Nucleic Acids Research, 1997, Vol. 25, No. 10 **2005–2011**

Analyses of frameshifting at UUU-pyrimidine sites

Reneé Schwartz and James F. Curran*

Department of Biology, Wake Forest University, PO Box 7325, Winston-Salem, NC 27109, USA

Received December 12, 1996; Revised and Accepted March 26, 1997

ABSTRACT

Others have recently shown that the UUU phenylalanine codon is highly frameshift-prone in the 3′ (rightward) direction at pyrimidine 3′ contexts. Here, several approaches are used to analyze frameshifting at such sites. The four permutations of the UUU/C (phenylalanine) and CGG/U (arginine) codon pairs were examined because they vary greatly in their expected frameshifting tendencies. Furthermore, these synonymous sites allow direct tests of the idea that codon usage can control frameshifting. Frameshifting was measured for these dicodons embedded within each of two broader contexts: the *Escherichia coli prfB* (RF2 gene) programmed frameshift site and a 'normal' message site. The principal difference between these contexts is that the programmed frameshift contains a purine-rich sequence upstream of the slippery site that can base pair with the 3′ end of 16 S rRNA (the anti-Shine–Dalgarno) to enhance frameshifting. In both contexts frameshift frequencies are highest if the slippery $tRNA^{Phe}$ is capable of stable base pairing in the shifted reading frame. This requirement is less stringent in the RF2 context, as if the Shine–Dalgarno interaction can help stabilize a quasi-stable rephased tRNA:message complex. It was previously shown that frameshifting in RF2 occurs more frequently if the codon 3′ to the slippery site is read by a rare tRNA. Consistent with that earlier work, in the RF2 context frameshifting occurs substantially more frequently if the arginine codon is CGG, which is read by a rare tRNA. In contrast, in the 'normal' context frameshifting is only slightly greater at CGG than at CGU. It is suggested that the Shine–Dalgarno-like interaction elevates frameshifting specifically during the pause prior to translation of the second codon, which makes frameshifting exquisitely sensitive to the rate of translation of that codon. In both contexts frameshifting increases in a mutant strain that fails to modify tRNA base A37, which is 3′ of the anticodon. Thus, those base modifications may limit frameshifting at UUU codons. Finally, statistical analyses show that UUU Ynn dicodons are extremely rare in *E.coli* genes that have highly biased codon usage.

INTRODUCTION

Much of what we know about frameshifting has come from studies of programmed frameshifting (reviewed in 1,2). At most of these sites frameshifting occurs by tRNA:message realignment (3–5), and frameshifting requires that the realigned complexes contain stable base pairs. In many programmed frameshifts, tRNA:message realignment may occur during ribosomal pauses (6–12). Furthermore, most programmed frameshifts have sequences near the slip site that enhance tRNA:message realignment. Unfortunately we know very little about how the features of programmed frameshift sites apply to frameshift errors during 'normal' translation. Here, we determine how sequence features shown to drive the *Escherichia coli prfB* (RF2 gene) programmed frameshift (13) affect frameshifting in a 'normal' context.

The RF2 programmed frameshift site (Fig. 1) has been extensively analyzed genetically and biochemically. The frameshift is part of an autoregulatory mechanism. The RF2 protein terminates translation at UGA and UAA codons (14). *prfB* contains a UGA near its 5′ end, and RF2 terminates synthesis of nascent RF2 polypeptide in a concentration-dependent manner (6,15). To synthesize RF2, ribosomes bypass the UGA termination codon by a rightward (+1) frameshift. The frameshift occurs by realignment of the peptidyl-tRNA from the codon immediately 5′ of the UGA onto the overlapping UUU triplet (13). The frameshift mechanism is shown in Figure 1; the four frameshift-enhancing features are numbered in the figure and described here. First, frameshifting is facilitated by slow translation of the UGA codon (6,15), which is consistent with the autoregulatory function of the frameshift. Frameshifting also occurs with codons substituted for the UGA (4,7,9), and frameshift frequency is inversely related to the rate of aa-tRNA selection at those codons (6,7,9,12). Second, the 3′ end of 16 S rRNA (the anti-Shine–Dalgarno sequence, ref. 17) base pairs with a run of purines upstream of the slip site to enhance frameshifting (18). The mechanism by which this interaction stimulates frameshifting is not known, but it is worth note that a Shine–Dalgarno-like interaction also stimulates leftward (–1) frameshifting in *E.coli dnaX* (19), and that frameshift direction and efficiency may be related to the size of the spacing between that interaction and the P site (4,7,19,20). Third, a G:U wobble base pair in the pre-shift codon:anticodon complex is associated with high-frequency frameshifting, as if this weak pair facilitates slippage from the initial frame (21). Weak pairing in the initial frame has also been shown to contribute to high frequency frameshifting at other

* To whom correspondence should be addressed. Tel:+1 910 759 5956; Fax: +1 910 759 6008; Email: curran@wfu.edu

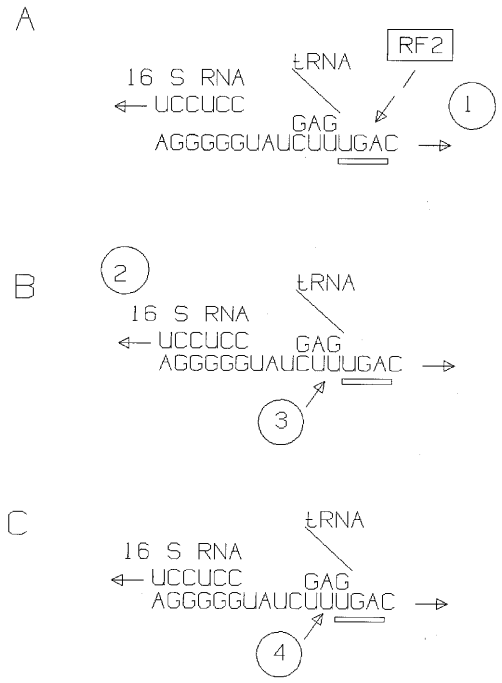

Figure 1. Model of the RF2 programmed frameshift mechanism. The mechanism is outlined illustrating the four features that contribute to high frequency frameshifting. These features are marked by circled numbers and are described below and in Introduction. (**A**) Prior to the shift, the ribosomal complex contains peptidyl-$tRNA^{Leu}$ in the P site base paired to the CUU triplet. The UGA codon (underlined) may either be translated normally, or a frameshift may occur. The first shifty feature is slow translation of the UGA (or of other codons substituted for it), which allows time for frameshifting. (**B**) The probability of a frameshift is increased if the 3′ end of 16 S RNA (the anti-Shine–Dalgarno) base pairs with the message (feature 2). Realignment is also facilitated by a G:U wobble base pair between the peptidyl-tRNA and the message (feature 3). (**C**) The fourth shifty feature is stable base pairing between the tRNA and message in the shifted frame.

programmed frameshift sites (5,22,23). Fourth, the realigned complex includes stable base pairs (4,21). This feature is also observed at virtually all other programmed frameshift sites.

Recently, Fu and Parker (24) show that a specific UUU UAC site near the 5′ end of *E.coli argI* undergoes a one nucleotide rightward (3′) frameshift with a frequency of several percent. Derivatives that preserve the UUU Ynn theme are frameshift prone, but changes of either the phenylalanine codon to UUC or the 3′ neighbor to adenosine inhibits frameshifting. Thus the sequence requirements for frameshifting at this site are at least superficially similar to those found for the extensively characterized RF2 programmed frameshift slippage site (the third and fourth features above). However, unlike the RF2 slippage site, the context surrounding the *argI* slippage site does not include any other element, such as a Shine–Dalgarno sequence, known to facilitate frameshifting at programmed frameshift sites. Furthermore, this frameshift has no apparent cellular utility: it does not generate a useful polypeptide, and it has no apparent regulatory function. It may, therefore, be a high frequency ‘error.’

To increase understanding of this error and of the RF2 frameshift, we compared the frameshifting properties of variants of the UUU Ynn theme in and out of the context of the RF2 programmed frameshift site. The variants were designed to test the effects of specific features shown to be important for frameshifting in RF2 (Fig. 1). In the RF2 context, the relative frameshift frequencies are essentially as expected from the earlier work described above. However, the stability of the rephased tRNA:message complex may be more important in the normal context than it is in RF2. In addition, the pause prior to translation of the second codon is less important in the normal context than it is in RF2. We attribute these differences to effects that the Shine–Dalgarno-like interaction has on frameshift stimulation in the RF2 context. We also find that, in both contexts, frameshifting is increased in stationary cultures if the slippery $tRNA^{Phe}$ lacks the modifications on base A37 3′ to the anticodon (the ms2io6A modifications, see ref. 25).

All organisms exhibit coding sequence biases that include the unequal usage of synonyms (reviewed in 26–28) and nonrandom codon contexts (reviewed in 29–31). There may be many factors that drive sequence bias, but we wonder whether at least some biases occur because they contribute to reading frame maintenance. Because UUU Ynn dicodons are frameshift-prone, it is conceivable that such sites are rare in genes. Statistical analyses show that such sites are indeed rare in *E.coli* genes, especially in genes that are expressed at high levels. It is at least plausible that such sites are avoided because of their inherent frameshifting tendencies.

MATERIALS AND METHODS

Strains and plasmids

Our standard *E.coli* K12 host is MY600 (32), which has the genotype: Δ*(lac-pro) ara thi*. MY599 contains a restrictive *rpsL* mutation, but is otherwise isogenic to MY600. It was made by curing S90C (33) of its F′ by acridine orange treatment as described by Miller (34). *Salmonella* strains GT522 and GT523 are *miaA*$^+$ and *miaA1*, respectively, and are otherwise prototrophic (35). These strains were generously provided by Dr Glenn Björk of the University of Umeä. Genetic manipulations of the *Salmonella* strains were performed using standard protocols (36). The culture media for β-galactosidase assays was Vogel and Bonner (37) minimal salts, supplemented with 0.5% casein hydrolysate (Difco), 0.5% glucose, and an appropriate antibiotic as described by Maniatis *et al*. (38). Assays were performed as previously described (32).

Plasmids are derivatives of pJC27 (32), pBR322 (39) and pRS20. pRS20 (Fig. 2) was made in two steps starting with pJC27 and pBR322. First, an *Eco*RI–*Sal*I fragment containing the *lacZ* gene from pJC27 was ligated to the large *Eco*RI–*Sal*I fragment from pBR322. Then, the *lacZ* promoter-ribosome binding region was replaced with the corresponding element from *ompA* on a *Sal*I–*Hin*dIII fragment. The *ompA* fragment was obtained from the *E.coli* genome as a PCR fragment in which the primers encode those restriction sites. The sequences of the primers are: 5′-GTCCGTCG-ACGATTAAACATACCTTATAC-3′ and 5′-CAGCAAGCTTTT-CATTTTTTGCGCCTCG-3′. The resulting construct fuses the second codon of *ompA* to the polylinker region of *lacZ*. Frameshift sites were then cloned as double-stranded oligonucleotides between the *Hin*dIII and *Bam*HI sites of the polylinker of pRS20 (Fig. 2). The sequence of the noncoding strand oligonucleotides used to make alleles with UUU CGG dicodons at the frameshift site: RF2 context: 5′-AGCTTCCTTAGGGGGTAT***TTTCGG***CTAGG-3′; Normal context: 5′-AGCTTAGCA***TTTCGG***TCGTAG-3′. The phe–arg dicodons are italicized and bolded. Other dicodon constructs were made with oligonucleotides having the corresponding sequence changes in those two codons.

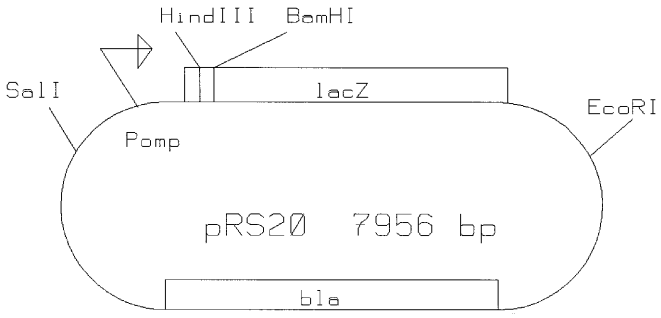

Figure 2. pRS20. pRS20 contains a fragment from the *E.coli ompA* gene including the promoter and translational start region fused to a polylinker at the 5′ end of a *lacZ* reporter. Oligonucleotides encoding frameshift sites were spliced between the *Hin*dIII and *Bam*HI sites in the polylinker. pRS20 was constructed as described in Materials and Methods.

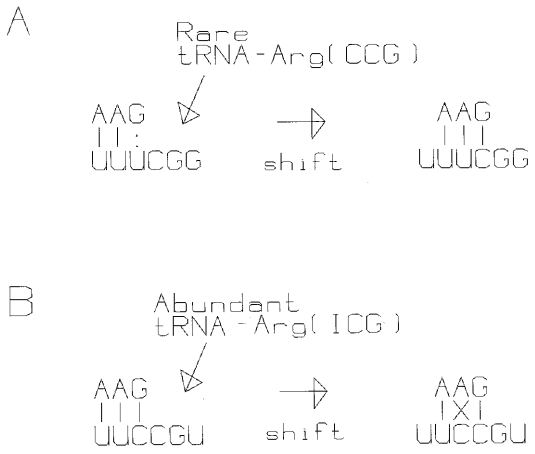

Figure 3. Comparisons of the slippery features of phe–arg dicodon sites. (**A**) and (**B**) show the pre-shift (at left) and post-shift (right) anticodon:message complexes for UUU CGG and UUC CGU, respectively. Watson–Crick base pairs are indicated by vertical lines; the G:U wobble pair is marked with a colon; and the C:A mispair is denoted with an 'x'. Codon pair UUUCGG contains shifty features 1, 3 and 4 from Figure 1: The pre-shift complex contains a G:U wobble pair, which may facilitate loss of the initial frame (feature 3); the post-shift complex is stabilized by three Watson–Crick base pairs (feature 4); and, because the tRNA that reads the CGG codon is very rare, there may be a greater pause during which realignment can occur (feature 1). In contrast, codon pair UUCCGU has none of these shifty features: The pre-shift complex includes no wobble base pairs; the post-shift complex includes a C:A mispair, which can destabilize the complex (21); and, because the CGU codon is read by an abundant tRNA, there may be little time available for tRNA:message realignment.

Sequence analyses

Computer programs were written Turbo Pascal, version 3 (Borland) and were run in a DOS environment. The highly expressed genes sequence database is derived from the ECO_H.DAT file from the EMBL TRANSTERM database (40). This file contains 106 *E.coli* coding sequences that have high CAI values. Expected numbers of codon pairs were calculated using a previously described joint probability equation (31). The weakly expressed genes sequence database is an expanded set derived from that used by Folley and Yarus (41).

RESULTS

Design of frameshift test sites

As described in the Introduction, Fu and Parker (24) show that a UUU codon is extremely frameshift-prone if it is followed by a pyrimidine-starting codon (UUU Ynn codon pairs). To better understand the mechanism(s) of this frameshift, we designed a set of four systematic variants of the UUU Ynn theme. Each member of the set was placed within two broader sequence contexts: the RF2 programmed frameshift site, and a short frameshift window that lacks programmed frameshift-enhancing features and which, therefore, may be representative of 'normal' message sites. Because the RF2 frameshift mechanism is very well-understood (see Fig. 1), comparisons of the two contexts may illuminate the frameshifting mechanism at the normal site.

The frameshift site of the RF2 programmed frameshift has been extensively characterized. For example, we have measured frameshifting for variants with 32 different triplets at the first codon (21) and 29 different triplets at the second codon (9). Three of the four features shown to be important for high frequency frameshifting (Fig. 1) are associated with these two codons. The first codon is more frameshift prone if it has a third-position uridine (feature 3 from Fig. 1), possibly because relatively weak wobble pairing with this U may facilitate slippage of the peptidyl-tRNA from the initial phase (21). In addition, efficient frameshifting requires that the tRNA that reads the first codon be capable of stable base pairing paring to the overlapping +1 triplet (feature 4 from Fig. 1; refs 4,21). Together, features 3 and 4 comprise the 'slipperiness' of the site. The other important feature is the pause prior to translation of the second codon (feature 1); the longer the pause, the greater the opportunity for tRNA:message realignment (6,7,9,12).

We chose to study the four permutations of codon pairs with UUU/C (phenylalanine) at the first position and CGG/U (arginine) at the second position because they are expected to have various combinations of the shift-site features identified from the extensive prior analyses described above. The features of the most (UUU CGG) and least (UUC CGU) frameshift-prone combinations are described in Figure 3. Briefly, codon pair UUU CGG is expected to be very slippery (features 3 and 4 from Fig. 1) and have a long pause (feature 1). In contrast, UUC CGU has none of these features. For the other two codon pairs, UUU CGU has only the slippery features (features 3 and 4), and UUC CGG is expected to have only a longer pause (feature 1). It should be noted that ribosomal pauses prior to aa-tRNA selection at CGG and CGU were not directly measured, but are assumed to depend inversely on the concentrations of the tRNAs that read them (Fig. 3).

Frameshift frequencies of the phenylalanine–arginine codon pairs

The four codon pairs were placed at the RF2 programmed frameshift site in RF2:*lacZ* fusions (Fig. 2). The β-galactosidase activities and percent frameshift frequencies are listed in Table 1 (the alleles with the RF2 prefix). The observed frameshift frequencies are completely consistent with the predictions. Dicodon UUU CGG, which has all of the shift-prone features, is ~18-fold more frameshift-prone than UUC CGU, which has none of the features. In addition, the two alleles that have only some of

the shift-prone features (UUU CGU and UUC CGG) have intermediate frameshift frequencies. And, because those intermediate values are essentially equivalent, the reduction in slipperiness by the UUU to UUC mutation, and the reduction in the pause by the CGG to CGU mutation, have equal effects RF2asy1on the RF2 programmed frameshift.

Table 1. Frameshifting frequencies of the UUU CGG/U constructs

Strain	Frameshift allele	β-gal activity	% Frameshifting
MY600 (pRS20)	WT	10 530	–
MY600(pRS31)	RF2/UUU CGG	555	5.3
MY600(pRS32)	RF2/UUU CGU	123	1.2
MY600(pRS33)	RF2/UUC CGG	131	1.2
MY600(pRS34)	RF2/UUC CGU	31	0.3
MY600(pRS21)	NSD/UUU CGG	129	1.2
MY600(pRS22)	NSD/UUU CGU	95	0.9
MY600(pRS23)	NSD/UUC CGG	3.5	0.03
MY600(pRS24)	NSD/UUC CGU	3.2	0.03

β-gal values are averages of from between four and six assays; standard errors of the mean are <10%. Percent frameshifting values are relative to the WT control.

We also assayed the same codon pairs within very short frameshift windows that lack any Shine–Dalgarno-like element (Fig. 2). Here, even without the Shine–Dalgarno the alleles with UUU at the first codon frameshift at frequencies that are ~2 orders of magnitude higher than the average frameshift frequency (42,43). In contrast, neither allele with UUC frameshifts significantly. These data are in general agreement with the earlier observations of Fu and Parker (24), who found that UUU Ynn, but not UUC Ynn, codon pairs were frameshift-prone. Thus, a slippery UUU at the first codon appears to be essential for frameshifting in 'normal' contexts. However, in contrast to the RF2 context, whether or not the second codon is read by a rare tRNA is relatively unimportant in the normal context (compare NSD/UUU CGG and NSD/UUU CGU). This suggests that in the normal context the pause prior to aa-tRNA selection at the second codon may be less important than it is at the RF2 programmed frameshift site.

Frameshifting with altered translational apparatus

At some sites frameshift frequencies can be affected by mutations and other changes in the translational apparatus. We wished to determine whether such changes would differentially affect the RF2 and NSD alleles. Fu and Parker (24) noticed that as cultures enter the stationary phase frameshifting increases at the UUU UAC site at the 5′ end of *argI*. Gallant and coworkers (44) also observe that frameshift frequencies can increase in stationary phase cultures. The changes in cell physiology that occur during stationary phase are not fully defined, and perhaps several changes could affect ribosomal kinetics and accuracy.

One change that may be important is the undermodification of tRNA, which can be induced by various stressful physiological conditions (45). The slippery $tRNA^{Phe}$ has the 2-methylthio and 6-isopentenyl modifications at adenosine 37 (ms2i6A in *E.coli* and ms2io6A in *Salmonella typhimurium)*, which is immediately 3′ of the anticodon. Formation of these modifications is dependent on the activity of the *miaA* gene (46). Absence of the *miaA* modifications decreases translational efficiency, probably because the modifications stabilize the tRNA:message complex (reviewed in ref. 25). Therefore, it is plausible that $tRNA^{Phe}$ might be more likely to slip on the message if it lacks the *miaA* modifications. To determine whether modification of base 37 and/or stationary growth differentially affects frameshifting by the RF2 and NSD alleles, we compared frameshift frequencies in isogenic *miaA*+/– *S.typhimurium* strains in both midlog and stationary phase (overnight) cultures. The *S.typhimurium* stains were the kind gifts of Dr Glenn Björk.

We measured the steady state β-galactosidase activities of the RF2/UUU CGG and NSD/UUU CGG alleles during expression in isogenic *miaA*+/– strains in both the midlog and the stationary culture conditions (Table 2). Assays made from log phase cultures at various densities $\leq 1.6 \times 10^9$ cells/ml are similar to within ~10% of each other, and the averages of eight such assays are reported in Table 2. Similarly, assays made from overnight cultures (densities $\geq 4 \times 10^9$ cells/ml) are also consistent, and the averages of those assays are also included in the table. The *miaA* mutation increases frameshifting of the RF2-based construct by nearly 2-fold (1.8- and 1.5-fold for low and high cell density, respectively). The NSD allele shows slightly greater increases (2.3- and 2.5-fold, for low and high cell density, respectively). These increases suggest that the *miaA* modifications reduce the inherent tendency of this tRNA to frameshift. Furthermore, the modifications may be slightly more important in the NSD context (the standard errors of these ratios are ~15%). We address the effect of the *miaA1* mutation on frameshifting in the Discussion.

Table 2. Frameshifting increases in an miaA strain at low and high cell densities

Strain	Frameshift allele	*miaA* allele	Cell density	
			Low	High
GT522(pRS31)	RF2/UUU CGG	*miaA*+	4.5	14.2
GT523(pRS31)	RF2/UUU CGG	*miaA1*	8.0	21.4
GT522(pRS21)	NSD/UUU CGG	*miaA*+	1.25	4.8
GT523(pRS21)	NSD/UUU CGG	*miaA1*	2.9	11.8

Reported are % frameshifting relative to the corresponding *Salmonella* host expressing WT lacZ from pRS20. Low cell density refers to cultures with viable cell concentrations between 2×10^8 and 1.6×10^9 cells/ml. High cell density refers to cultures with viable cell concentrations between 4×10^9 and 6.5×10^9 cells/ml.

All of the stationary cultures show increased frameshifting. For the RF2-based allele, frameshifting increases ~3-fold (2.8-fold for miaA+ and 3.2-fold for miaA1). The NSD alleles is affected ~4-fold (3.8-fold for miaA+ and 4.1-fold for miaA1). Thus, with borderline significance the NSD allele appears to be slightly more sensitive to the stationary culture condition. Finally, because the stationary culture condition increases frameshifting equally for the *miaA*+ and *miaA1* strains, the stationary-induced increase is not simply due to an inefficient *miaA* modification. We have not further explored the molecular and physiological bases for stationary-induced frameshifting.

We also studied the effect of a restrictive *rpsL* (streptomycin resistance) mutation on frameshifting at UUU Ynn. Restrictive *rpsL* mutations inhibit or 'restrict' aminoacyl-tRNA selection

(47–50). Sipley and Goldman (12) and we (51) have shown that restrictive *rpsL* mutations can increase frameshifting for certain RF2/*lacZ* fusions, presumably because restricted aa-tRNA selection at the second codon increases the time available for tRNA:message realignment. We also observed that the response to *rpsL* can differ among RF2 alleles, and *rpsL* sensitivity correlates with weak codon:anticodon pairing at the second codon (51). To determine whether a restrictive *rpsL* mutation affects frameshifting at UUU Ynn, we assayed constructs from Table 1 in MY599. None of the constructs showed significant increases in frameshift frequency in the *rpsL* background (not shown), which suggests that instability of the tRNA:message complexes at the *second* codon is not a factor in frameshifting at these sites. In any case, *rpsL* does not distinguish the RF2 and NSD alleles.

UUU Ynn dicodons are strongly avoided

It is firmly established that highly expressed genes show strong sequence biases, and sequence bias is thought to contribute to high levels of expression (26–31). Therefore, if UUU Ynn codon pairs are generally frameshift-prone, then such sites should be underrepresented in genes that show strong sequence biases. We searched for the occurrences of UUU Ynn codon pairs in *E.coli* genes that have high degrees of codon usage bias (40). Assuming that codons are randomly distributed, then 122 UUU Ynn codon pairs are expected; however, only 25 such sites occur (see Table 3). Assuming a binomial probability, a difference from expectation this great has a chance likelihood of only 5×10^{-18}. We also searched for these sites in a collection of weakly expressed genes (41). The UUU Ynn codon pairs are also underrepresented in these genes, but the difference from expectation is of only borderline significance. Interestingly, for both highly and weakly expressed genes, the 'missing' UUU Ynn sites are almost exactly matched by 'extra' UUC Ynn sites (97 fewer UUU Ynn and 92 more UUC Ynn in the 'Hi' set; and 31 fewer UUU Ynn and 33 more UUC Ynn in the 'Lo' set). These observations are consistent with there being random distributions of phenylalanine-Ynn dicodons, accompanied by selection for UUU→UUC mutations at those sites. Further, because this apparent selection is much greater for highly biased genes, this bias may contribute to high levels of expression.

Table 3. Statistical analyses of the occurrences of UUUY tetranucleotides

Sites	Number of sites		Binomial probability
	Expected	Observed	
Hi-UUU Ynn	122	25	5×10^{-18}
Hi-UUC YNN	332	424	1.5×10^{-6}
Lo-UUU Ynn	169	138	2×10^{-2}
Lo-UUC YNN	105	138	2×10^{-3}
Hi-vUU UYn	143	146	0.4
Hi-nnU UUC and Hi-nnU UUU Rnn	303	240	5×10^{-4}

The prefix 'Hi-'refers to genes with high levels of sequence bias; 'Lo-' means weakly expressed genes.

A formal alternative, however, is that the scarcity of UUU Ynn codon pairs in highly biased genes is due to some physical property of UUUY tetranucleotides (or the corresponding DNA sequence) that is unrelated to protein synthesis. In this were true, then UUUY tetranucleotides should be rare in all three reading frames. We tested this by searching the highly biased gene set for sites that include the UUUY tetranucleotides phased in the alternate reading frames (e.g., the dicodon gUU UCa includes a UUUY tetranucleotide, but it is phased one nucleotide rightward relative to the translated frame). One complication inherent to studies nucleotide runs, like UUUY, is that certain codon pairs can include them in more than one reading frame. For example, dicodons UUU UYn include UUUY tetranucleotides in both the initial and the rightward phases. Because above we show that UUUY tetranucleotides are rare in the initial frame (above), we were careful to exclude sequences that include UUUY in the initial frame from these control analyses. For example, to study the rightward phase we examined codon pairs of the type AUU UYN plus GUU UYN plus CUU UYN (collectively noted as vUU UYn).

In the rightward phase, vUU UYn dicodons, occur almost exactly as expected for randomly distributed codons (Table 3). These data show that the UUUY tetranucleotides are not universally excluded from coding sequences, as would be expected for sequences that have some adverse structural anomaly. Therefore, the extreme scarcity of UUU Ynn codon pairs may be related to a translational property(s) such as frameshifting.

Interestingly, the UUUY tetranucleotides are relatively rare in the leftward reading frame (nnU UUC + nnU UUU Rnn codon combinations, Table 3), though the effect is not nearly as strong as it is for the in-phase UUU Ynn dicodon set. Gallant and co-workers have shown that nnU UUC dicodons can be shift-prone into the leftward frame, but in those experiments frameshifting requires starvation for the amino acid specified by the codon followed the phenylalanine codon (52) or the stationary culture condition (44). Leftward frameshifting also occurs at such sites in several programmed frameshifts (1,2). Clearly, leftward frameshifting can occur at phenylalanine codons at least under certain growth conditions and/or contexts. Thus, the modest avoidance of the leftward UUUY tetranucleotide might also be related to an inherent shiftiness.

DISCUSSION

We know a great deal about the sequence features that contribute to frameshifting at the *prfB* (RF2) programmed frameshift site (Fig. 1), but we know very little about how those features related to frameshift errors during normal translation. It was previously shown that a naturally occurring UUU UAU codon pair and derivatives that retain a UUU Ynn theme are ~100 times more frameshift-prone than the average message site (24). To help understand the mechanism of this apparently errant translation, we analyzed the frameshifting tendencies of variants of the UUU Ynn theme placed within two broader contexts: the RF2 programmed frameshift site, which contains a frameshift-enhancing element (the Shine–Dalgarno-like element, ref. 23, see Fig. 1), and a site that lacks frameshift-enhancing features. Extensive preliminary analyses of the RF2 frameshift enabled the design of a set of four codon pairs that contain various combinations of frameshift-prone features. Our data show that predictions about

frameshift tendency in RF2 are partially applicable to a non-programmed slippery site. The disagreements between the RF2 and the normal contexts may result from the effects of the Shine–Dalgarno-like interaction, which occurs only for the RF2-based alleles.

In both contexts, high frequency frameshifting requires that the rephased tRNA:message complex be stable, but the partially-mismatched complex at UUC CGG is relatively frameshift-prone only in the RF2 context (Table 1). Together with previous work, these data suggest that the Shine–Dalgarno may be able to help hold the shifted phase following tRNA:message realignment. We previously show that many RF2 alleles frameshift with considerable frequencies despite forming moderately mismatched complexes in the rightward frame (21). The intermediate frameshift frequency for RF2-UUC CGG (Table 1), which has a single mismatch in the rightward frame (an A:C in the middle position), is consistent with that earlier work. In contrast, the NSD alleles that have UUC as the phenylalanine codon essentially do not frameshift (Table 1). The simplest explanation is that those alleles, which can form neither a perfectly realigned tRNA:message complex nor a Shine–Dalgarno interaction, cannot stabilize the rephased complex.

Another effect of the Shine–Dalgarno-like interaction may be to highly elevate frameshifting only when the second codon is available for translation in the A site (i.e., after translocation of peptidyl-tRNA^{Phe}). An implicit requirement for RF2 autoregulation is that frameshift probability is increased only while the regulatory UGA is available for recognition by RF2 (6,7,13). Most probably, translocation of the message simultaneously places the regulatory codon into the A site and the stimulatory AGGGG to within reach of the ribosomal anti-Shine–Dalgarno. This model makes the programmed frameshift totally dependent on the kinetic competition between Shine–Dalgarno association and translation of the A site codon (7). Another implication of this model is that because only post-translocation frameshifting is stimulated by the Shine–Dalgarno, other mechanisms do not contribute significantly to observed β-galactosidase activity.

In contrast, the frameshifting observed for alleles that lack the Shine–Dalgarno could conceivably result from rephasing before, during and/or after translocation. Because the second codon has only a minor effect on frameshifting (compare UUU CGG and UUU CGU in Table 1), most frameshifting on the NSD alleles may occur before the second codon becomes available for decoding. In addition, that the normal sites is also slightly more sensitive to the *miaA1* mutation and to the stationary phase also supports the notion that frameshifting at the normal site differs mechanistically from that of RF2. Plausible other mechanisms include selecting phe-tRNA^{Phe} directly into the rightward phase, and/or slippage of the tRNA:message complex prior to or during translocation. Unfortunately, it is not yet possible to distinguish these mechanisms.

The lack of the *miaA*-dependent modifications at A37 increases frameshifting for both the RF2 and NSD alleles. Others have observed that the absence of a different modification at position 37 (the lack of the 1-methyl modification at G37) of one or more proline tRNAs is associated with rightward frameshifting (53,54). There, increased frameshifting is consistent with either increased slippage because of reduced tRNA:message stability and/or out-of-phase pairing between the unmodified G37 and the message in the rightward phase (54,55). There is abundant evidence that the absence of the *miaA* modifications decreases tRNA:message stability. For example, in a model system in which tRNAs are paired through complementary anticodons (56), complexes are less stable if they include a tRNA that is missing its *miaA*-dependent modifications (57). In addition, unmodified tRNAs are less efficient during nonsense suppression and missense reading *in vivo* (58–62). It was suggested that the mutation makes tRNA:message complexes less stable because the unmodified adenosine does not stack strongly onto the codon:anticodon complex, and that this relatively weak interaction increases the likelihood that the suppressor will be rejected during ribosomal selection (60). This suggestion is supported by the observation that *miaA* unmodified tRNAs are aggressively proofread and are poor competitors with modified tRNAs during *in vitro* polypeptide synthesis (63). It seems very likely that an *miaA* induced increased in frameshifting is caused, at least in part, by an instability in the pre-shift complex which leads to an increased probability of tRNA:message slippage.

The reasons for biased codon and context usage are not yet known (26–31), but they are very likely to include selection against sites that are either inefficiently translated and/or error-prone. Our statistical analyses show that UUU Ynn dicodons are extremely rare in the translated phase of *E.coli* genes that exhibit strong codon usage biases (Table 3). This apparent avoidance of UUU Ynn codon pairs is very likely due to one or more translational properties; it seems at least plausible that a strong, inherent frameshifting tendency could be such a property. We previously showed that the CGA codon, which is decoded with an inosine:adenosine wobble base pair, is both extremely rare and inefficiently decoded (51). It seems very likely that the avoidance of CGA codons and the scarcity of UUU Ynn codon/contexts are driven, at least in part, by the poor translational qualities of these sequences.

ACKNOWLEDGEMENTS

We are grateful to Ms Jennifer Miller for technical assistance and to Dr Glenn Björk for the kind gift of *Salmonella* strains. We are also grateful to Drs Jonathan Gallant and Valery Lim for unpublished data and for helpful comments on the manuscript. This work was supported by NIH grant GM52643 to JFC.

REFERENCES

1 Farabaugh, P.J. (1996) *Micro. Rev.* **60**, 103–134.
2 Gesteland, R.F., and Atkins, J.F. (1996) *Ann. Rev. Biochem.* **65**, 741–768.
3 Jacks, T., Madhani, H.D., Masiarz, F.R., and Varmus, H.E. (1988) *Cell* **55**, 447–458.
4 Weiss, R.B., Dunn, D.M., Atkins, J.F., and Gesteland, R.F. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52**, 687–693.
5 Brierley, I., Jenner, A.J., and Inglis, S.C. (1992) *J. Mol. Biol.* **227**, 463–479.
6 Craigen, W.J., and Caskey, C.T. (1986) *Nature* **322**, 273–275.
7 Curran, J.F., and Yarus, M. (1988) *J. Mol. Biol.* **203**, 75–83.
8 Belcourt, M.F., and Farabaugh, P.J. (1990) *Cell* **62**, 339–352.
9 Curran, J.F., and Yarus, M. (1989) *J. Mol. Biol.* **209**, 65–77.
10 Tu, C., Tzeng, T.H., and Bruenn, J.A. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8636–8640.
11 Tsuchihashi, Z. (1991) *Nucleic Acids Res.* **19**, 2457–2462.
12 Sipley, J., and Goldman, E. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 2315–2319.
13 Craigen, W.J., Cook, R.G., Tate, W.P., and Caskey, C.T. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3616–3620.
14 Scolnick, E., Tompkins, R., Caskey, C.T., and Nirenberg, M. (1968) *Proc. Natl. Acad. Sci. USA* **61**, 768–774.

15 Adamski, F.M., Donly, B.C., and Tate, W.P. (1993) *Nucleic Acids Res.* **21**, 5074–5078.
16 Poole, E.S., Brown, C.M. and Tate, W.P. (1995) *EMBO J.*, **14**, 151–158.
17 Shine, J., and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 1342–1346.
18 Weiss, R.B., Dunn, D.M., Dahlberg, A.E., Atkins, J.F., and Gesteland, R.F. (1988) *EMBO J.* **7**, 1503–1507.
19 Larsen, B., Wills, N.M., Gesteland, R.F., and Atkins, J.F. (1994) *J. Bacteriol.* **176**, 6842–6851.
20 Larsen, B., Peden, J., Matsufuji, S., Matsufuji, T., Brady, K., Maldonaldo, R., Wills, N.M., Fayet, O., Atkins, J.F., and Gesteland, R.F. (1995) *Biochem. Cell Biol.* **73**, 1123–1129.
21 Curran, J.F. (1993) *Nucleic Acids Res.* **21**, 1837–1843.
22 Tsuchihashi, Z., and Brown, P.O. (1992) *Genes Dev.* **6**, 511–519.
23 Weiss, R.B., Dunn, D.M., Shuh, M., Atkins, J.F., and Gesteland, R.F. (1989) *New Biologist* **1**, 159–169.
24 Fu, C., and Parker, J. (1994) *Mol. Gen. Genet.* **243**, 434–441.
25 Björk, G.R. (1995) *Prog. Nucleic Acids Res. Mol. Biol.* **50**, 263–338.
26 deBoer, H.A., and Kastelein, R.A. (1986) In Reznikoff, W., and Gold, L. (eds), *Maximizing Gene Expression*. Butterworths Publishers, Stoneham, MA, pp. 225–285.
27 Andersson, S.G.E., and Kurland, C.G. (1990) *Microbiol. Rev.* **54**, 198–210.
28 Ikemura, T. (1992) In Hatfield, D.L., Lee, B.J., and Pirtle, R.M. (eds), *Transfer RNA in Protein Synthesis.* CRC Press, pp. 87–111.
29 Buckingham, R.H. (1994) *Biochimie* **76**, 351–354.
30 Hatfield, G.W., and Gutman, G.A. (1992) In Hatfield, D.L., Lee, B.J., and Pirtle, R.M. (eds), *Transfer RNA in Protein Synthesis.* CRC Press, Boca Raton, FL, pp. 157–189.
31 Yarus, M., and Curran, J.F. (1992) In Hatfield, D.L., Lee, B.J., and Pirtle, R.M. (eds.), *Transfer RNA in Protein Synthesis*. CRC Press, Boca Raton, FL, pp. 319–365.
32 Curran, J.F., and Yarus, M. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6538–6542.
33 Miller, J.H. and Albertini, A.M. (1983) *J. Mol. Biol.* **164**, 59–71.
34 Miller, J.H. (1972) *Experiments in Molecular Genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
35 Ericson, J.U., and Bjork, G.R. (1986) *J. Bacteriol.* **166**, 1013–1021.
36 Davis, R.W., Botstein, D., and Roth, J.R. (1980) *A Manual for Genetic Engineering: Advanced Bacterial Genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
37 Vogel, H., and Bonner, O. (1956) *J. Biol. Chem.* **218**, 97–106.
38 Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
39 Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heynecker, H.L., Boyer, H.W., Crosa, J.H., and Falkow, S. (1977) *Gene* **2**, 95–113.
40 Brown, C.M., Stockwell, P.A., Dalphin, M.E., and Tate, W.P. (1993) *Nucleic Acids Res.* **22**, 3119–3123.
41 Folley, L.S., and Yarus, M. (1989) *J. Mol. Biol.* **209**, 359–378.
42 Atkins, J.F., Elseviers, D., and Gorini, L. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1192–1195.
43 Kurland, C.G. (1979) In Celis, J. and Smith, J.D. (eds), *Nonsense Mutations and tRNA Suppressors.* Academic Press, N.Y. pp. 95–108.
44 Barak, Z., Gallant, J., Lindsley, D., Kwieciszewki, B., and Heidel, D. (1996) *J. Mol. Biol.* **263**, 140–148.
45 Björk, G.R. (1987) In Neidhardt, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M., and Umbarger, H.E. (eds), *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology.* ASM Press, Washington, D.C., pp. 719–731.
46 Connolly, D.M., and Winkler, M.E. (1989) *J. Bacteriol.* **171**, 3233–3246.
47 Gorini, L. (1971) *Nature New Biol.* **234**, 261–264.
48 Bohman, K., Ruusala, T., Jelenc, P.C., and Kurland, C.G. (1984) *Mol. Gen. Genet.* **198**, 90–99.
49 Thompson, R.C., Dix, D.B., Gerson, R.B., and Karim, A. (1981) *J. Biol. Chem.* **256**, 6676–6681.
50 Bilgin, N., Claesens, F., Pahverk, H., and Eherenberg, M. (1992) *J. Mol. Biol.* **224**, 1011–1027.
51 Curran, J.F. (1995) *Nucleic Acids Res.* **23**, 683–688.
52 Gallant, J.A., and Lindsley, D. (1992) *J. Mol. Biol.* **223**, 31–40.
53 Björk, G.R., Wikström, P.M., and Byström, A.S. (1989) *Science* **244**, 986–989.
54 Hagervall, T.G., Tuohy, T.M.F., Atkins, J.F., and Björk, G.R. (1993) *J. Mol. Biol.* **232**, 756–765.
55 Lim, V.I. (1997) *J. Mol. Biol*., **266**, 877–890.
56 Grosjean, H.J., DeHenau, S., and Crothers, D.M. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 610–614.
57 Vacher, J., Grosjean, H., Houssier, C. and Buckingham, R.H. (1984) *J. Mol. Biol.* **177**, 329–342.
58 Petrullo, L.A., Gallagher, J.P., and Elseviers, D. (1983) *Mol. Gen. Genet.* **190**, 289–294.
59 Bouadloun, F., Srichaiyo, T., Isaksson, L.A., and Björk, G.R. (1986) *J. Bacteriol.* **166**, 1022–1027.
60 Ericson, J., and Björk, G.R. (1991) *J. Mol. Biol.* **218**, 509–516.
61 Esberg, B., and Björk, G.R. (1995) *J. Bacteriol.* **177**, 1967–1975.
62 Bjornsson, A., and Isaksson, L. (1993) *J. Mol. Biol.* **232**, 1017–1029.
63 Diaz, I., and Eherenberg, M. (1991) *J. Mol. Biol.* **222**, 1161–1171.