Molecular Microbiology (1994) 11(2), 303–313

The function of a ribosomal frameshifting signal from human immunodeficiency virus-1 in *Escherichia coli*

Elizabeth Yelverton, Dale Lindsley, Phil Yamauchi and Jonathan A. Gallant*
University of Washington, Department of Genetics, SK-50, Seattle, Washington 98195, USA.

Summary

A 15–17 nucleotide sequence from the *gag–pol* ribosome frameshift site of HIV-1 directs analogous ribosomal frameshifting in *Escherichia coli*. Limitation for leucine, which is encoded precisely at the frameshift site, dramatically increased the frequency of leftward frameshifting. Limitation for phenylalanine or arginine, which are encoded just before and just after the frameshift, did not significantly affect frameshifting. Protein sequence analysis demonstrated the occurrence of two closely related frameshift mechanisms. In the first, ribosomes appear to bind leucyl-tRNA at the frameshift site and then slip leftward. This is the 'simultaneous slippage' mechanism. In the second, ribosomes appear to slip before binding aminoacyl-tRNA, and then bind phenylalanyl-tRNA, which is encoded in the left-shifted reading frame. This mechanism is identical to the 'overlapping reading' we have demonstrated at other bacterial frameshift sites. The HIV-1 sequence is prone to frameshifting by both mechanisms in *E. coli*.

Introduction

Ribosomes normally maintain a constant reading frame from AUG to the finish, but they are capable of slipping into an alternative reading frame at an average frequency of the order of 10^{-4} (Atkins *et al.*, 1972; J. A. Gallant *et al.*, unpublished). In certain special cases, much higher frequencies of ribosome frameshifting occur. These cases include production of polypeptide release factor 2 of *Escherichia coli*, which depends upon a rightward frameshift within the coding sequence (Craigen *et al.*, 1985; Craigen and Caskey, 1987; Weiss *et al.*, 1987; Curran and Yarus, 1988); translation of the reverse transcriptase of the yeast Ty element, which also depends upon a rightward frameshift (Clare *et al.*, 1988); and translation of the RNA of several retroviruses, which express *gag–pol* and *gag–pro–pol* polyproteins by means of leftward frameshifts (reviewed by Hatfield and Oroszlan, 1990; Cattaneo, 1989).

Ribosomal frameshifting in both rightward and leftward directions has also been shown to occur at certain 'hungry' codons whose cognate aminoacyl-tRNAs are in short supply (Gallant and Foley, 1980; Weiss and Gallant, 1983; 1986; Gallant *et al.*, 1985; Kurland and Gallant, 1986). Not all hungry codons are equally prone to shift: in a survey of 21 frameshift mutations of the rIIB gene of phage T4, Weiss and Gallant (1986) found that only a minority were phenotypically suppressible when challenged by limitation for any of several aminoacyl-tRNAs.

The context rules governing ribosome frameshifting at hungry sites are under investigation, and have been defined in a few cases (Weiss *et al.*, 1988; Gallant and Lindsley, 1992; Peter *et al.*, 1992; Kolor *et al.*, 1993; Lindsley and Gallant, 1993). So far these sequences do not resemble any of the naturally occurring shifty sites summarized in the first paragraph above. In order to find out whether these two categories of ribosome frameshifting are mechanistically related, we have tested the susceptibility of a well-studied retroviral frameshift site to manipulation by aminoacyl-tRNA limitation in *E. coli*. We have directed our analysis to the shifty site at the *gag–pol* junction of HIV-1 both because of its clinical interest, and because certain features render it convenient for analysis.

In some viral systems, baroque secondary structures in the mRNA downstream of the frameshift site are required to augment frameshifting levels (Jacks *et al.*, 1988b; Brierley *et al.*, 1989). In the case of HIV-1, however, although a stem–loop structure might exist downstream of the frameshift site (Jacks *et al.*, 1988a), direct modification or elimination of the stem–loop sequence has little effect on the rate of frameshifting (Madhani *et al.*, 1988; Weiss *et al.*, 1989). Moreover, Wilson *et al.* (1988) demonstrated that a short (21 nucleotide) sequence of HIV-1 without the stem–loop was sufficient to direct a high level of frameshifting in heterologous *in vitro* systems.

The site of ribosomal frameshifting at the slippery sequence U–UUU–UUA has been directly established by amino acid sequencing of frameshifted proteins (Jacks *et al.*, 1988b), and the participation of certain aminoacyl-tRNAs has been clearly implicated by mutagenesis of the monotonous tract of uridines (Jacks *et al.*, 1988b; Wilson *et al.*, 1988). Our purpose was to discover

Received 13 July, 1993; revised and accepted 27 September, 1993.
*For correspondence. Tel. (206) 543 8235; Fax (206) 543 0754.

whether the ribosomal frameshifting directed by a very short sequence in HIV-1 could be reproduced by *E. coli* ribosomes *in vivo*, and, if so, whether we could alter the rate of frameshifting by regimens that change the relative abundance of key aminoacyl tRNAs encoded at or near the frameshift site. Weiss *et al.* (1989) have also reported that a 52 nucleotide fragment from HIV-1 is sufficient to direct ribosomal frameshifting in an *E. coli* system. In this report we present evidence that a much shorter 15-17 nucleotide sequence derived from HIV-1 is sufficient to direct the same ribosomal frameshift event in *E. coli* as in eukaryotes. We also show that in *E. coli* the rate of ribosomal frameshifting on that sequence can be increased by limitation for leucine, the amino acid encoded at the frameshift site. Protein sequence analysis of the product indicates the occurrence of two slightly different mechanisms of shifting.

Results

The strategy behind the construction of our assay system for ribosomal frameshifting may be understood with reference to Fig. 1. When eukaryotic ribosomes decode the HIV mRNA sequence . . . U UUU UUA GGG . . ., shown as nucleotides 1–10 in Fig. 1A, the adenine at position 7 appears to be read twice: first, as the third position of a leucine codon (UUA) and then again as the first position in the overlapping arginine codon (AGG). The resultant leftward ribosomal frameshift, most often referred to as a (−1) frameshift in the retroviral literature, has been estimated

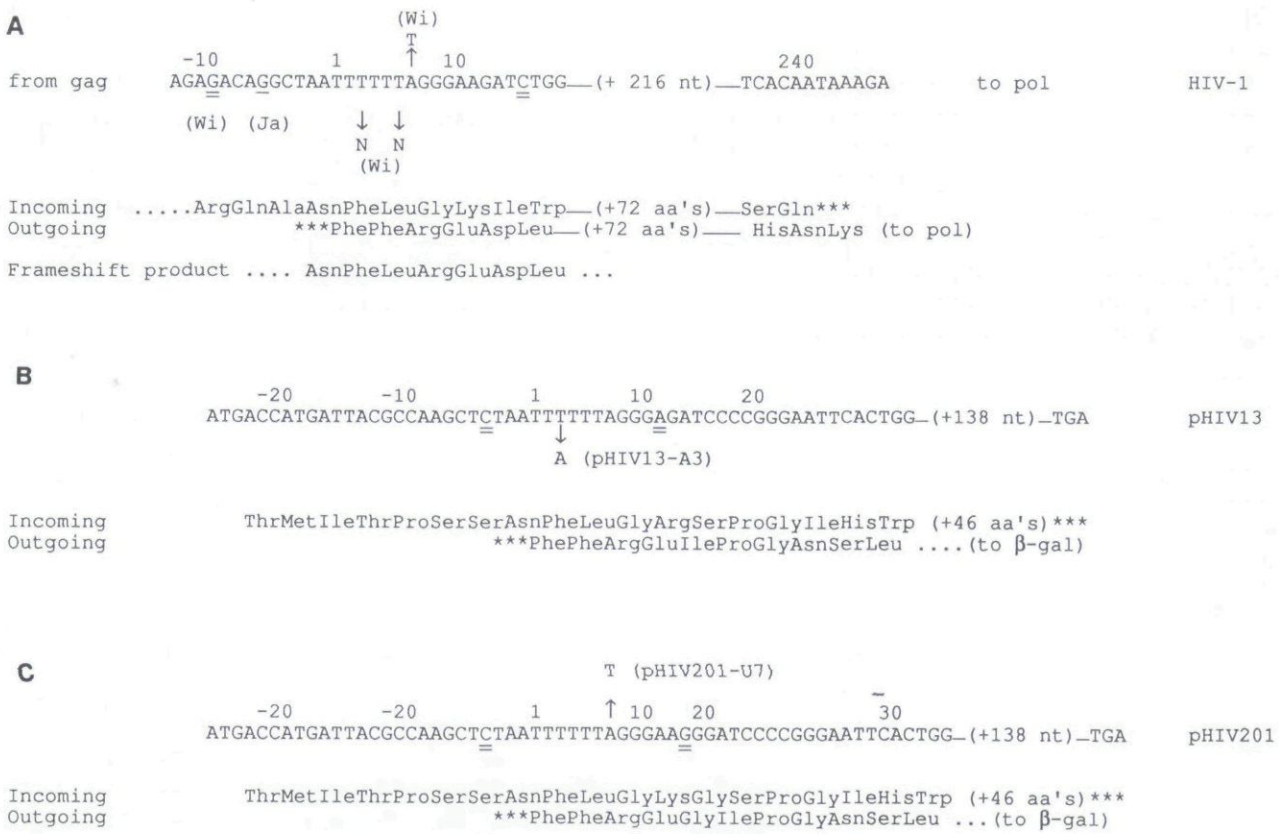

Fig. 1. A portion of the *gag–pol* junction in the HIV-1 genome is depicted. For clarity, the wild-type nucleotide sequence and mutations derived from it are presented as they occur in mRNA. Arbitrary nucleotide position 1 marks the beginning of the *gag–pol* junction; this corresponds to nucleotide number 1629 of Ratner *et al.* (1985). In a heterologous mammalian *in vitro* translation system, most of the frameshift product has the amino acid sequence . . . Asn–Phe–Leu–Arg (Jacks *et al.*, 1988b), where Leu is encoded by positions 5–7 and Arg is encoded by positions 7–9. Some mutations that result in increased or decreased expression of frameshift products in a heterologous test system are shown above and below the nucleotide sequence, respectively (Wilson *et al.*, 1988). 'N' signifies a mutation to any non-U base. Double underlines at positions −10 and +16 mark the boundaries of a fragment that directs the synthesis of a frameshift protein product in a heterologous yeast system (Wilson *et al.*, 1988). The singly underlined G at position −6 is the 5′ boundary of a fragment that directs the synthesis of a frameshift protein product in a mammalian *in vitro* system (Jacks *et al.*, 1988b).
B. and C. A portion of the mRNA sequences expressed from *lacZ* frameshift alleles HIV13, HIV13-A3, HIV201, and HIV201-U7 are depicted. Numbers above the nucleotide sequence correspond to analogous positions of the HIV-1 *gag–pol* junction. Doubly underlined nucleotides mark the boundaries of sequence that is conserved with respect to HIV-1. Synthesis of β-galactosidase from the alleles requires a leftward frameshift. Amino acids are shown for the mature protein, after *in vivo* cleavage of the initiating *N*-terminal formyl-met residue.

Table 1. Differential rates of β-galactosidase synthesis in exponential cultures.

Construct	Differential rate (EU mg^{-1} protein) ± SEM (n)
pBW1100 (*lacZ*$^+$)	173 ± 25 (12)
pHIV13 (U UUU UUA)	2.6 ± 0.46 (13)
pHIV13-A3 (U UAU UUA)	0.55 ± 0.025 (4)
pHIV201 (U UUU UUA)	1.84 ± 0.17 (22)
pHIV201-U7 (U UUU UUU)	4.4 ± 0.36 (4)

Constructs pHIV13 and pHIV201, and their variants pHIV13-A3 and pHIV201-U7, are described in Fig. 1. (The sequence of the critical heptanucleotide at the frameshift site is shown in parenthesis after each construct's designation.) All constructs were transformed into a derivative of CP79 (*relA2 thr*$^-$ *leu*$^-$ *his*$^-$ *arg*$^-$ *thi*$^-$) carrying a complete deletion of *lacZ*. Methods of cultivation, and enzyme and protein assay were as described previously (Gallant and Lindsley, 1992; Peter *et al.*, 1992). Cells were grown into exponential phase in M63-glucose medium supplemented with all required amino acids plus Ile and Val. The *lac* promoter was induced (2 mM IPTG and 2.5 mM cAMP) for about one doubling. Data are reported ± standard error of the mean, with the number of replicate induced cultures in parentheses. These values include all the unstarved control cultures from the various starvation experiments.

to take place in about 10% of ribosomal transits (Jacks *et al.*, 1988b). In HIV-1, the outcome of the leftward ribosomal frameshift is the successful production of the *gag–pol* fusion protein. In the assay system we have devised, the outcome of an analogous leftward frameshift by *E. coli* ribosomes will be the successful production of the enzyme β-galactosidase from genetically frameshifted alleles of the *lacZ* gene. We have previously used an assay system of similar design to demonstrate that lysyl-tRNA starvation can amplify ribosomal frameshifting in either direction at lysine codons, given certain context rules (Gallant and Lindsley, 1992; Peter *et al.*, 1992; Lindsley and Gallant, 1993).

Alleles to be tested were constructed by the ligation of paired complementary oligonucleotides into the *Hin*dIII-*Bam*HI site of pBW1100, as described in Gallant and Lindsley (1992). Figure 1 shows the sequence of the translated strand from the region of our constructs that reproduces the *gag–pol* frameshift signal from HIV-1. The *lacZ* frameshift alleles carried on plasmids pHIV13 and pHIV201 are constructed so that a shift to the left by one base, as in the expression of the *gag–pol* fusion of HIV, is required to generate active enzyme. The two constructs both carry a short sequence identical to the region around the frameshift site in the *gag–pol* overlap of HIV-1 for 15 nucleotides in pHIV13 and 17 nucleotides in pHIV201 (see Fig. 1); they differ slightly from one another several bases downstream of the frameshift site. Host cells carrying either of these plasmids produce active enzyme at about 1% of the efficiency of cells carrying a control *lacZ*$^+$ plasmid (Table 1). This basal value is close to the value (1.8%) observed by Weiss *et al.* (1989) for frameshifting on a much longer HIV-derived sequence in a similar β-galactosidase reporter. It is also much higher than the frequency of leftward frameshifting (0.03–0.2%) we observed previously at sequences unrelated to HIV (Gallant and Lindsley, 1992). The presence of the HIV sequence in our reporter thus leads to an unusually high frequency of leftward frameshifting.

Modification of the critical heptanucleotide sequence from U UUU UUA to U UAU UUA in plasmid pHIV13-A3 decreased frameshifting about fivefold, while modification of the heptanucleotide to U UUU UUU in pHIV201-U7 increased frameshifting by two- to threefold (Table 1). These genetic results are analagous to earlier findings in other reporter systems (Jacks *et al.*, 1988b; Wilson *et al.*, 1988; Weiss *et al.*, 1989) and suggest that the heptanucleotide is the predominant site of leftward frameshifting in our reporter system as well.

Chemical evidence that this is indeed the case was obtained from protein sequence analysis. The protein encoded by pHIV201 was purified and subjected to automated Edman sequence analysis, as previously described (Gallant and Lindsley, 1992; Peter *et al.*, 1992). The results (Fig. 2) are readily interpreted with reference to the nucleotide sequence and coding potential shown in Fig. 1. For the first 10 cycles, the amino acid sequence corresponds to normal translation in the (0) reading frame: T-M-I-T-P-S-S-N-F-L. After this position, the signal (R-E-G-I-P) clearly corresponds to the leftward reading frame. Thus, a leftward shift occurred after incorporation of the leucine at position 10, exactly the frameshift characteristic of HIV translation in eukaryotic systems (Jacks *et al.*, 1988b). Similar results were obtained with plasmid pHIV13 (data not shown). Thus, our short cloned sequence reproduces the specificity of the HIV frameshift signal when translated by *E. coli* ribosomes.

We asked whether the HIV sequence's rather high level of shiftiness could be modulated by stalling the ribosome at hungry codons within the frameshift region. Phe and leu are the last two amino acids encoded in the initial reading frame before the shift, and Arg is the first encoded in the new reading frame after the shift. We imposed limitation for phe-tRNA by inhibiting the synthetase with the analogue phenylalanine-hydroxamate (PHX). Graded limitations for arginine or leucine were imposed by growing auxotrophic host cells on a range of concentrations of slowly utilized methyl esters of these amino acids (Buston and Bishop, 1955). Tables 2 and 3 present the results of all these limitation regimens.

Considering first Arg and Phe (Table 2), we observe that both limitations substantially reduced the differential rate of enzyme synthesis expressed from both the frameshift construct and the *lacZ*$^+$ control. The reduction was about the same in both genotypes, indicating that neither arginine limitation nor phe-tRNA limitation had a specific effect on the frequency of frameshifting within the HIV sequence.

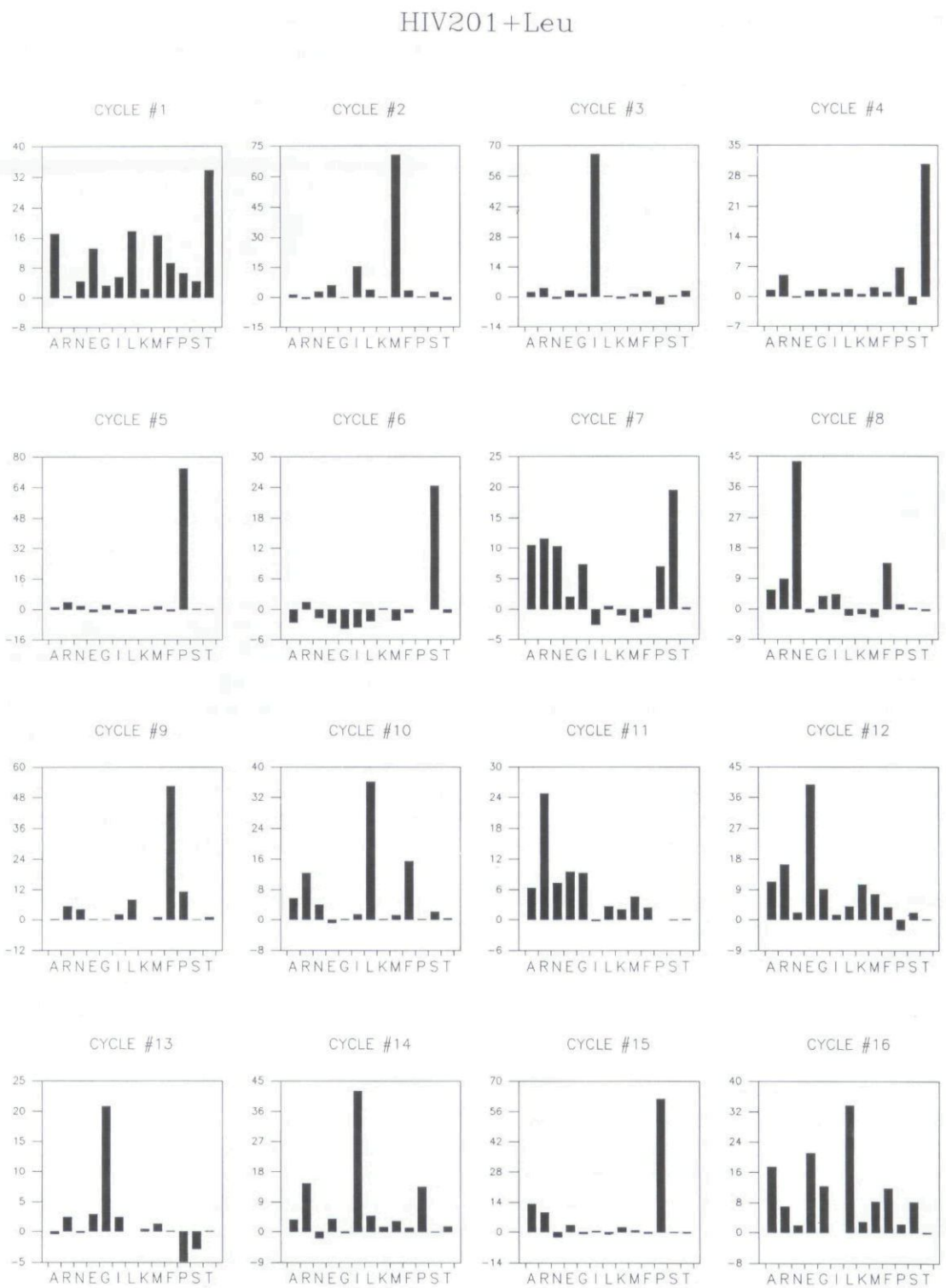

Fig. 2. Protein sequence analysis of 201 β-galactosidase made during exponential growth. In order to ensure leucine sufficiency, the host was a leucine prototroph isogenic with our normal host, and it was grown nevertheless in the presence of 100 $\mu g\,ml^{-1}$ leucine. The *lac* promoter was induced for one doubling and two litres of cells were harvested. Beta-galactosidase was isolated as described previously (Gallant and Lindsley, 1992; Peter *et al.*, 1992) and subjected to automated Edman sequencing. The raw data for picomole amounts of PTH amino acids in each cycle were corrected for background and lag using Applied Biosystems software, as described (Gallant and Lindsley, 1992). The figure presents the sum of three independent sequencing runs, one performed at Illinois and two at Riverside. In every cyle, the most abundant PTH-amino acid released predominates over others by at least a factor of two, affording a clear signal of the primary sequence. Secondary signals reflect a variety of artefacts. One is minor peaks which anticipate the next cycle (for example, F in cycle 8, R in cycle 10, E in cycle 11, and G in cycle 12), presumably owing to a minority of the protein that is one amino acid shorter on the *N*-terminus than the bulk of the material.

Table 2. Growth rates and differential rates of β-galactosidase synthesis during aminoacyl-tRNA limitations.

Growth regimen	201 (Frameshift) relative growth rate	201 (Frameshift) diff. rate EU mg^{-1}	201 (Frameshift) DR rel. to control culture	1100 (*lacZ*$^+$) relative growth rate	1100 (*lacZ*$^+$) diff. rate EU mg^{-1}	1100 (*lacZ*$^+$) DR rel. to control culture
AME (μg ml^{-1})						
Control	≡1.0	1.54	≡1.0	≡1.0	258	≡1.0
40	0.43	0.63	0.41	0.41	53	0.205
20	0.35	0.46	0.30	0.26	37	0.14
10	0.125	0.54	0.35	0.16	68	0.26
PHX (μg ml^{-1})						
Control	≡1.0	1.7	≡1.0	≡1.0	154	≡1.0
50	0.45	0.78	0.46	0.41	60	0.39
100	0.265	0.83	0.49	0.24	49	0.32
200	0.16	0.25	0.15	0.14	41	0.27

Exponential cells carrying plasmid pHIV201 or the control *lacZ*$^+$ plasmid pBW1100 were subjected to limitation for phenylalanyl-tRNA by addition of phenylalanine-hydroxamate (PHX) at the indicated concentrations; and for arginine by centrifuging and washing the cells, and resuspending them in growth medium containing the indicated concentrations of arginine-methyl-ester (AME), with a control culture containing arginine at our standard concentration of 100 μg ml^{-1}. The *lac* promoter was induced as in Table 1, and cells harvested for assay after about one doubling. Cultures are listed in order of decreasing growth rate. Growth rate varies directly with the concentration of AME, which serves as a source of required arginine, and inversely with the concentration of PHX, which inhibits phenylalanine activation.

In the case of leucine limitation, a different outcome was suggested in preliminary screening by a crude, qualitative Petri plate assay (Kurland and Gallant, 1986). Quantitative data on log-phase cells are shown in Table 3. In the *lacZ*$^+$ control, leucine limitation progressively restricted enzyme synthesis, reducing it very severely at low growth rates. In the HIV construct, on the other hand, there was no such restriction, but rather a small stimulation of enzyme synthesis.

If this effect is caused by frameshifting at the leucine codon within the HIV sequence, then it should be eliminated by eliminating that leucine codon. As an additional control, therefore, we tested 201-U7, a construct that is identical to 201 except for the conversion of the leucine UUA codon to UUU (see Fig. 1). In this construct, leucine limitation did not stimulate enzyme synthesis as in 201, but rather reduced it, as it did in the wild type (Table 4). The reduction was a good deal less than in the wild type, suggesting that some frameshifting at various positions still occurs in 201-U7, so as to offset partly the effect of reduced *lacZ* expression. However, the difference between the response of 201 and that of 201-U7 demonstrates that the leucine codon eliminated in the latter must be a major determinant of frameshifting induced by leucine limitation.

Decisive identification of the frameshift site during leucine limitation was obtained from protein sequence analysis, as described earlier. The results are shown in

Table 3. Relative growth rates and differential rates of β-galactosidase synthesis during leucine limitation.

Growth regimen	201 (Frameshift) relative growth rate	201 (Frameshift) relative diff. rate	1100 (*lacZ*$^+$) relative growth rate	1100 (*lacZ*$^+$) relative diff. rate
LME (μg ml^{-1})				
30	0.52	1.27 ± 0.16 (4)	0.44	0.55 ± 0.06 (3)
10	0.26	1.35 ± 0.11 (6)	0.24	0.13 ± 0.031 (4)
5	0.175	1.48 ± 0.16 (3)	0.165	0.076 ± 0.0032 (3)
3	0.13	2.0 ± 0.135 (4)	0.135	0.041 ± 0.011 (3)

Exponential cells carrying plasmid pHIV201 were spun, washed, and resuspended in medium supplemented with the indicated concentrations of leucine-methyl-ester (LME). These were grown in parallel with a control aliquot returned to growth in leucine at 100 μg ml^{-1}. The entries show growth rates and differential rates of enzyme synthesis relative to the control leucine culture, expressed as mean values ± standard error (number replicate experiments). The mean control differential rates in these experiments were 2.3 ± 0.16 (6) for pHIV201 and 191 ± 44 (4) for pBW1100.

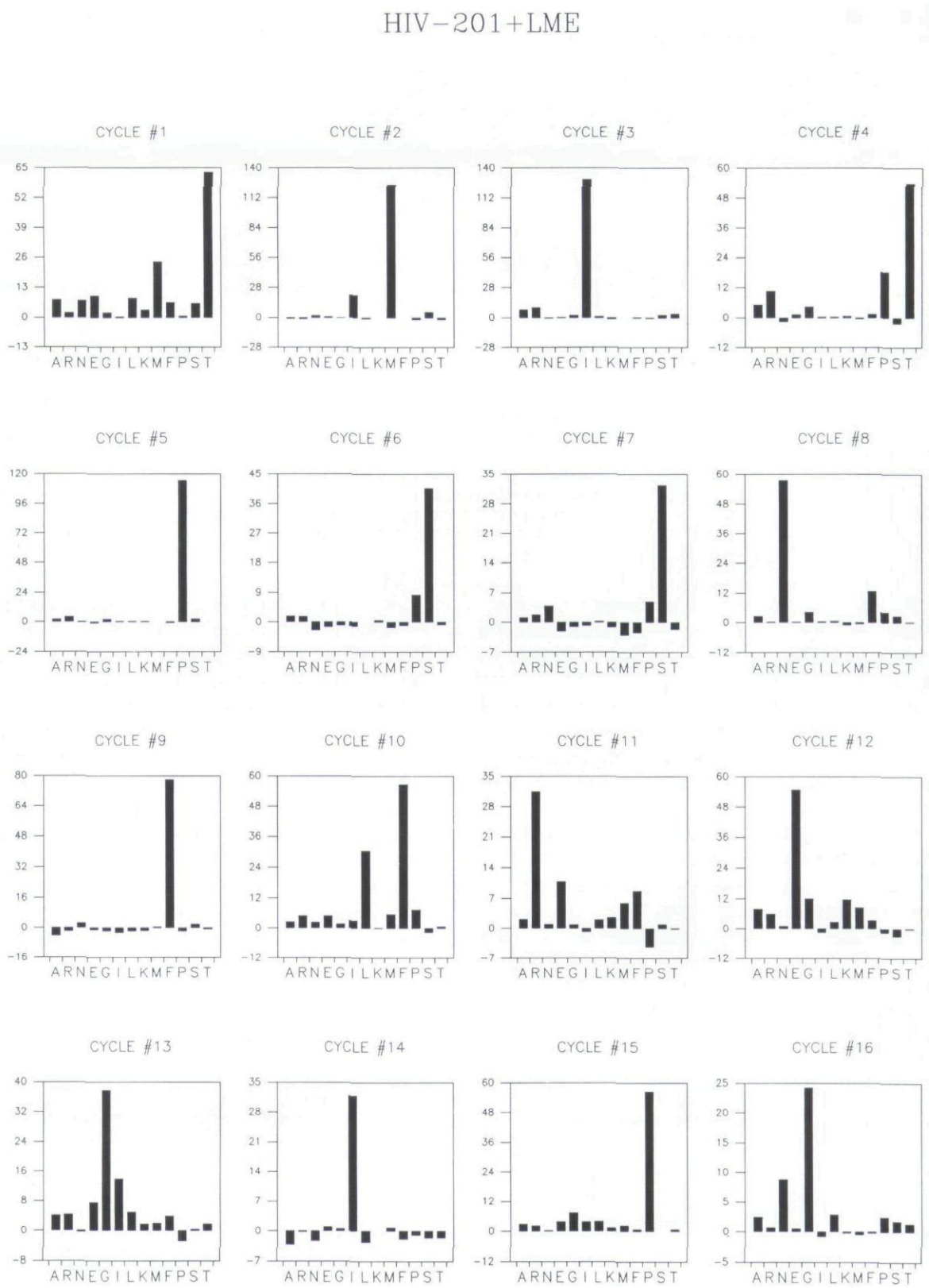

Fig. 3. Protein sequence analysis of 201 β-galactosidase made during leucine limitation. Two litres of a late-log culture were centrifuged, washed, and resuspended in 10 l of medium containing 10 $\mu g\,ml^{-1}$ leucine-methyl-ester in a bench-top fermentor. The cells were induced and grown for about one doubling, harvested, and β-galactosidase isolated and sequenced as in Fig. 2. The figure presents the sum of four independent sequencing runs, two at Illinois and two at Riverside.

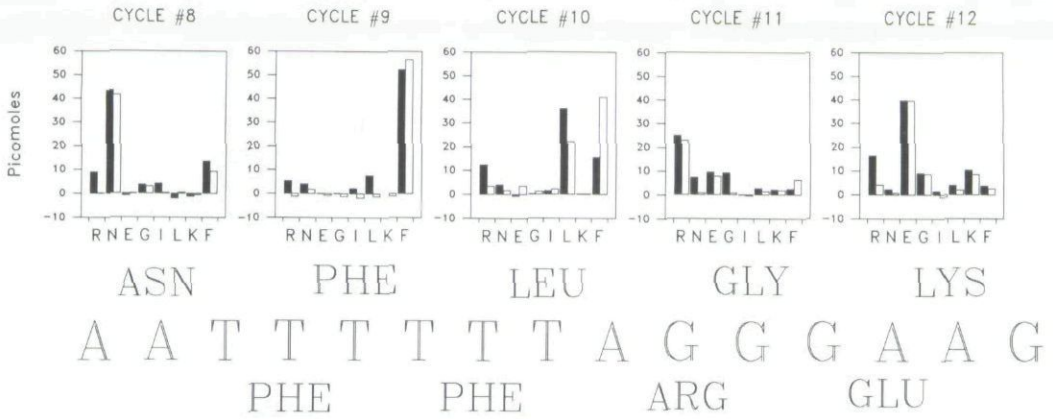

Fig. 4. Comparison of protein sequences at and around the frameshift junction. The data are taken from cycles 8 to 12 of Fig. 2 and Fig. 3. In order to normalize to the same number of picamoles, all the values for the leucine-starved protein were multiplied by the ratio (unstarved/starved) of the sums of the predominant peaks (N in cycle 8, F in 9, R in 11, and E in 12). The ratios of leucine to phenylalanine in cycle 10 of the two sets of data were as follows. For the unstarved case, the ratios in three runs were 2.55, 1.9, and 3.5, yielding an average of 2.65 with a standard error of 0.46. For the leucine-starved case, the ratios in four runs were 0.52, 0.58, 0.68, and 0.51, yielding an average of 0.57 with a standard error of 0.039. The difference is highly significant (P less than 0.005) by the τ_d test (Crow *et al.*, 1969). The nucleotide sequence in the region of interest is shown below the cycle bar plot, with the expected amino acids read in the initial (0) reading frame shown above the nucleotide sequence, and the expected amino acids read in the leftward reading frame shown below. Solid bars = HIV-201 unstarved (+Leu); open bars = HIV-201 starved (+LME) normalized to unstarved values.

Fig. 3. The protein sequence is very similar to that of the protein from unstarved cells (Fig. 2), with one significant exception. For the first nine cycles, the initial T-M-I-T-P-S-S-N-F sequence of the normal (0) reading frame is clear, as it was in the protein from unstarved cells. From cycle 11 onwards the signature of the leftward reading frame, R-E-G-I-P-G, is equally clear. The one difference between the starved and unstarved proteins is at cycle 10, the position of the leucine codon. The predominant amino acid in the starved protein is F (phenylalanine) rather than L at this position.

Three considerations argue that carryover of F from cycle 9, the preceding position, does not contribute significantly to the peak of F in cycle 10 of the leucine-starved protein. First, the data in every cycle have been corrected for carryover by means of the Applied Biosystems lag-correction algorithm (Hunkapiller, 1986). The effectiveness of this correction can be seen in the very low levels of carryover throughout the cycle bar plots of Figs 2 and 3, and of our earlier publications (Peter *et al.*, 1992; Gallant and Lindsley, 1992). Second, amino acids differ in their susceptibility to carryover from one cycle to the next (proline, for example, being notoriously bad), and it is carryover of F that we need to assess in considering these data. In fact, Figs 6 and 7 of Gallant and Lindsley (1992) present exactly the needed control: F was the amino acid present in cycle 7 of both sequences, and cycle 8 corresponded to a 'hungry' codon, just as in the present case. In those two sets of data, which represent the average of four independent sequencing runs in each case, carryover of F from cycle 7 to the next cycle averaged 8.3% (standard error = 2.8%), far less than the relative level of F found in cycle 10 of the present cycle bar plots.

Finally, the starved and unstarved proteins are virtually identical at every position except cycle 10, in which the difference in regard to phenylalanine and leucine was replicated in several independent sequencing runs on protein of each type. A summary of this comparison of the starved and unstarved proteins in and around cycle 10, and the statistics for this cycle, are presented in Fig. 4.

Phenylalanine corresponds to the UUU triplet overlapping the hungry leucine codon in the leftward reading frame. We conclude that most ribosomes which produced β-galactosidase slipped leftward when stalled at the leucine codon before aminoacyl-tRNA binding, and then bound phe-tRNA cognate to the overlapping phenylalanine codon. However, cycle 10 contains a significant

Table 4. Relative growth rates and differential rates of β-galactosidase synthesis in 201-U7.

	201-U7	
Growth regimen	relative growth rate	relative differential rate
LME ($\mu g\,ml^{-1}$)		
30	0.41	1.01 ± 0.20 (3)
10	0.23	0.51 ± 0.10 (3)
5	0.17	0.43 ± 0.088 (3)
3	0.11	0.36 ± 0.056 (3)

Exponential cells carrying plasmid pHIV201-U7 were treated as in Table 3. The mean control differential rate was 4.7 ± 0.20 (3).

secondary peak of leucine as well, suggesting that a minority of ribosomes bound leucyl-tRNA and then shifted left, just as in unstarved cells.

Discussion

Our wild type *lacZ*$^+$ control shows a marked decrease in β-galactosidase synthesis during limitation for each of the amino acids tested. This response is typical of *relA*$^-$ cells, and is primarily caused by a dependence of *lacZ* transcription on ppGpp during amino acid limitation (Primakoff and Artz, 1979; Yang *et al.*, 1979; Primakoff, 1981; Foley *et al.*, 1982), compounded by errors in translation at some hungry codons (reviewed by Gallant, 1979; Cashel and Rudd, 1987; Parker, 1989). Expression of the *lacZ* gene carrying the HIV frameshift signal should be subject to the same *relA*$^-$ defects and is indeed reduced to the same extent as wild-type during limitation for arginine or for phenylalanyl-tRNA (Table 2).

During leucine limitation, in contrast, enzyme synthesis is increased rather than decreased in cells carrying the frameshift allele pHIV201 (Table 3). In 201-U7, which is identical to 201 except for the absence of the leucine codon at the frameshift site, leucine limitation provokes a decrease in enzyme synthesis (Table 4), as it does in the *lacZ*$^+$ control. The fact that enzyme synthesis increases with leucine limitation in 201 means that an increase in frameshifting more than compensates for the sharp decrease in *lacZ* expression found in the control experiments with the wild type and with 201-U7. The increased frameshifting attributable to this one leucine codon can be estimated as the ratio of the effect of leucine limitation on 201 to that on 201-U7. In the former case, enzyme synthesis was two times greater in 3 μg ml^{-1} of leucine-methyl-ester than in leucine, whereas in the latter case it was three times less. By this estimate, therefore, the frequency of frameshifting at the critical leucine codon was increased about sixfold at 3 μg ml^{-1} of the leucine analogue, a limitation regimen which reduced growth about sevenfold.

The protein sequence of the frameshifted material (Fig. 3) demonstrates that the ribosomes read the sequence in a normal fashion for the first nine codons, up to the hungry codon calling for leucine at position 10. At this position, the predominant amino acid was F (phenylalanine) rather than L (leucine), implying that the ribosomes slipped leftward to decode the TTT triplet for phenylalanine, which overlaps the hungry codon from the left. This overlapping mode of reading is entirely analogous to the case of leftward frameshifting at a different hungry codon we have described earlier (Gallant and Lindsley, 1992).

However, there is also a significant secondary peak of leucine at position 10, amounting to one third of the total signal (Fig. 3). This is the amino acid encoded in the initial or (0) reading frame. The presence of leucine here strongly suggests that a minority of ribosomes bound leucyl-tRNA and then slipped leftward, just as a majority of ribosomes did under conditions of leucine sufficiency.

Thus, the HIV sequence is subject to starvation-promoted frameshifting on *E. coli* ribosomes, and by two seemingly different mechanisms. One is overlapping reading at the hungry codon, like the cases we have analysed elsewhere (Weiss and Gallant, 1986; Gallant and Linsdley, 1992). The second is zero-frame reading and subsequent slippage—inserting leucine at position 10 in this case—analagous to other studies of the translation of HIV (Jacks *et al.*, 1988b; Weiss *et al.*, 1989) or rous sarcoma virus (RSV) (Jacks *et al.*, 1988a) in heterologous systems.

In fact, a close reading of the data of Jacks *et al.* (1988b) reveals that the HIV sequence also displayed both modes of frameshifting in the eukaryotic cell-free translation system they analysed. The predominant amino acid at the frameshift site in their study was leucine, but there was also about 20–25% phenylalanine (see Fig. 2 of Jacks *et al.*, 1988b), indicative of overlapping reading. Our results in unstarved cells (our Fig. 2) indicate about 30% phenylalanine.

These two modes of frameshifting share fundamental similarities, most strikingly in regard to their sequence requirements in the four positions to the left of the frameshift site. In the case of the mechanism of (0) frame tRNA binding followed by slippage, mutation studies in this region strongly suggest that the frameshift is facilitated by a leftward slip of peptidyl-tRNA in the P-site (Jacks *et al.*, 1988a,b; Wilson *et al.*, 1988; Weiss *et al.*, 1989). In the case of the overlapping reading mechanism, our mutation studies suggest a similar involvement of a peptidyl shift (Kolor *et al.*, 1993). Thus, in both mechanisms the leftward slippage at the ribosome's A-site is facilitated if the adjacent peptidyl-tRNA finds complementary pairing with a left-shifted triplet at the P-site.

The distinction between the two mechanisms seems, at first thought, to reside in the order in which aminoacyl-tRNA binding and slippage of the message occur. However, this distinction may be more apparent than real. It could be that the aminoacyl-tRNA that reads the overlapping codon first enters the A-site in the (0) frame by way of a mismatch. In the present case, this would mean phe-tRNA pairing initially with the UUA leucine codon, an interaction which demands one mismatch in the third position. This non-cognate interaction would of course be favoured by a shortage of leucyl-tRNA, particularly in our *relA*$^-$ host cells which suffer increased recognition errors at hungry codons (Gallant, 1979; Cashel and Rudd, 1987; Parker, 1989).

Subsequently, leftward slippage of the phe-tRNA in the

A-site to the overlapping UUU triplet would provide complementarity at all three positions, and thus improve stability. In this way, an initial recognition error in the (0) reading frame could be resolved by way of a reading frame error, an example of what has elsewhere been termed 'error coupling' (Kurland and Gallant, 1986).

The *pol* gene product expressed from intact HIV in human host cells has yet to be characterized at the amino acid sequence level. Consequently, it is unknown whether the amino acid at the frameshift junction is leucine or phenylalanine (or a mixed population), and it is therefore unknown which mode(s) of frameshifting operates under natural conditions. The overlapping reading mechanism is worth bearing in mind because of the obvious possibilities it presents for physiological control.

In prokaryotic systems, frameshifting by overlapping reading is a natural outcome of imbalances in the aminoacyl-tRNA pools (Atkins *et al.*, 1979; Gallant and Foley, 1980; Weiss and Gallant, 1983; 1986; Kurland and Gallant, 1986; Bruce *et al.*, 1986; Weiss and Gallant, 1986). Our present results demonstrate that a translational frameshift signal from HIV-1 can be regulated in this fashion in response to changing levels of leucyl-tRNA. Similar regulation might be a simple but effective method for establishing physiological control of the expression of any frameshifted gene. Many such genes function to replicate and spread viruses, transposable elements, and retroviruses (reviewed by Hatfield and Oroszlan, 1990; Cattaneo, 1989). In these cases, ribosome frameshifting during aminoacyl-tRNA imbalance might represent a mechanism for linking element movement to the state of the cellular economy.

The degree of aminoacylation of tRNAs provides a sensitive difference signal of cellular activity. Anabolism generates the amino acids which charge the tRNAs, while translation drains these amino acids into growing proteins. Hence, the tRNA charging level reflects the balance of these two flows. In bacterial cells, tRNA aminoacylation levels operate both specific attenuation control circuits (Landick and Yanofsky, 1987) and the highly pleiotropic stringent control system (Gallant, 1979; Cashel and Rudd, 1987), and may regulate the expression of individual translation factor genes (Grunberg-Manago, 1987).

This speculation might provide a rationale for the various different shifty sequences found in the *gag–pol* and *gag–pro–pol* overlaps of different retroviruses, and the comparable regions of different coronaviruses. Perhaps each unique shifty sequence has been evolutionarily optimized to respond to a tRNA deficiency that is the most sensitive difference signal for its particular host cells. Likewise, evolutionary pressure to maintain a response to a cell-specific aminoacyl-tRNA signal might also explain why retroviruses maintain non-optimal translational frameshift sequences which must in some cases be augmented by baroque mRNA secondary structures downstream (Weiss *et al.*, 1989).

Finally, we might note that our data do not distinguish between a simultaneous slippage mechanism at the leucine codon (as proposed by Jacks and Varmus) and an overlapping reading mechanism at the next codon. If ribosomes in unstarved cells paused at the GGG codon following the leucine codon, then arg-tRNA could enter by overlapping reading and shift the reading frame leftward (see Fig. 4). The amino acid sequence in this case would be exactly that ascribed to simultaneous slippage, and which we have demonstrated in unstarved *E. coli* cells. We have no independent reason to expect ribosome stalling at the GGG codon in growing cells, except that it is a rare codon and calls for a rare tRNA. Experiments designed to investigate this possibility are under way.

Experimental procedures

Method for allele construction

Frameshift alleles were constructed by modification of the plasmid vector pBW1100. pBW1100, a kind of gift of Bob Weiss, is a modified pBR322 in which nucleotides 1–1417 were replaced by nucleotides 1–4625 (NIH GenBank co-ordinates) of the *E. coli lac* operon. The *lac* operon fragment has been modified by the deletion of nucleotides 641–1070, the removal of the natural *Eco*RI site within the *lacZ* gene, and by the addition of the pUC9 polylinker (Vieira and Messing, 1982) between codons 4 and 5. The resultant pBW1100 is $LacI^{-}Z^{+}Y^{-}A^{-}$, and confers resistance to ampicillin. In typical constructions, about 1 μg of pBW1100 was cut to completion with restriction endonucleases *Hin*dIII and *Bam*HI (Bethesda Research Labs) at their single sites within the pUC9 polylinker. About 100 ng of the prepared vector was used without purification in a 20 μl ligation containing a 10-fold molar excess of paired complementary oligonucleotide primers, kindly prepared by Yim Foon Lee (Howard Hughes Institute, University of Washington) on an Applied Biosystems 380B synthesizer. Ligations were carried out at room temperature for 3–12 h, then transformed into a lac^{-} strain for screening on Xgal indicator plates. Plasmid DNA from (leaky) lac^{-} candidates was examined by restriction analysis and by DNA sequencing of the salient region of the *lacZ* gene. Sequences of the oligonucleotide pairs were: HIV13/HIV13-A3, 5′-AGCTCTAATT$^{A}_{T}$TTAGG-GA-3′; and 5′-GATCTCCCTAAA$^{T}_{A}$AATTAG-3′; HIV201, 5′-AGCTCTAATTTTTTAGGGAAGG-3′ and 5′-GATCCCTTCC-CTAAAAAATTAG-3′; HIV201-U7, 5′-AGCTCTAATTTTTTT-GGGAAGG-3′ and 5′-GATCCCTTCCCAAAAAAATTAG-3′.

Bacterial methods

Bacterial strains and culture conditions were as described previously (Weiss *et al.*, 1988), except that minimal glucose-M63 medium was supplemented only with isoleucine and valine in addition to the strains' growth requirements. The *lacZ* plasmids were kept under selection through the presence of 1 mg ml^{-1} carbenicillin. In most experiments,

the host cells also carried pPY1011, a compatible plasmid (spectinomycin resistant, streptomycin resistant) carrying a complete *lacI*^{q} gene. Selection for this plasmid was maintained through the presence of 50 μg ml^{-1} spectinomycin in the overnight cultures used to inoculate our experimental cultures, but spectinomycin was not present during induction of *lacZ*.

Method for assaying β-galactosidase

Cell-free extracts were made by resuspending cells at a final protein concentration of 1–2 mg ml^{-1}, and sonicating for 20 s at ice-water temperature in a Braunsonic 1510 instrument; cell debris was centrifuged out at 26 000 × *g* in a Sorvall refrigerated centrifuge. Beta-galactosidase activity was measured by incubating up to 50 μl of extract with 250 μl ONPG (0.8 mg ml^{-1} in 50 mM sodium phosphate buffer, pH 7.0) at room temperature, and monitoring the accumulation of *o*-nitrophenol at A_{420} using a continuously recording Gilford 2400-2 spectrophotomer. One enzyme unit is defined as the amount of enzyme required to produce a change in absorbance of one per minute. Protein concentrations were determined by the method of Lowry *et al.* (1951) using egg albumin as a protein standard. Differential rates in enzyme units per mg of protein were computed for the period of induction (approximately one doubling in 2 mM IPTG, 2.5 mM cAMP) with or without growth limitation (as indicated).

Determination of amino acid sequence

Purification of β-galactosidase by immunoprecipitation with specific antiserum was as described in Gallant and Lindsley (1992). *N*-terminal amino acid analysis of the purified protein was performed by automated Edman degradation on Applied Biosystems 477 A pulsed-liquid phase sequencer and 470 A gas phase sequencer at University of Illinois Biotechnology Center and University of California, Riverside. Data were analysed using Applied Biosystems software for background and lag correction.

Acknowledgements

This work was supported by Grant GM13626 and from Grant GM07735 from the National Institutes of Health.

References

Atkins, J.F., Elseviers, D., and Gorini, L. (1972) Low activity of β-galactosidase in frameshift mutants of *Escherichia coli*. *Proc Natl Acad Sci USA* **69**: 1192–1195.

Atkins, J.F., Gesteland, R.F., Reid, B.R., and Anderson, C.W. (1979) Normal tRNAs promote ribosomal frameshifting. *Cell* **18:** 1119–1131.

Brierley, I., Digard, P., and Inglis, S.C. (1989) Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* **57**: 537–547.

Bruce, A.G., Atkins, J.F., and Gesteland, R.F. (1986) tRNA anticodon replacement experiments show that ribosomal frameshifting can be caused by doublet decoding. *Proc Natl Acad Sci USA* **83:** 5062–5066.

Buston, H.W., and Bishop, J. (1955) Synthetic α-amino-β-hydroxycaproic acids. *J Biol Chem* **215**: 217–220.

Cashel, M., and Rudd, K.E. (1987) *The Stringent Response.* In Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology.* Vol. 2. Neidhardt, F., Ingraham, J., Low, B., Magasanik, B., Schaechter, M., and Umbarger, E. (eds). pp. 1410–1438.

Cattaneo, R. (1989) How 'hidden' reading frames are expressed. *Trends Biochem Sci* **14**: 165–167.

Clare, J.J., Belcourt, M., and Farabaugh, P.J. (1988) Efficient translational frameshifting occurs within a conserved sequence of the overlap between two genes of a yeast Tyl transposon. *Proc Natl Acad Sci USA* **85**: 6816–6820.

Craigen, W.J., and Caskey, C.T. (1987) Translational frameshifting: where will it stop? *Cell* **50**: 1–2.

Craigen, W.J., Cool, R.G., Tate, W.P., and Caskey, C.T. (1985) Bacterial peptide chain release factors: conserved primary structure and possible frameshift regulation of release factor 2. *Proc Natl Acad Sci USA* **82**: 3616–3620.

Crow, E.L., Davis, Frances, A., and Maxfield, M.W. (1960) *Statistics Manual.* New York: Dover Publications.

Curran, J.F., and Yarus, M. (1988) Use of tRNA suppressors to probe regulation of *Escherichia coli* release factor 2. *J Mol Biol* **203:** 75–83.

Foley, D.P., Dennis, P., and Gallant, J. (1982) Mechanism of the *rel* defect in beta-galactosidase synthesis. *J Bacteriol* **145:** 641–643.

Gallant, J. (1979) Stringent Control in *E. coli. Annu Rev Genet* **13:** 393–415.

Gallant, J., and Foley D. (1980) On the causes and prevention of mistranslation. In *Ribosomes: Structure, Function and Genetics.* (ed.). Chambliss, G. *et al.* Baltimore: University Park Press, pp. 615–638.

Gallant, J., and Lindsley, D. (1992) Leftward ribosome frameshifting at a hungry codon. *J Mol Biol* **223**: 31–40.

Gallant, J., Weiss, R., Murphy, J., and Brown, M. (1985) Some puzzle of translational accuracy. In *Molecular Biology of Bacterial Growth.* Schaechter, M., Neidhardt, F.C., Ingraham, J.L., and Kjeldgaard, N.O. (eds). pp. 92–107.

Grunberg-Manago, M. (1987) Regulation of the expression of aminoacyl-tRNA synthetases and translation factors. In Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology.* Vol. 2. Neidhardt, F.C., Ingraham, J.L., Low, B., Magasanik, B., Schaechter, M., and Umbarger, E. (eds). pp. 1386–1409.

Hatfield, D., and Oroszlan, S. (1990) The *where, what, and how* of ribosomal frameshifting in retroviral protein synthesis. *Trends Biochem Sci* **15**:186–190.

Hunkapiller, M.W. (1986) Automated amino acid sequence assignment: Development of a fully automated protein sequencer using edman degradation. In *Methods in Protein Sequence Analysis.* (ed.) Walsh, K.A. pp. 367–384.

Jacks, T., Madhani, H.D., Masiarz, F.R., and Varmus, H.E. (1988a) Signals for ribosomal frameshifting in the Rous sarcoma virus *gag–pol* region. *Cell* **55**: 447–458.

Jacks, T., Power, M.D., Masiarz, F.R., Luciw, P.A., Barr, P.J., and Varmus, H.E. (1988b) Characterization of ribosomal

frameshifting in HIV-1 *gag–pol* expression. *Nature* **331**: 280–283.

Kolor, K., Lindsley, D., and Gallant, J. A. (1993) On the role of the P-site in leftward ribosome frameshifting at a hungry codon. *J Mol Biol* **230**:1–5.

Kurland, C. and Gallant, J. (1986) The secret life of the ribosome. In *Accuracy in Biology*. Galas, D. (ed). London: Chapman and Hall.

Landick, R., and Yanofsky, C. (1987) Transcription Attenuation. In Escherichia coli *and* Salmonella Typhimurium: *Cellular and Molecular Biology*. Vol. 2. (eds) Neidhardt, X., Ingraham, X., Low, X., Magasanik, X., Schaechter, X., and Umbarger, X. pp. 1276–1301.

Lindsley, D., and Gallant, J.A. (1993) On the directional specificity of ribosome frameshifting at a hungry codon. *Proc Natl Acad Sci USA*, in press.

Lowry, D.H., Rosebrough, N.J., Farr, A.L., and Randall, R. (1951) Protein measurement with the Folin phenol reagent. *J Biol Chem* **193**: 265–274.

Madhani, H.D., Jacks, T., Varmus, H.E. (1988) Signals for expression of HIV *pol* gene by ribosomal frameshifting. In *The Control of Human Retroviral Gene Expression*. Franza, R., Cullen, B., Wong-Staal, F. (eds). Cold Spring Harbor, New York: Cold Spring Harbor, Laboratory Press pp. 119–125.

Parker, J. (1989) Errors and alternatives in reading the universal genetic code. *American Society for Microbiology* **53**: 273–291.

Peter, K., Lindsley, D., Peng, L., and Gallant, J. (1992) Context rules of rightward overlapping reading. *New Biologist* **4**: 1–7.

Primakoff, P. (1981) *In vivo* role of the *relA*$^+$ gene in the regulation of the *lac* operon. *J Bacteriol* **145:** 410–416.

Primakoff, P., and Artz, S.W. (1979) Positive control of *lac* operon expression *in vitro* by guanosine 5′-diphosphate 3′-diphosphate. *Proc Natl Acad Sci USA* **76:** 1726–1730.

Ratner, L., Haseltine, W., Patarca, R., Livak, K., Starcich, B., Josephs, S., Doran, E., Rafalski, J., Whitehorn, E., Baumeister, K., Ivanoff, L., Petteway, Jr., S., Pearson, M., Lautenberger, J., Papas, T., Ghrayeb, J., Chang, N., Gallo, R., and Wong-Staal. F. (1985) Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* **313**: 277–284.

Vieira, J., and J. Messing, J. (1982) The pUC plasmids, a M13mp7 derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19:** 259.

Weiss, R., and Gallant, J. (1983) Mechanism of ribosome frameshifting during translation of the genetic code. *Nature* **302**: 389–393.

Weiss, R., and Gallant, J. (1986) Frameshifting suppression in aminoacyl-tRNA limited cells. *Genetics* **112**:727–739.

Weiss, R.B., Dunn, D.M., Atkins, J.F., and Gesteland, R.F. (1987) Slippery runs, shifty stops, backward steps, and forward hops: −2, −1, +1, +2, +5, and +6 ribosomal frameshifting. *Cold Spring Harbor Symp Quant* **52:** 687–693.

Weiss, R., Lindsley, D., Falahee, B., and Gallant, J. (1988) On the mechanism of ribosomal frameshifting at hungry codons. *J Mol Biol* **203**: 403–410.

Weiss, R.B., Dunn, D.M., Shuh, M., Atkins, J.F., and Gesteland, R.F. (1989) *E. coli* ribosomes re-phase on retroviral frameshift signals at rates ranging from 2 to 50 percent. *The New Biologist* **1**:159–169.

Wilson, W., Braddock, M., Adams, S.E., Rathjen, P.D., Kingsman, S.M., Kingsman, A.J. (1988) HIV expression strategies: ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems. *Cell* **55**: 1159–1169.

Yang, H., Heller, K., Gellert, M., and Zubay, G., (1979) Differential sensitivity of gene expression *in vitro* to inhibitation of DNA gyrase. *Proc Natl Sci USA* **76:** 3304–3308.