structures as a canonical frameshift signal (Supplemental Table S1). For example, manual checking of the *Poleoviruses* found by using HMM successfully identifies a pseudoknot three to nine nucleotides downstream from the heptamer site.

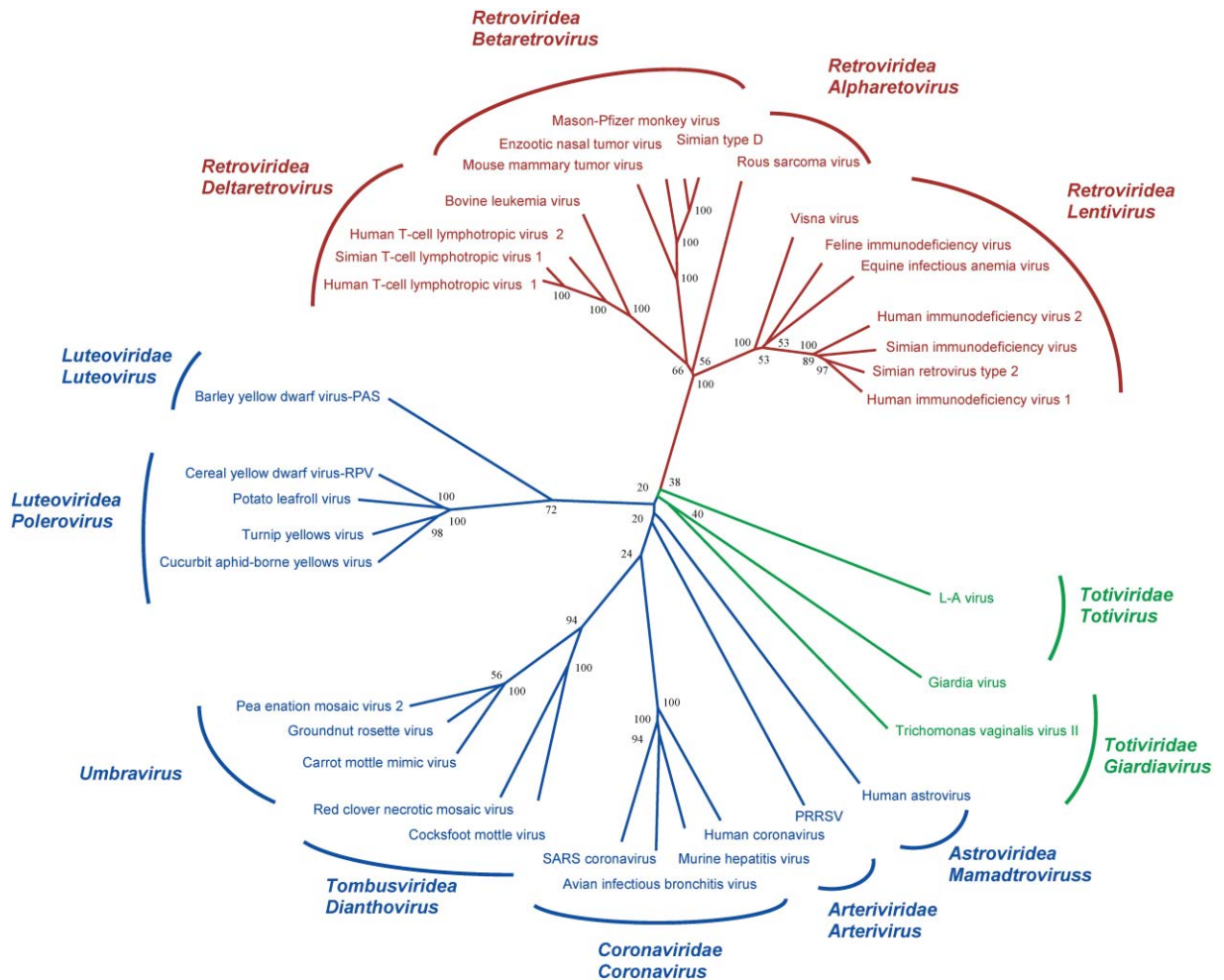


Figure 1. Unrooted Neighbor-Joining Phylogeny Based on the Amino Acid Sequence of the *Pol* Protein
Labels at branch tips represent taxa as provided by the Universal Virus Database, ICTVdB (<http://phene.cpmc.columbia.edu>; Büchen-Osmond, 2003). Numbers at branch nodes indicate the percentage of bootstrap support for that node based on 1000 replications. Color codes are according to the clade, with the following scheme: ssRNA-RT viruses (red), positive-sense ssRNA viruses (blue), and dsRNA viruses (green). Abbreviations: PRRSV, *Porcine reproductive and respiratory syndrome virus*.

In the *Totiviridae* family (dsRNA virus), only the *Totivirus* genus, represented by the L-A and LB-C viruses (two yeast viruses), has a -1 frameshift signal. Moreover, all viruses from the *Totivirus* and the *Giardiavirus* genera appear to exhibit such sites. The only exception is the *Ustilago maydis* virus H1 (*Totivirus* genus): it shows a perfectly conserved canonical slippery sequence and a strong pseudoknot, despite an in-frame *gag-pol* gene. This could be due to either a sequencing/annotation error or to an unusual configuration where canonical decoding would be responsible for the synthesis of the Gag-Pol protein, whereas the frameshift would lead to the production of only the Gag domain, reminiscent of the control of the *dnaX* gene expression in *E. coli* (Tsuchihashi and Kornberg, 1990). Another ambiguous case concerns the *Helminthosporium victoriae* virus 190S, where previous experimental investigation of gene expression concluded that translation of the second ORF is initiated on its own internal AUG codon (Huang and Ghabrial, 1996). However, this does not exclude the possibility that both mechanisms are at play to express

different polypeptides involved in polymerase activity, as observed in some bacterial transposons (Fayet et al., 1990).

Upstream Bias

As mentioned above, the alignment of the initial set of 20 viruses -1 frameshift signal sequences revealed that the base composition around the slippery sequence follows a preferential use of nucleotides. Composition bias in the spacer has been previously reported for first nucleotides (Bekaert et al., 2003; Bertrand et al., 2002); the second part of the bias has been reported in relation with the first stem composition bias (ten Dam et al., 1990). Bias in upstream sequences of the slippery regions was accurately detected in the larger scale data derived from the 74 virus sequences identified through an order 1 HMM search where the probability of a given nucleotide is dependent on the identity of the previous nucleotide. Accordingly, Figure 2 shows the bias of dinucleotide distribution. The χ^2 score for the last dinucleotide position before the heptamer is 80 with 15 degrees

Table 1. Nucleic Acid Alignment of Heptameric Sequence

Virus Acronym	FS-1		Slippery Region		Genbank
BChV	15.8% ± 2%	cacacucgcC	GGgAAAu	gGacuGaGcG	NC_002766
BLV gag/pro	8.1% ± 1%	cccucaaaUC	aaAAAAC	UaauAGaGGG	NC_001414
BWYV	12.0% ± 1%	ccaagagcUC	GGgAAAC	gGGagaGcGG	NC_004756
BYDV	12.2% ± 1%	uugacucugu	GGguuuu	UaGagGGGcu	NC_002160
CABYV	17.5% ± 1%	aaucgagUC	GGgAAAC	gGGcAGGcGG	NC_003688
EIAV	7.0% ± 1%	gaaguguucC	aaAAAAC	gGGagcaaGG	NC_001450
FIV	9.0% ± 1%	gaaagaauUC	GGgAAAC	UGGaAGGcGG	NC_001482
HIV1	6.0% ± 1%	gacaggcuaa	uuuuuuu	gGGaAGaucu	NC_001802
IBV	19.3% ± 1%	auaagaauUa	uuuAAAC	gGGuAcGGGG	NC_001451
L-A	10.0% ± 1%	guacucagca	GGguuuu	gGaguGGuaG	NC_003745
L-BC	13.0% ± 2%	cugagaagUu	GGauuuu	cGuguaGcaG	NC_001641
LDV	13.1% ± 1%	aggcaucggC	uuuAAAC	UGcuAGccac	NC_002534
MMTV gag/pro	20.2% ± 2%	cugaaaaUC	aaAAAAC	UuGuAaaGGG	NC_001503
PEMV1	31.0% ± 2%	ccagacgcUC	GGgAAAC	gGauuuucc	NC_003629
PLRV	19.0% ± 1%	caaacaagcC	GGgAAAu	gGGcAaGcGG	NC_001747
PLRV-W	17.8% ± 2%	caaacaagcC	uuuAAAu	gGGcgaGcGG	Y07496
PRRSV	15.7% ± 1%	agggagcagUg	uuuAAAC	UGcuAGccGc	NC_001961
SARS	10.3% ± 1%	caucaacgUu	uuuAAAC	gGGuuuGcGG	NC_004718
ScYLV	0.7% ± 0%	cuccagacca	GGgAAAu	gaGccaaGuG	NC_000874
SRV1 gag/pro	13.0% ± 2%	cacccaUCA	GGgAAAC	gGacuGaGGG	NC_001551
<i>Pseudoconsensus</i>		xxxxxxxUC	GGAAAAC	UGGxAGGGGG	

Nucleotides in agreement with the functional pseudoconsensus inferred from the HMM profile are in uppercase. Acronyms are as follows: BChV, *Beet chlorosis virus*; BLV, *Bovine leukemia virus* (gag/pro junction); BWYV, *Beet western yellows virus*; BYDV, *Barley yellow dwarf virus*; CABYV, *Cucurbit aphid-borne yellows virus*; EIAV, *Equine infectious anemia virus*; FIV, *Feline immunodeficiency virus*; HIV1, *Human immunodeficiency virus 1*; IBV, *Avian infectious bronchitis virus*; L-A, *Saccharomyces cerevisiae virus L-A*; L-BC, *Saccharomyces cerevisiae virus L-BC*; LDV, *Lactate dehydrogenase-elevating virus*; MMTV, *Mouse mammary tumor virus* (gag/pro junction); PEMV1, *Pea enation mosaic virus 1*; PLRV, *Potato leafroll virus*; PLRV-W, *Potato leafroll virus*, Germany strain (Wageningen); PRRSV, *Porcine reproductive and respiratory syndrome virus*; SARS, *SARS coronavirus*; ScYLV, *Sugarcane yellow leaf virus*; and SRV1, *Simian type D virus 1* (gag/pro junction).

of freedom, which makes it significant for a p value of 6.4×10^{-11} . The -4/-5 position also seems biased, but this has not been analyzed further.

To determine the role of this dinucleotide in frameshifting, we constructed dual-reporter vectors with the 16 possible sequences within the context of the wild-type (wt) frameshift signal of the *Avian infectious bronchitis virus* (IBV), because it has been extensively used as a model virus for -1 frameshifting studies (Brierley et al., 1991, 1992). Table 3 shows that a 3.3-fold variation was found between the frameshifting efficiencies directed by these 16 IBV variant sites. Compared to the wt sequence, the frameshifting level is significantly reduced (p value < 10^{-4}) in ten of the mutants.

Role of Base 39 of tRNA

The dinucleotide situated 5' of the heptamer corresponds to the first two nucleotides of the preceding codon; its impact can thus be interpreted as an effect either of the amino acid, the codon, or the decoding tRNA. Because it was previously shown that tRNA modi-

fication can affect recoding efficiency (Lecoite et al., 2002; Licznar et al., 2003), we looked for a bias in modifications of tRNA involved in decoding high and low frameshifting constructs. A correlation between the presence of pseudouridine at position 39 (Ψ 39) of the tRNA anticodon domain was observed (Table 3): all constructs that exhibited a high-frameshifting level use a cognate (or near-cognate) tRNA carrying the Ψ 39 modification. Conversely, the sequences that do not involve a codon decoded by a tRNA with the Ψ 39 modification direct low-frameshifting efficiency. This observation prompted us to investigate the effect of the mutation of

Table 2. Detailed Distributions of the HMM Profile Results from GenBank Compared to the Initial Data Set (RECODE database)

Status	GenBank (1500 virus)	RECODE (35 viruses)
New viruses	10	-
New annotations	32	15
Frameshift localized	12	8
Already annotated	20	12
	74	35

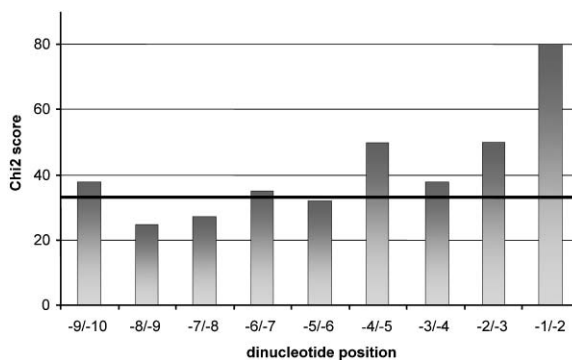


Figure 2. Estimation of Dinucleotide Usage Bias

χ^2 values were calculated at each position upstream of the slippery heptamer. Black line represents the threshold score for a p value of 1% with 15 degrees of freedom. Higher values are more indicative of a significant bias. The first dinucleotide is strongly significant (p value = 6.4×10^{-11}).

Table 3. Effect of the Upstream Dinucleotide on -1 Frameshifting Efficiency

Plasmids	Modified Sequence	tRNA (Anticodon loop)	Frameshift
pAC.5.AA	aau <u>AAu</u> uua aac	ACU GUU t ⁶ AA Ψ	18.0% \pm 2%
pAC.5.AC	aau <u>ACu</u> uua aac	Am ⁵ CU IGU t ⁶ AA Ψ	22.0% \pm 1%
pAC.5.UA	aau <u>UAu</u> uua aac	ACU GΨA i ⁶ AA Ψ	19.3% \pm 1%
pAC.5.UC	aau <u>UCu</u> uua aac	A Ψ U IGA i ⁶ AA Ψ	22.1% \pm 2%
pAC.5.UG	aau <u>UGu</u> uua aac	A Ψ U GCA i ⁶ AA Ψ	19.0% \pm 2%
pAC.5.UU	aau <u>UUu</u> uua aac	ACmU GmAA YA Ψ	21.0% \pm 1%
pAC.5.AG	aau <u>AGu</u> uua aac	CCU GCU AAG	9.3% \pm 1%
pAC.5.AU	aau <u>AUu</u> uua aac	GCU IAU t ⁶ AAC	7.0% \pm 1%
pAC.5.CA	aau <u>CAu</u> uua aac	G Ψ U GUG m ¹ GCC	9.0% \pm 1%
pAC.5.CC	aau <u>CCu</u> uua aac	CUU AGG GUG	9.0% \pm 1%
pAC.5.CG	aau <u>CGu</u> uua aac	GCU ICG AAC	7.5% \pm 1%
pAC.5.CU	aau <u>CUu</u> uua aac	GUC GAG GUC	10.0% \pm 2%
pAC.5.GA	aau <u>GAu</u> uua aac	C Ψ U GUC m ¹ GCG	9.4% \pm 1%
pAC.5.GC	aau <u>GCu</u> uua aac	CUU IGC m ¹ Ψ G	8.0% \pm 1%
pAC.5.GG	aau <u>GGu</u> uua aac	G Ψ U GCC A Ψ C	8.5% \pm 1%
pAC.5.GU	aau <u>GUu</u> uua aac	C Ψ U IAC ACG	6.7% \pm 1%

The FY strain was transformed with one of the plasmids harboring the test sequence as indicated (from 5' to 3'). Frameshifting efficiencies were measured at 30°C, and the data are expressed as percentages. Codons including the dinucleotides are underlined and anticodon of tRNA anticodon loops are in bold (Lecoite, 2002); heptamers are in bold and dinucleotides in uppercase.

the *PUS3* gene, whose product is specifically responsible for the Ψ 39 modification (Lecoite et al., 1998). If Ψ 39 is actually involved, inactivation of *PUS3* should result in a lower frameshift efficiency.

Two low- and two high-frameshift rate constructs were tested in modification mutants (Table 4). With the low-frameshifting rate subset (frameshift efficiency lower than 10%), *pus3* Δ mutants show no significant effect (Table 4). In contrast, with the high-frameshifting rate subset (frameshift efficiency higher than 18%), which involves decoding by a Ψ 39 modified tRNA, a reduced frameshifting frequency was observed in *pus3* Δ mutants. This frequency was similar to that directed by the low-frameshifting rate subset, indicating that most of the effect was reversed in the mutant. We verified that the effect is actually due to the modifying activity of Pus3p and not to a possible chaperone-like activity by using the *pus3*[D151A] mutant, which harbors a mutation in the active site of the PUS3 protein. In this mutant, the high-frameshifting constructs yield lower frameshifting efficiency, as in a *pus3* Δ mutant context (Table 4).

The effect of the dinucleotide upstream of the heptamer suggests that the three ribosomal site tRNAs are involved in the mechanism of -1 frameshifting. However, although the mechanism of frameshifting in eukaryotes is thought to involve mostly tandem slippage of the tRNAs occupying the A and P sites, single slippage at the P site has been reported to occur (Jacks et al.,

1988). If this is the case in the experimental system used here, there is no tRNA in the A site at the time of slippage (Baranov et al., 2004; Leger et al., 2004). To test the occurrence of single slippage, we used a mutant site in which the UUUAAAC heptamer was mutated to UUU AUAC. In this case, tandem slippage should be inefficient due to the presence of two mismatches after re-pairing of the A site tRNA in the -1 frame, but single slippage would not be affected. The frameshifting efficiency obtained with this construct was <0.1%, similar to the background level. This result demonstrates that in these experiments, frameshifting actually occurred through a tandem tRNA slippage mechanism. This implies that the three sites are involved in ribosomal frameshifting (see below).

The Ψ 39 modification is conserved over the tree of life; its role on -1 frameshifting could thus be similar in a broad spectrum of organisms. This is consistent with the fact that the bias at the two positions upstream of the heptamer was deduced from a wide variety of viruses of different origins. However, each host cell, like the yeast strains used here, carries a specific tRNA pool that differs from one organism to another. This could explain the different dinucleotide usage observed between viruses; however, not enough sequence data are available to assess this point. In any case, the existence of a bias indicates an important role of tRNA modification on -1 frameshifting in eukaryotes. A role of tRNA

Table 4. Effect of *PUS3* Gene Deletion on -1 Frameshifting Efficiency

Plasmids	Wt	<i>pus3</i> Δ	<i>pus3</i> Δ + pRS315	<i>pus3</i> Δ + <i>PUS3</i>	<i>pus3</i> Δ + <i>pus3</i> [D151A]
pAC.5.CG	5.3% \pm 1%	5.9% \pm 1% (1.1)	5.8% \pm 0% (1.1)	5.5% \pm 1% (1.0)	6.0% \pm 1% (1.1)
pAC.5.GA	7.6% \pm 1%	8.0% \pm 1% (1.1)	8.1% \pm 1% (1.1)	7.2% \pm 0% (0.9)	7.3% \pm 1% (1.0)
pAC.5.UA	21.7% \pm 1%	12.8% \pm 1% (0.6)	11.5% \pm 1% (0.5)	19.3% \pm 1% (0.9)	11.8% \pm 1% (0.5)
pAC.5.UC	19.5% \pm 1%	10.3% \pm 1% (0.5)	10.2% \pm 1% (0.5)	18.7% \pm 2% (1.0)	12.7% \pm 1% (0.7)

Wild-type (wt) and *pus3* Δ mutants of 74-D694 strains were transformed with the test plasmids. The 74-D694 *pus3* Δ strain was also transformed with empty pRS315 or the same plasmid containing the *PUS3* gene or the mutant *pus3*[D151A] gene, as indicated. -1 frameshifting efficiencies were measured at 30°C, and the data are expressed as percentages. Numbers in parentheses correspond to ratios of recoding efficiency in the wt strain over the recoding efficiency in the *pus3* Δ derivative strain. No significant difference can be expected by a Mann-Whitney statistical test, except between pAC.5.UA/UC in wt or *pus3* Δ + *PUS3* compared to other transformed strains (p value < 0.005).

modifications on +1 frameshifting has been previously described both in *E. coli* and in *S. cerevisiae* (Bjork et al., 1989; Lecointe et al., 2002; Urbonavicius et al., 2001). For -1 frameshifting, a few examples have been reported in *E. coli* (Brierley et al., 1997; Licznar et al., 2003), but not in eukaryotes. In these cases, the tRNAs involved were acting at the A or P site.

Overall, these results demonstrate that the effect of the upstream context of the heptamer is directed by the modification status of the tRNA decoding the -1 codon.

Conclusions

Viral Frameshifting Signals

It is striking that all members of a genus (or family, in some cases) use a frameshifting event to produce their Pol protein but that phylogenetic analyses of frameshift sequences give rise to patterns inconsistent with accepted trees (data not shown). Inconsistency of frameshifting patterns with accepted phylogenetic trees is not surprising taking into account the recombinant nature of many viruses; functional requirements probably account for both this complete conservation and the variability of the frameshifting site sequences. Indeed, in the *Retroviridae* family, the *Alpharetrovirus* genus is exceptional because some members exhibit frameshift signals but others do not. In fact, this genus is subdivided in two categories: replication-competent viruses, which possess the *pol* gene, and defective viruses, which do not. Logically, frameshift signals are found only in the latter category. It is even more interesting that despite their position among the *Totiviridae*, the *Leishmanivirus* genus members do not carry -1 frameshift sites but, rather, use +1 frameshifting to express their polymerase domain. This suggests that strong biological constraints are at play in the selection of a recoding event in the life cycle of these viruses, possibly related to the incorporation of the polymerase as a fusion protein in the viral particle.

Role of E Site in Frameshifting

An interesting feature of the results presented here is the involvement of an extended nonanucleotide signal in ribosomal frameshifting. As demonstrated above, no single slippage is observed in the experimental system used here; this nonanucleotide-directed frameshifting thus involves classical tandem slippage where both A and P site tRNAs slip by one nucleotide upstream. This implies that the three ribosomal sites are involved in -1 frameshifting. Two hypotheses can be proposed to account for the role of the E site tRNA in frameshifting. Firstly, frameshifting might be enhanced by the absence of a tRNA in the E site. In this case, the Ψ 39 modification would destabilize the tRNA:E site interaction. Secondly, Ψ 39 might interfere directly or indirectly with the interaction of the P site tRNA with the mRNA, decreasing pairing stability.

The first hypothesis is supported by recent results in which premature release of the E site tRNA from the ribosome has been shown to be coupled with high-level +1 frameshifting at the *prfB* gene, encoding the prokaryotic termination factor RF2 (Marquez et al., 2004). Likewise, in eukaryotes, Ψ 39 may induce an unusual E site conformation. If this is the case, one would predict that the Ψ 39 modification induces a higher fre-

quency of release of the tRNA from the E site. If E-tRNA normally helps prevent tRNA slippage in the P site, this could explain the different susceptibilities of a given heptamer to slippage. Probably the E-tRNA is released during the accommodation step of the A-tRNA and not during the preceding decoding reaction (Nierhaus, 1990; Noller et al., 2002). However, in the case of a -1 frameshift event, E-tRNA release at the decoding step would facilitate the slippery event of A and P site tRNAs, and this precisely might be the effect of Ψ 39. Biochemical experiments will be required to clarify this point.

The second hypothesis is supported by structural data on the prokaryotic ribosome (Ramakrishnan and Moore, 2001; Yusupov et al., 2001) and inferred cryo-EM reconstruction of the yeast ribosome (Spahn et al., 2001) that strongly suggest that the E-tRNA interacts with several partners. The closest distance between the anticodon stem backbones of the P- and E-tRNAs is about 6 Å, which is closer than the distance separating the A- and P-tRNAs. The two tRNAs are not in direct contact but are linked by the 16S rRNA helices H24, H28, and H29, and loops 690 and 790, both of which they directly interact with through their anticodon loops (Yusupov et al., 2001). Another link between E and P sites is through the mRNA. A single possible contact was noted between the mRNA and E-tRNA in the crystal structure, but the latter was noncognate. Even this noncognate E site anticodon was close enough to the codon, such that cognate interaction would be structurally plausible; moreover, there is biochemical evidence for codon-anticodon specificity in the E site (Lill and Wintermeyer, 1987; Rheinberger et al., 1986). E site tRNA is thus sufficiently connected to the P site to suggest that it very likely plays a role in promoting the stability of P site codon-anticodon pairing. Ψ 39 modification can be expected to improperly fill the E site during the slippage-prone state, probably resulting in an unstable P site codon-anticodon interaction and enhanced -1 frameshifting. This is reminiscent of the role played by a particular context of a bacterial tmRNA resume codon. In this case, an unusual E site conformation destabilizes the P site codon-anticodon interaction and induces frameshifting (Trimble et al., 2004).

The results presented here demonstrate that the slippery component of -1 frameshift signals, at least in yeast, is more complex than previously anticipated. Compared to the initial model of Jacks et al. (1988), sequence elements of both the 3' and 5' heptamer elements are now shown to participate in frameshifting efficiency through interactions between tRNA, mRNA, and the ribosome. Similarly, downstream secondary structures can directly or indirectly influence frameshifting. A combinatorial use of upstream codons, heptamer sequences, downstream codons, and stimulatory secondary structures permit a given frameshifting efficiency for a given virus in a given host. Whether or not these different sequence elements act independently remains to be established.

Experimental Procedures

Polymerase Tree

A ClustalW 1.83 (Thompson et al., 1994) alignment of viral polymerase amino acid sequences retrieved from GenBank was used. It was

employed to deduce a neighbor-joining tree with 1000 bootstrap replications (Saitou and Nei, 1987) by using Mega 2.1 package (Kumar et al., 2001), which provides a graphical representation of selected viruses. Pairwise distances were calculated as mean observed substitutions per site. The unrooted tree is shown in Figure 1 and is color coded to mark each clade.

Profile Construction

Frameshift sites—the heptamers surrounded by ten nucleotides on both sides—from the selected viruses were used to construct and calibrate an HMM by using the HMMER package 2.3.2 (Eddy, 1998). Each sequence was aligned on the shifty heptamer and the HMM established. Sequences from viruses not selected as representative subset but reported as frameshifting viruses were used to validate our profile. All sites were found (data not shown).

Searches with Profile

With this HMM profile, searches against the publicly available viral genome database (GenBank, downloaded February 10th, 2004) were carried out. All searches against nucleotide databases were performed with the HMMER 2.3.2 package. As threshold, we assigned a minimal *e* value of 0.5. Subsequently, an enhanced HMM profile was established with the HMMER package 2.3.2. The frameshift sites found—the heptamers surrounded by ten nucleotides on either side—were used to construct and calibrate the enhanced HMM profile.

Bias

Dinucleotide biases were estimated by counting each dinucleotide of the 74 sequences and comparing the distribution with an equiprobability model. Because we compared different viruses from different hosts, we should use the lesser bias model where the frequency of each dinucleotide is 1/16. This estimation used a χ^2 probability with 15 degrees of freedom.

Yeast Strains and Media

The *S. cerevisiae* strains used were FY1679-18B (Mat α *his3* Δ 200, *trp1* Δ 63, *ura3*-52, and *leu2* Δ 1), 74-D694 (Mat α , *ade1*-14, *trp1*-289, *his3* Δ 200, *leu2*-3, 112, and *ura3*-52), and its derivative *pus3* Δ (Mat α , *ade1*-14, *trp1*-289, *his3* Δ 200, *leu2*-3, 112, *ura3*-52, and *pus3* Δ ::*KAN*). Strains were grown in minimal media (0.67% yeast nitrogen base, 2% glucose) supplemented with the appropriate amino acids to allow maintenance of the different plasmids under standard growth conditions. Yeast transformations were performed by the lithium acetate method (Ito et al., 1983).

Plasmids and Molecular Biology Methods

pAC99 derivatives were constructed by cloning the synthetic oligonucleotides of interest at the unique *MscI* site of pAC99 (Bidou et al., 2000). For viral frameshift sites, heptamer, spacer, and secondary structure surrounded by ten nucleotides on each side were inserted. For mutants of the upstream region of the heptamer, the IBV frameshift site was used and the wt sequence (UA) was changed to the 15 other possible sequences (see Supplemental Table S2). Plasmids containing the *PUS3* gene or the mutant *pus3*[D151A] gene were from Lecointe et al. (2002). All constructs were verified by sequencing the region of interest.

Quantification of -1 Frameshifting Efficiency

Luciferase and β -galactosidase activities were assayed in the same crude extract as previously described (Stahl et al., 1995). The assays were carried out at least five times by using two independent transformants grown in the same conditions. The luciferase β -galactosidase ratio obtained with test constructs was normalized to the ratio obtained with the in-frame control and expresses frameshift efficiency.

Acknowledgments

We would like to thank Dominique Fourmy, Henri Grosjean, and Olivier Namy for helpful discussions and suggestions and François Lecointe for providing us with the *pus3* Δ stains and plasmids. We thank members of the Génétique Moléculaire de la Traduction labo-

ratory and the “frameshift team” for numerous stimulating discussions. We are especially grateful to Anne-Lise Haenni for critically reading the manuscript. This work was supported by the Association pour la Recherche sur le Cancer (contract 4699).

Received: July 20, 2004

Revised: September 15, 2004

Accepted: October 26, 2004

Published: January 6, 2005

References

- Baranov, P.V., Gurchich, O.L., Fayet, O., Prere, M.F., Miller, W.A., Gesteland, R.F., Atkins, J.F., and Giddings, M.C. (2001). RECODE: a database of frameshifting, bypassing and codon redefinition utilized for gene expression. *Nucleic Acids Res.* 29, 264–267.
- Baranov, P.V., Gesteland, R.F., and Atkins, J.F. (2004). P-site tRNA is a crucial initiator of ribosomal frameshifting. *RNA* 10, 221–230.
- Bekaert, M., Bidou, L., Denise, A., Duchateau-Nguyen, G., Forest, J.P., Froidevaux, C., Hatin, I., Rousset, J.P., and Termier, M. (2003). Towards a computational model for -1 eukaryotic frameshifting sites. *Bioinformatics* 19, 327–335.
- Bertrand, C., Prere, M.F., Gesteland, R.F., Atkins, J.F., and Fayet, O. (2002). Influence of the stacking potential of the base 3' of tandem shift codons on -1 ribosomal frameshifting used for gene expression. *RNA* 8, 16–28.
- Bidou, L., Stahl, G., Hatin, I., Namy, O., Rousset, J.P., and Farabaugh, P.J. (2000). Nonsense-mediated decay mutants do not affect programmed -1 frameshifting. *RNA* 6, 952–961.
- Bjork, G.R., Wikstrom, P.M., and Bystrom, A.S. (1989). Prevention of translational frameshifting by the modified nucleoside 1-methylguanosine. *Science* 244, 986–989.
- Brierley, I., Rolley, N.J., Jenner, A.J., and Inglis, S.C. (1991). Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.* 220, 889–902.
- 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.
- Brierley, I., Meredith, M.R., Bloys, A.J., and Hagervall, T.G. (1997). Expression of a coronavirus ribosomal frameshift signal in *Escherichia coli*: influence of tRNA anticodon modification on frameshifting. *J. Mol. Biol.* 270, 360–373.
- Büchen-Osmond, C. (2003). The universal virus database ICTVdB. *Comput. Sci. Eng.* 5, 16–25.
- Chamorro, M., Parkin, N., and Varnus, H.E. (1992). An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. *Proc. Natl. Acad. Sci. USA* 89, 713–717.
- den Boon, J.A., Snijder, E.J., Chirside, E.D., de Vries, A.A., Horzinek, M.C., and Spaan, W.J. (1991). Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. *J. Virol.* 65, 2910–2920.
- Dinman, J.D., and Wickner, R.B. (1992). Ribosomal frameshifting efficiency and gag/gag-pol ratio are critical for yeast M1 double-stranded RNA virus propagation. *J. Virol.* 66, 3669–3676.
- Eddy, S.R. (1998). Profile hidden Markov models. *Bioinformatics* 14, 755–763.
- Fayet, O., Ramond, P., Polard, P., Prere, M.F., and Chandler, M. (1990). Functional similarities between retroviruses and the IS3 family of bacterial insertion sequences? *Mol. Microbiol.* 4, 1771–1777.
- Herold, J., and Siddell, S.G. (1993). An ‘elaborated’ pseudoknot is required for high frequency frameshifting during translation of HCV 229E polymerase mRNA. *Nucleic Acids Res.* 21, 5838–5842.
- Huang, S., and Ghabrial, S.A. (1996). Organization and expression of the double-stranded RNA genome of *Helminthosporium victoriae* 190S virus, a totivirus infecting a plant pathogenic filamentous fungus. *Proc. Natl. Acad. Sci. USA* 93, 12541–12546.
- Ito, H., Fukuda, Y., Murata, K., and Kimura, A. (1983). Transformation

- of intact yeast cells treated with alkali cations. *J. Bacteriol.* **153**, 163–168.
- Jacks, T., Madhani, H.D., Masiarz, F.R., and Varmus, H.E. (1988). Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* **55**, 447–458.
- Jonassen, C.M., Jonassen, T.O., and Grinde, B. (1998). A common RNA motif in the 3' end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. *J. Gen. Virol.* **79**, 715–718.
- Kim, Y.G., Su, L., Maas, S., O'Neill, A., and Rich, A. (1999). Specific mutations in a viral RNA pseudoknot drastically change ribosomal frameshifting efficiency. *Proc. Natl. Acad. Sci. USA* **96**, 14234–14239.
- Kontos, H., Naphthine, 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.
- Kumar, S., Tamura, K., Jakobsen, I.B., and Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244–1245.
- Lecoite, F. (2002) Etude d'Enzymes de Modification de Nucléotides des ARNt et Leurs Fonctions dans le Métabolisme Cellulaire Chez *Saccharomyces cerevisiae* (Orsay, France: Université Paris XI).
- Lecoite, F., Simos, G., Sauer, A., Hurt, E.C., Motorin, Y., and Grosjean, H. (1998). Characterization of yeast protein Deg1 as pseudouridine synthase (Pus3) catalyzing the formation of psi 38 and psi 39 in tRNA anticodon loop. *J. Biol. Chem.* **273**, 1316–1323.
- Lecoite, F., Namy, O., Hatin, I., Simos, G., Rousset, J.P., and Grosjean, H. (2002). Lack of pseudouridine 38/39 in the anticodon arm of yeast cytoplasmic tRNA decreases in vivo recoding efficiency. *J. Biol. Chem.* **277**, 30445–30453.
- Leger, M., Sidani, S., and Brakier-Gingras, L. (2004). A reassessment of the response of the bacterial ribosome to the frameshift stimulatory signal of the human immunodeficiency virus type 1. *RNA* **10**, 1225–1235.
- Licznar, P., Mejlhede, N., Prere, M.F., Wills, N., Gesteland, R.F., Atkins, J.F., and Fayet, O. (2003). Programmed translational –1 frameshifting on hexanucleotide motifs and the wobble properties of tRNAs. *EMBO J.* **22**, 4770–4778.
- Lill, R., and Wintermeyer, W. (1987). Destabilization of codon-anticodon interaction in the ribosomal exit site. *J. Mol. Biol.* **196**, 137–148.
- Marczinke, B., Fisher, R., Vidakovic, M., Bloys, A.J., and Brierley, I. (1998). Secondary structure and mutational analysis of the ribosomal frameshift signal of rous sarcoma virus. *J. Mol. Biol.* **284**, 205–225.
- Marquez, V., Wilson, D.N., Tate, W.P., Triana-Alonso, F., and Nierhaus, K.H. (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, J.F., and Neuhard, J. (1999). Ribosomal –1 frameshifting during decoding of *Bacillus subtilis* cdd occurs at the sequence CGA AAG. *J. Bacteriol.* **181**, 2930–2937.
- Namy, O., Rousset, J.P., Naphthine, S., and Brierley, I. (2004). Reprogrammed genetic decoding in cellular gene expression. *Mol. Cell* **13**, 157–168.
- Nierhaus, K.H. (1990). The allosteric three-site model for the ribosomal elongation cycle: features and future. *Biochemistry* **29**, 4997–5008.
- Noller, H.F., Yusupov, M.M., Yusupova, G.Z., Baucom, A., and Cate, J.H. (2002). Translocation of tRNA during protein synthesis. *FEBS Lett.* **514**, 11–16.
- Ramakrishnan, V., and Moore, P.B. (2001). Atomic structures at last: the ribosome in 2000. *Curr. Opin. Struct. Biol.* **11**, 144–154.
- Rheinberger, H.J., Sternbach, H., and Nierhaus, K.H. (1986). Codon-anticodon interaction at the ribosomal E site. *J. Biol. Chem.* **261**, 9140–9143.
- Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
- Shehu-Xhilaga, M., Crowe, S.M., and Mak, J. (2001). Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *J. Virol.* **75**, 1834–1841.
- Shigemoto, K., Brennan, J., Walls, E., Watson, C.J., Stott, D., Rigby, P.W., 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.
- Spahn, C.M., Beckmann, R., Eswar, N., Penczek, P.A., Sali, A., Blobel, G., and Frank, J. (2001). Structure of the 80S ribosome from *Saccharomyces cerevisiae*-tRNA-ribosome and subunit-subunit interactions. *Cell* **107**, 373–386.
- Stahl, G., Bidou, L., Rousset, J.P., and Cassan, M. (1995). Versatile vectors to study recoding: conservation of rules between yeast and mammalian cells. *Nucleic Acids Res.* **23**, 1557–1560.
- ten Dam, E.B., Pleij, C.W., and Bosch, L. (1990). RNA pseudoknots: translational frameshifting and readthrough on viral RNAs. *Virus Genes* **4**, 121–136.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
- Trimble, M.J., Minnicus, A., and Williams, K.P. (2004). tRNA slippage at the tmRNA resume codon. *RNA* **10**, 805–812.
- 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.
- Urbonavicius, J., Qian, Q., Durand, J.M., Hagervall, T.G., and Bjork, G.R. (2001). Improvement of reading frame maintenance is a common function for several tRNA modifications. *EMBO J.* **20**, 4863–4873.
- Walter, A.E., Turner, D.H., Kim, J., Lyttle, M.H., Muller, P., Mathews, D.H., and Zuker, M. (1994). Coaxial stacking of helices enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl. Acad. Sci. USA* **91**, 9218–9222.
- Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., and Cate, J.H. (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883–896.