<table>
<thead>
<tr>
<th>Title</th>
<th>Two groups of phenylalanine biosynthetic operon leader peptides genes: a high level of apparently incidental frameshifting in decoding <em>Escherichia coli</em> pheL.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Gurvich, Olga L.; Näsvall, S. Joakim; Baranov, Pavel V.; Björk, Glenn R.; Atkins, John F.</td>
</tr>
<tr>
<td>Publication date</td>
<td>2011</td>
</tr>
<tr>
<td>Type of publication</td>
<td>Article (peer-reviewed)</td>
</tr>
</tbody>
</table>
| Link to publisher's version | https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkq1272  
http://dx.doi.org/10.1093/nar/gkq1272  
Access to the full text of the published version may require a subscription. |
| Rights      | © 2010, the Authors. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited. |
| Item downloaded from | http://hdl.handle.net/10468/5025                                                                                                                                                               |

Downloaded on 2018-12-18T04:59:45Z
Two groups of phenylalanine biosynthetic operon leader peptides genes: a high level of apparently incidental frameshifting in decoding *Escherichia coli* pheL

Olga L. Gurvich1, S. Joakim Näsvall2, Pavel V. Baranov3, Glenn R. Björk2 and John F. Atkins1,4,*

1Department of Human Genetics, University of Utah, Salt Lake City, UT, 84112-5330, USA, 2Department of Molecular Biology, Umeå University, S-901 87 Umeå, Sweden, 3Department of Biochemistry and 4BioSciences Institute, University College Cork, Cork, Ireland

Received July 5, 2010; Revised and Accepted November 23, 2010

ABSTRACT

The bacterial pheL gene encodes the leader peptide for the phenylalanine biosynthetic operon. Translation of pheL mRNA controls transcription attenuation and, consequently, expression of the downstream pheA gene. Fifty-three unique pheL genes have been identified in sequenced genomes of the gamma subdivision. There are two groups of pheL genes, both of which are short and contain a run(s) of phenylalanine codons at an internal position. One group is somewhat diverse and features different termination and 5′-flanking codons. The other group, mostly restricted to Enterobacteria and including *Escherichia coli* pheL, has a conserved nucleotide sequence that ends with UUC_CCC_UGA. When these three codons in *E. coli* pheL mRNA are in the ribosomal E-, P- and A-sites, there is an unusually high level, 15%, of +1 ribosomal frameshifting due to features of the nascent peptide sequence that include the penultimate phenylalanine. This level increases to 60% with a natural, heterologous, nascent peptide stimulator. Nevertheless, studies with different tRNAPro mutants in *Salmonella enterica* suggest that frameshifting at the end of pheL does not influence expression of the downstream pheA. This finding of incidental, rather than utilized, frameshifting is cautionary for other studies of programmed frameshifting.

INTRODUCTION

The shift-prone sequences at which mRNA:tRNA realignment occurs during programmed frameshifting, are generally avoided in highly expressed genes except where the resultant frameshifting is utilized for gene expression. In poorly expressed genes, where deleterious effects at the protein product or mRNA structural/stability levels are minimal, such sequences do not seem to be rare—at least as extrapolated from the occurrences in *Escherichia coli* of the −1 and +1 shift-prone sequences, A_AAA_AAG, CCC_UGA, AGA_AGA and AGG_AGG (1,2). However, when specific frameshifting is selected for gene expression, there are generally stimulatory signals that elevate the level of frameshifting at the shift site. These stimulatory, or recoding signals, are often particular mRNA sequence 3′ of the shift site that elevate the level of frameshifting at the shift site. Other characterized frameshift stimulatory signals are specific 5′-mRNA sequences. Bioinformatic evidence indicates that particular nascent peptide sequences can promote utilized frameshifting (3). Furthermore, a nascent peptide stimulator for the programmed bypassing of 50 nt in decoding phage T4 gene 60 acts by causing peptidyl-tRNA codon:anticodon dissociation. Not surprisingly this nascent peptide can induce frameshifting on synthetic constructs (4).

Specific nascent peptide sequences that act within the exit tunnel of the ribosome can trigger changes within the peptidyl transferase centre (PTC) and/or induce ribosomal pausing. Mediation of these effects, in some cases,
involves exogenous factors. Inducible expression of erythromycin-resistance genes relies on ribosomal stalling, which is dependent on drug binding in the exit tunnel (5); translational arrest during SecM synthesis, which regulates translation of the co-transcribed SecA, is responsive to the secretion status of the cell (6); ribosomal stalling during translation of TnaC, which regulates expression of tryptophanase operon, is sensitive to the level of free tryptophan (7,8). The mechanism of action of several nascent peptide signals, including that of SecM, has been studied in detail (9–12).

Stop codons are slow to decode and UGA is, overall, the least efficient terminator. The extent of the pause at a termination codon is influenced by the identity of the 3’-adjacent base (13) and when the stop codon is preceded by a proline codon, its decoding is especially slow, perhaps because the C-terminal Pro residue facilitates trans- to cis-isomerization of the final peptide bond, thereby inhibiting termination (14). A C-terminal proline has a variety of translational consequences (15–17); for instance, it is a key feature of the TnaC nascent peptide-mediated effect (18). Additionally, it can facilitate frameshifting. When a slow-to-decode stop codon in the ribosomal A-site is paired with a P-site peptidyl-tRNA that has good potential for re-pairing to mRNA in an overlapping frame, the combination is especially shift-prone, hence the term ‘shifty stop’ (19,20). Thus, not surprisingly, the sequence CCC_UGA is especially +1 shift-prone (21–23). Frameshifting at CCC_UGA is utilized for expression of antizyme in some eukaryotes (24) and in the expression of the gene for a major tail component in several Listeria phages (25,26). Codons other than UGA are also used in the A-site; some phages use +1 frameshifting at the CCC_UAA, i.e. the Lactobacillus phage Q54 (27) and Bacillus phage SPP1 (28), while in a Listeria phage (26) and the E. coli cheA gene (29), the slow-to-decode UGA stop codon is substituted by a sense codon whose tRNA is severely limiting.

The frameshifting that occurs at one of the most shift-prone sites in E. coli, A_AAA_AAG (30,31), involves a weak interaction of anticodon base 34 of the plentiful sole tRNA with the third codon bases in the A-site (32,33). More striking is the importance of lack of full Watson–Crick codon complementarity of the P-site tRNA during Saccharomyces cerevisiae Ty3 frameshifting (34,35). The cognate tRNA for CCC (36,37) is also often rare, even in E. coli and Salmonella that do not have a low GC content. To a small extent, this increases the chance of acceptance at CCC-containing ribosomal A-sites, of a more abundant near-cognate proline tRNA with elevated frameshifting consequences. Mutants of tRNAPro selected for enhanced frameshifting at CCC led to the isolation of partially debilitating mutants of the CCC-decoding isoacceptor that substantially increased the chance of CCC being decoded by a near-cognate tRNA. After its transfer to the P-site, a near-cognate tRNA is more prone than a cognate tRNA to dissociate from mRNA and realign before re-pairing to mRNA (38,39). Some of these mutants are important for the present work.

Of the 19 E. coli genes that terminate with CCC_UGA, one, pheL, shows a dramatically higher level of frameshifting, 15% in a relA strain, than any of the others (1). Although the level of frameshifting is slightly higher in relA strains than in WT stringent cells, the effect is small (2,40). This frameshift efficiency is much more than can be accounted for by just CCC_UGA, which is generally not more than 2% (1,21).

pheL is co-transcribed with the 3’ pheA whose product catalyses the first two steps of phenylalanine biosynthesis from chorismate (41). Translation of the short pheL leader peptide-encoding sequence, in particular the run(s) of phenylalanine codons within it, governs transcription attenuation and, consequently, expression of the downstream pheA. When Phe is limiting, the attenuation mechanism (42–44), Figure 1, results in greatly elevated pheA mRNA synthesis and expression. However, there is a basal level of expression even when cells are replete for Phe. With abundant Phe, this basal level is ∼10% of the level under Phe starvation conditions. The basal level is dependent on the efficiency of ribosome release from the pheL UGA stop codon—release factor mutants can decrease the basal level of pheA expression (45). The possibility that frameshifting at CCC_UGA may influence basal expression of pheA has not been addressed and is unknown in biosynthetic operon expression.

Frameshifting at the CCC_UGA of pheL gene results in the synthesis of an extra product. This product can potentially have a separate function. However, since frameshifting allows ribosomes to travel past the stop codon of pheL and interfere with formation of secondary structures, it could influence attenuation of transcription and pheA expression. If so, it would be the first case where the significance of frameshifting is in mediating ribosomal progression to a novel location instead of in synthesizing an extra protein product (46). [The only specific suggestion for such a concept (47) has not been experimentally investigated.]

**MATERIALS AND METHODS**

**Plasmid constructions**

The pheL and pdxH gene sequences were amplified by PCR from E. coli genomic DNA or from previous constructs (1) using primers with appropriate overhanging restriction sites. All subsequent nucleotide changes were introduced using PCR with complementary oligonucleotides carrying appropriate changes as primers. All plasmid constructions were confirmed by DNA sequencing.

For analysis of frameshift stimulators, sequences were cloned into the BamH1/EcoRI sites of pGPHM57 (4) in-between GST and maltose binding protein (MBP), so that MBP is in the +1 frame relative to GST. Construction of HLLH, HLH5 and HLH3 was performed by a two-step PCR. In the first step, DNA sequences corresponding to pheL, 5’ and 3’ of pdxH, were amplified separately. To amplify the pheL sequence, the pheL/pdxH chimera primers HLa(GGAAGATTTAGCTCTTTCGAAATGAAAACAC ATACGTTTTTCT) and HLb(GGAAGATTTTGCGATCTTTAGCGTCTTAAAGGCCCGCGATTG) were used
for the HLLH construct; HLa and LH2(GCATCTTTTC
AGGGGAGG) were used for HLH5; HL1(CCTTCCCC
TGAAAAGATGC) and HLb were used for HLH3. The
5’ pdxH sequence was amplified using the
pdxH primer CT4t(ATAGGATCCAGCGTGAAAATGATGCGTGG
AAG) and the pheL/pdxH chimera primers: LHb(CGGT
ATGTGTTTCATTGCAAGACGATCAATC) for
HLLH and HLH5 constructs and HL2(CTCCCATTCA
GGGTGCAAG) for the HLH3 construct. The 3’
pdxH sequence was amplified using the pdxH primer
CT4b(TAT GAATTCGAGTACCAGCGATTAAAGCAAG) and
the pheL/pdxH chimera primers LHa(CGGGGGGCCT
TTAAAGATGCAAAAATCTTGC) for HLLH and
HLH5 constructs and HL2(CCTCCCATCTGAAAAGATGC) for
HLH3. DNA fragments from the first PCR step were
gel-purified and a second PCR step was performed using
CT4a and CT4b primers and the appropriate mix of frag-
ments obtained in the first step. Changes were further
introduced into the HLH5 construct to make MP, PM
and constructs with separate codon mutations.

Construct CT5+ was made by cloning a synthetic oligo-
nucleotide encompassing pheL sequence from the AUG
codon to the CCC_UGAA stop, into the HindIII/ApaI
sites between the GST-encoding sequence and lacZ of
psKAGS (48), so that lacZ was in the +1 frame relative
to GST (frameshift reporter); for CT5IF, the cloned
pheL sequence ended with CCC_TGG (to eliminate the UGA
stop codon) and was placed in pSKAGS so that lacZ was
in the 0-frame relative to the GST in-frame control).

For attenuation control experiments, the regulatory
part of the phenylalanine operon was amplified by PCR
from E. coli genomic DNA using the primers ATACTGC
AGTTGACAGCGTGAAAACAGTAC and GGGATA
TCGCATATGTGTCATAGTGTTGCCTTTTTGTTAT
C. The PCR product was digested with NdeI and PstI
restriction enzymes and cloned into corresponding sites
of the pLA2 vector, removing the P araB promoter (49).

The resulting construct pPZWT has
lacZ expression
driven from the phenylalanine promoter and the initiation
codon of pheA. Its expression is under
pheL control and
the regulatory secondary structures (Figure 7A).

Pulse-chase analysis in E. coli

Overnight cultures of DH5alpha E. coli strains expressing
the appropriate construct were grown in MOPS-glucose
(50) containing 100 μg/ml ampicillin and all amino acids
was used as selectable marker for strain constructions. pKD3 inserted 56 bp downstream of the chloramphenicol-resistance cassette derived from plasmid genes encoding tRNAProS. enterica recombineering (57). The (omitted from the media used to generate material for the pheR (56). The (55), which also causes +1 frameshifting at CCC codons were diluted 1:50 in 300 μl of 150 mM Tris–HCl; pH 7.2) and heated at 95°C for 5 min. Aliquots of 5 μl were loaded on 4–12% NuPAGE Gels (Invitrogen Inc.) and electrophoresed, under conditions recommended by the manufacturer, in MOPS–SDS buffer (Invitrogen Inc.). Gels were exposed overnight and visualized with a Molecular Dynamics PhosphorImager. The amounts of termination and frameshift products were quantified by ImageQuant. The frameshifting efficiency was estimated as the ratio of the amount of frameshift product to the sum of the termination and frameshift products. At least three independent experiments were performed with each construct on separate days and the frameshifting levels presented in Figures 4–6 are the average values obtained, with the error bars indicating standard deviations.

Salmonella strains, genetic procedures and growth conditions

All strains (Table 1) are derivatives of S. enterica serovar Typhimurium strain LT2. As solid medium, LA (10 g of Trypticase peptone, 5 g of yeast extract, 5 g of NaCl and 15 g of agar per litre) was used. LB (5 l) and rich MOPS (50) were used as liquid media. Antibiotics were used at the following concentrations: carbenicillin (Cb): 50 mg/l; kanamycin (Km): 100 mg/l; chloramphenicol (Cm): 30 μg/ml.

To transfer chromosomal markers or plasmids between strains of S. enterica, transductions were performed as described elsewhere (52) using a derivative of bacteriophage P22 containing the mutations HT105/I (53) and int-201 (54).

proM2219 was isolated in a selection for frameshift suppressor derivatives of tRNAProemoSUGG that cause +1 frameshifting at CCC codons (38). It changes a G at position 31 in tRNAProemoSUGG to an A, resulting in an A–C mismatch instead of a G–C base pair between positions 31 and 39 in the lower part of the anticodon stem. zhe-2533:cat is a chloramphenicol-resistance cassette derived from plasmid pKD3 inserted 56 bp. downstream of the proM gene. It was used as selectable marker for strain constructions. ΔproL is a deletion of the gene encoding tRNAProSUGG (55), which also causes +1 frameshifting at CCC codons (56). The pheR mutant was constructed using λ-red recombineering (57). The pheR gene (one of the two S. enterica genes encoding tRNAProGAA) was replaced by the kanamycin-resistance cassette from plasmid pKD4, to yield pheR::kan allele. To enable introduction of plasmid pPHWT (KmR) into pheR strains, the kanamycin-resistance cassette was removed by expression of Flp recombinase from plasmid pCP20, resulting in the pheR::frt allele (and the simultaneous conversion of zhe-2533::cat into zhe-2533::frt).

β-Galactosidase assays

Cultures for β-galactosidase assays were grown to mid-exponential phase (OD600~0.4–0.5) in a Beckman Coulter DU 730 spectrophotometer) in rich MOPS medium at 37°C and assayed (58), using the alternative protocol with SDS and chloroform instead of toluene to

<table>
<thead>
<tr>
<th>Table 1. Strains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strain</td>
</tr>
<tr>
<td>DHHspha</td>
</tr>
<tr>
<td>GT8060</td>
</tr>
<tr>
<td>GT8001</td>
</tr>
<tr>
<td>GT8066</td>
</tr>
</tbody>
</table>

DH5alpha is the only E. coli strain used in this study. All other strains listed in the table are S. enterica strains. *Mutations that are most experimentally relevant are in bold.
permeabilize the cells. The frameshifting efficiencies reported in Figure 7A were calculated as below:

\[
FS(\%) = 100 \times \frac{\text{beta gal. activity from CT5}}{\text{beta gal. activity from CT5IF}}
\]

RESULTS

Conservation of pheL and the frameshift site

Since the pheL gene is very short, it is not annotated in the majority of sequenced genomes. To identify pheL genes, beyond those previously characterized in E. coli, S. typhi, Klebsiella pneumoniae, Yersinia pestis, Erwinia carotovora, Serratia oneidensis and Vibrionales (59), we extracted 200-nt upstream of annotated pheA genes in all sequenced bacteria. These were examined for the presence of at least five phenylalanine codons (where at least two of them are adjacent) occurring in-frame within a 60-nt region with no stop codons in the same translational phase. The intergenic regions between the resulting putative pheL genes and pheA genes were checked for the presence of a transcription termination signal and for potential termination/antitermination secondary structures. Those pheL genes that contain such signals and structures were classified as true positives. They were identified in a number of enteric bacteria including Escherichia, Shigella, Salmonella, Erwinia, Klebsiella, Dickeya, Enterobacter, Citrobacter and Serratia species (Figure 2). The pheL gene was also found in Yersinia species; however, the

Figure 2. The pheL sequences from different bacteria. When the nucleotide sequence is identical in several strains or species (e.g. E. coli K-12 and S. flexneri 2a) it is included only once in the alignment. Phenylalanine codons are highlighted in yellow; TGA stop codons—blue; TAA and TAG stop codons—in red; and Proline codons—in green. The sequence sources are in the supplementary material. (A) Alignment of the pheL sequences terminating with CCC_UGA. The position of the /C0 10 Proline codon that has a 2-fold effect on frameshifting, is indicated by black shading. For each sequence, continuation of translation in the +1 frame after CCC_UGA is described in terms of where it terminates relative to the frameshift site, termination hairpin or attenuation site (run of ‘T’s) or whether it proceeds past the pheA initiation codon and in which frame. (B) Alignment of the pheL sequences that do not end with CCC_UGA.
region between pheL and pheA is disrupted by an insertion sequence, IS1541, and the transcription attenuation mechanism might not be utilized to regulate pheA levels in Yersinia. In fact, in Y. pestis, the pheA gene itself is interrupted by another mobile element, IS100, and is unlikely to be functional. The pheL gene was also found in non-enteric bacteria, Vibrio and Shewanella species, of the gamma proteobacteria class. Only a few bacteria outside the proteobacteria class, Bacteriodes and Desulfobacterium autotrophicum, potentially use a pheL gene to regulate pheA expression.

All pheL genes can be separated into two groups. One group, limited to enterobacteria and Shewanella amazonesis, has highly conserved nucleotide sequence and all end with CCC UGA (Figure 2A). In the other group, pheL sequences do not end with CCC UGA and exhibit a lesser degree of conservation (Figure 2B). Plus-one frameshifting at CCC UGA at the end of pheL would result in synthesis of a longer peptide. The amino acid sequence of this peptide, however, is not conserved among different enteric bacteria. The length of the peptide also differs. In E. carotovora [Pectobacter carotovorum, (60)] the stop codon in the +1 frame is only six codons 3′ of the shift site, on the 3′-side of the pause stem loop. In most other bacteria, the stop codon is located downstream of the pause structure (Figure 1B). In Salmonella species, the +1 frame stop codon overlaps with the AUG initiation codon of pheA. In E. coli, it is located 80 nt downstream of the pheA initiation codon. Moreover, in K. pneumoniae and E. sakazakii, +1 frameshifting on CCC UGA would allow ribosomes to enter the pheA ORF. Since there are no stop codons separating the two genes in this frame, frameshifting in these bacteria would result in synthesis of a pheL-pheA fusion product. Thus, although the +1 frameshifting on CCC UGA at the end of pheL is likely conserved in enteric bacteria, the length and the amino acid sequence of the synthesized peptide is not. Therefore, the peptide is unlikely to have a separate function.

Features in pheL that are responsible for frameshifting

The level of frameshifting in pheL gene decoding is much higher than in the 18 other E. coli genes that also end with CCC UGA (1). An investigation of the features of pheL mRNA, and its encoded product responsible for the elevated frameshift levels, was undertaken. The frameshift reporter utilized encodes GST (glutathione-S-transferase)-MBP fusion protein (1). The pheL sequence to be tested was cloned so that its 5′ part is in-frame with GST, while MBP was placed in the +1 frame relative to GST. Thus, translation termination at CCC UGA results in GST-PheL synthesis. If, however, +1 frameshifting at CCC UGA occurs, then a GST-PheL-MBP fusion protein is synthesized. Frameshifting efficiencies were determined from expression levels of both proteins assayed by [35S]-methionine pulse-chase labelling of total protein.

The possibility that sequence 3′ of CCC UGA in E. coli pheL mRNA stimulates +1 frameshifting was addressed first. The original construct contained 33 nt of the natural sequence 3′ of CCC UGA (1). Here (see Figure 3A), the 3′-sequence was shortened to only 9 nt 3′ of the CCC UGA or extended to either 66 nt (including the RNA structures, but excluding the transcription termination site) or to 102 nt (including the start codon of pheA) 3′ of the CCC UGA. No major change in frameshift efficiency was detected with the first two constructs (Figure 3B, first and second lane). With the construct that has 102 nt 3′ of CCC UGA, the band corresponding to the frameshift product was faint (Figure 3B, third lane). Theoretically, sequences downstream of the CCC UGA can preclude frameshifting. Alternatively, since the transcription termination site is located prior to the MBP sequence, lower levels of the full-length GST–pheL–MBP mRNA can be expected with this construct in the presence of phenylalanine (with high levels of the mRNA corresponding to GST-pheL). Consequently, a lower amount of the frameshift product with this construct can be explained not by lower frameshift levels per se, but by the absence of full-length mRNA containing sequences downstream of the frameshift site. To distinguish between the possibilities, the growth media was depleted of phenylalanine (to decrease transcription attenuation).

![Figure 3](image-url)

Figure 3. Influence of sequences 3′ of CCC UGA frameshift site on efficiency of frameshifting in pheL. (A) Scheme of the reporter constructs. Arrows indicate the positions of the 3′-ends of sequences cloned in pHM57 relative to regulatory structures and sequences in pheL-pheA. Numbers correspond to the number of nucleotides 3′ of CCC UGA. (B) SDS-PAGE of protein products produced from pheL constructs with different 3′-lengths. The termination, GST-PheL, and the frameshift, GST-PheL-MBP, products are indicated by arrows. The (−) lane contains protein products from an uninduced control. (C) Effect of phenylalanine depletion on production of the frameshift product from the constructs with (102) and without (66) the transcription attenuation site.
This resulted in an increase of the frameshift product from the construct with the transcription termination site (Figure 3C, lanes 1 and 2). Phenylalanine depletion did not increase the amount of the frameshift product from the construct without the transcription termination site (Figure 3C, lanes 3 and 4). Thus, lower levels of GST–PheL–MBP frameshift product, from the construct containing the transcription attenuation site, are most likely due to a lower amount of the full-length mRNA produced. Therefore, most likely the 3′-context of the pheL mRNA does not influence, either positively or negatively, frameshifting on CCC_UGA.

To test this presumption, an additional experiment was performed. Sequence 5′ or 3′ of the pheL CCC_UGA was exchanged for the corresponding sequence 5′ or 3′ of CCC_UGA of the pdxH gene (Figure 4A). The pdxH coding sequence also ends with CCC_UGA, but a product derived from frameshifting during expression of this gene was not detected [(1); Figure 4B, lane 2]. First, the sequence from the pheL construct with 66 nt downstream of the frameshift site was moved in place of the CCC_UGA in the pdxH-containing construct (Figure 4A, third construct). This resulted in the 5′-pdxH–pheL–CCC_UGA–pdxH-3′ chimera construct (HLLH). The frameshifting efficiency on the CCC_UGA, with this context, was comparable with that exhibited by the original pheL construct (Figure 4B and C, lane 3). Next, either the 5′ or the 3′ pheL sequence was removed from the above chimera construct. The frameshifting efficiency with the construct 5′-pdxH–pheL–CCC_UGA–pdxH-3′ (HLH5; Figure 4B and C, lane 4) was comparable with that obtained for the wild-type pheL. The frameshifting efficiency, however, dropped to a marginal level with the construct 5′-pdxH–CCC_UGA–pheL–pdxH-3′ (HLH3; Figure 4B and C, lane 5). This confirms that the main stimulatory signals for frameshifting are located 5′ of the sequence CCC_UGA in pheL, with the identity of the 3′-sequence being unimportant.

The 5′-stimulator can act at the nucleotide level and/or at the level of the encoded amino acids in the nascent peptide. To distinguish between these possibilities, two types of experiments were performed with the HLH5 construct described above. In one set of experiments, single nucleotide insertions and deletions were introduced in the pheL sequence so that ribosomes translating the 5′-region were routed in and out of different frames. The synthesized peptide thus had a completely different amino acid sequence, but the nucleotide sequence was maximally preserved. Two series of constructs were made for this purpose: PM0-3 and PM0-3 (Figure 5A). In the PM series, a nucleotide was inserted at the start codon of pheL and a compensatory nucleotide deletion was made prior to the sequence corresponding to the CCC_UGA frameshift site. In the MP series, 4 nt, ATGA, were deleted at the beginning of the pheL and a compensatory insertion prior to the frameshift site routed ribosomes back into the zero frame. Numbers 0 through 3 indicate the number of wild-type pheL codons in a construct immediately 5′ of the CCC_UGA. Constructs PM0 and MP0, in which ribosomes were routed back in-frame just prior to CCC_UGA, showed a significant drop in frameshifting efficiency, to 10 and 30% of WT pheL, respectively (Figure 5B and C). The frameshifting efficiency gradually increased with the addition of WT codons upstream of the CCC_UGA. Constructs MP3 and PM3, which had three wild-type codons upstream of the CCC_UGA, exhibited wild-type levels of frameshifting.

In the other set of mutants, codons throughout the pheL ORF in the HLH5 reporter were altered separately, or in blocks, to either synonymous or non-synonymous codons (Figure 6A). None of the changes to synonymous codons affected frameshifting efficiency (Figure 6B–G), while non-synonymous changes of a few codons had a much stronger effect. The identity of the amino acid, phenylalanine, encoded 5′ adjacent to CCC_UGA is crucial for the observed frameshifting levels (Figure 6D and E). Only changing the third position of this phenylalanine codon, UUC, to the synonymous UUU did not alter the frameshifting level, whereas changing it to UUG (leucine) reduced frameshifting to <20% of the WT level (as did changing the UUC to CUG, arginine). Changing just the second nucleotide of UUC so the codon became UAC (tyrosine) reduced the frameshifting level to ~50% of WT (a similar level to changing the UUC to a different tyrosine codon (UAU)). Changing just the first nucleotide of UUC so the codon became CUC (leucine) also drastically reduced frameshifting. Thus, frameshifting was unaffected by minimal changes that preserved codon meaning but reduced when the identity of the encoded amino acid was altered. Changing the ACC (threonine) codon at the −2 position with respect to the CCC_UGA

![Figure 4](image-url)

**Figure 4.** Effect of pheL and pdxH 5′ and 3′ sequences on frameshifting at CCC_UGA. (A) Scheme of the analysed constructs. (B) SDS-PAGE of proteins expressed from pheL, pdxH, HLLH, HLH5 and HLH3 constructs. The termination, GST-PheL, and the frameshift, GST–PheL–MBP, products are indicated by arrows. (C) Quantitation of frameshifting efficiency during translation of pheL, pdxH, HLLH, HLH5 and HLH3 constructs from (B). The level of frameshifting obtained with the WT pheL sequence is set to 100% of wild-type.
to the synonymous codon ACA did not affect frameshifting levels (Figure 6B and C), but the non-synonymous alteration to a glycine (GGC) codon resulted in a 50% increase in frameshifting (Figure 6B and C).

A single nucleotide change in the 10th codon of the shift site, CCG to CCU (both proline) did not affect the frameshifting level, whereas changing the CCG to ACG (threonine) or CUG (serine) or any of the other four non-proline codons tested, led to a 2-fold diminution of frameshifting. The other changes that resulted in alteration of the encoded amino acid also resulted in a similar reduction of frameshifting (Figure 6F and G).

[The role of the proline codon at position 10 was investigated further by changing it to a different proline codon, CCC, in the MP0 and MP1 constructs, that encode a scrambled peptide sequence (see above and Figure 5A). No significant change in frameshifting efficiency was observed (Figure 5B and D).]

Taken together, these results provide suggestive evidence that the frameshift stimulator in the *pheL* gene is mainly composed of the amino acids encoded in the mRNA sequence. A different peptide stimulator acts in the 50% efficient translational bypassing of the 50 non-coding nucleotides between codons 46 and 47 of bacteriophage T4 mRNA (61–63). The nascent peptide stimulator acts to promote dissociation of peptidyl-tRNA anticodon pairing to codon 46 (4,64). Although in bypassing the anticodon re-pairing to
mRNA is not at an overlapping codon, and in gene 60 bypassing the sequence of the coding gap is not scanned, the key feature of P-site codon:anticodon dissociation is shared by at least most cases of programmed frameshifting. Accordingly, we tested whether the gene nascent peptide would stimulate +1 frameshifting at CCC_{UGA}. The 5'_{0} pheL sequence in the HLH5 construct was substituted with the 135 nt (45 codons) preceding the

![Figure 6](image_url)

Figure 6. Analysis of the peptide stimulator by synonymous and non-synonymous mutations in pheL. In (C), (E) and (G), white bars correspond to synonymous and grey to non-synonymous mutations. In (B), (D) and (F) the (−) lane contains protein products from an uninduced control. The termination, GST–PheL, and the frameshift, GST–PheL–MBP, products are indicated by arrows. (A) Peptide mutations (in lower case letters). The wild-type pheL sequence is in capital letters. Negative numbers above the codons of the pheL peptide represent the position of a codon relative to the CCC_{UGA}. Codons mutated simultaneously in a single construct are under- (or over-) lined. Nucleotides differing from the wild-type sequence are underlined. Numbers in brackets indicate the number of a construct and correspond to the numbers of constructs analysed in (B) and (C). (B) SDS–PAGE of protein products from pulse-chase experiment with the ‘numbered’ constructs from (A). (C) Relative quantitation of frameshifting efficiencies in the constructs analysed in (B). All frameshifting efficiencies are relative to that of WT pheL, which is set at 100%. (D) SDS–PAGE of proteins from pulse-chase analysis of constructs with changes in the −1 codon of pheL. Grey arrow indicates a putative tmRNA-tagged termination product discussed in the text. (E) Relative quantitation of frameshifting efficiencies in the constructs analysed in (D). (F) SDS–PAGE of proteins produced from the constructs with changes in the −10 codon of pheL. (G) Relative quantitation of frameshifting efficiencies in the constructs analysed in (F).
codon 46 GGA take-off site of gene 60. Remarkably, ~60% +1 frameshifting was observed on CCC_UGA with the gene 60 nascent peptide stimulator (Figure 6D, right-most lane).

**Role of frameshifting at the end of pheL**

The amino acid sequence encoded by the new frame after the +1 ribosomal frameshift is not conserved (Figure 2A). Sequence comparisons, therefore, do not provide support for a functional role for the protein product. Frameshifting at the end of pheL might influence the levels of pheA expression. Ribosomes that escape the pheL stop codon and continue translating the pheL–pheA mRNA, would preclude formation of the attenuator structure and might contribute to antitermination of pheA transcription.

To test this, we availed of mutants that showed higher or lower frameshifting than wild-type at CCC_UGA. In S. enterica, +1 frameshifting at CCC codons is known to be enhanced by various mutations affecting tRNA\textsubscript{Pro}\textsubscript{GGG} (sufB2, ΔproL) or tRNA\textsubscript{Pro}\textsubscript{GGG} (sufA6). Most, or all, of the frameshifting in these mutants occurs when the structurally normal near-cognate peptidyl-tRNA, tRNA\textsubscript{Pro}\textsubscript{cmo5UGG}, occupies the P-site of the ribosome with a CCC codon (39,56,65). In addition, a number of mutants of tRNA\textsubscript{Pro}\textsubscript{cmo5UGG} (proM) have recently been isolated as +1 frameshift mutant suppressors acting at CCC codons (38).

We used plasmids CT5+ and CT5IF, which carry gst–pheL–lacZ fusions, to determine the frequencies of +1 frameshifting in the various mutants. In these plasmids, lacZ is in the +1 frame (CT5+), or the 0 frame (CT5IF), compared with gst–pheL, and frameshifting was monitored by β-galactosidase assays. The ΔproL mutants caused a 4.4-fold increase in frameshifting compared with wild-type (1.9% in wt, versus 8.4% in ΔproL), and the most efficient proM mutant, proM2219, increased frameshifting 2.5-fold to 4.7% (Figure 7A). The combination of proM2219 and ΔproL led to a 14.5-fold increase in frameshifting (27.5%) as the mutant tRNA\textsubscript{Pro}\textsubscript{cmo5UGG} did not have to compete with any other tRNA for CCC-codons. Other frameshift suppressor mutants (sufA6, sufB2) were tested and caused frameshifting at about the same level as the ΔproL mutant. A cmoB2 mutant, known to decrease frameshifting at other CCC_UGA sites, did not decrease frameshifting below the wild-type level (data not shown).

To test the functional effect of frameshifting at the pheL CCC_UGA on attenuation control, a new reporter was constructed. The nucleotide sequence encompassing the promoter region, pheL and the pheL–pheA intergenic region, including the AUG start codon of pheA, was cloned upstream of lacZ in the pLA2 vector (49). The lacZ gene was placed in-frame with the pheA initiation AUG. Therefore, in this construct, pPZWT, β-galactosidase expression was driven by the phenylalanine operon promoter and was under similar control (such as pheL regulatory RNA structures and an initiation stimulatory Shine–Dalgarno) as endogenous pheA.

Whether wild-type cells are grown on minimal media or minimal media supplemented with the aromatic amino acids, the levels of pheA expression remain essentially the same (41). The pheA gene can be 2- to 20-fold de-repressed by phenylalanine starvation in chemostat conditions and/or with specific mutants (66,67), such as those that cause decreased amount or function of Phe-tRNA\textsubscript{Phe}. One such mutant is miaA, which lacks the ms\textsuperscript{3}(o)\textsubscript{A}37 modification present in tRNA\textsubscript{Phe} (68). Another is pheR, which was first thought to be a repressor protein, but was later shown to lack one of the two genes encoding tRNA\textsubscript{Phe} (69). In our hands, the miaA1 mutation did not cause any de-repression of β-galactosidase activity from the pPHWT plasmid (data not shown), whereas with
DISCUSSION

pheL sequence conservation

Nearly half of the identified pheL coding sequences terminate with UGA, and in these there is an impressive, although not absolute correlation, with both phenylalanine and proline codons being encoded by the penultimate and 5′-adjacent codons to this stop codon. This contrasts with those that terminate with UAG or UAA. The 19 sequences in Figure 2A are highly related while the majority of those in Figure 2B are quite diverse. Some role for poor termination in those depicted in Figure 2A likely awaits discovery. It is tempting to deduce that selection has not only been acting at the amino acid level. In all 19 pheL genes in Figure 2A, the nucleotide sequence is UUC CCC UGA whereas in all other positions, except for the initiator AUG and regulatory phenylalanine codons, variations do occur. Our results obtained with S. enterica mutants, however, argue against selection of this sequence for frameshifting purposes.

While the length of pheL and identity of most of its codons are variable, as expected, the regulatory phenylalanine codons are substantially conserved (Figure 2). A proline codon commonly occurs in the 15th (+ or −1) position and may, via a pausing effect, also be important for the regulation. However, this region is key for formation of both antitermination and pause-structure hairpins and the ‘C’s of the proline codons are complementary to the downstream ‘G’s. In the few sequences that do not have proline codons in those positions, the nucleotide sequence is pyrimidine rich, probably to allow the base pairing with the ‘G’s. Changes to purine nucleotides in this region are highly unlikely, because they would require compensatory changes on the other sides of both antitermination and pause stems and those requiring further compensatory changes in the terminator hairpin.

pheL and tnaC

Despite the obvious differences between the biosynthesis of phenylalanine and catabolism of tryptophan, some comparison between the present pheL work and tnaC is merited. Although C-terminal proline is important for E. coli tnaC expression, it is also just semi-conserved (18). In Proteus vulgaris, there are two extra codons 3′ of the proline codon before the termination codon and, yet, pausing still occurs there. Among the 34 sequences listed in Figure 2B, 20 have a proline codon close to the same position in relation to the initiation codon as do the 19 in Figure 2A that terminate with UUC CCC UGA, although several have extra codons after the proline codon. Of those, seven unique but related sequences from the Shewanella species, have tandem proline codons. Nevertheless, there are no tandem proline codons in any of the 23 sequences when a proline codon is 5′-adjacent to a stop codon. It is tempting to infer similarities to the pausing location in Proteus vulgaris tnaC. [An Asp, eight residues N-terminal of the proline residue, is important for tnaC pausing (18). There is no Asp codon in pheL, and, in the great majority of the sequences, there is a Phe codon at the equivalent position to the Asp codon in tnaC, or else adjacent to it.] E. coli tnaC terminates with CCU UGA rather than CCC UGA in E. coli pheL (the 3′-proline codon in 9 of the 34 sequences in Figure 2B is not CCC by contrast to the likely frameshift-prone CCC-containing sequences in Figure 2A). The tRNA that decodes CCC also decodes CCU and a possible role for frameshifting in expression of E. coli tnaC, was considered at an early stage in the analysis of that operon, but dismissed (70). In the pheL context when CCC UGA is mutated to CCU UGA, the frameshifting level dropped by two-thirds (1). More importantly, the N-terminal amino acid adjacent to Pro in TnaC is Arg, and placing an arginine codon at the corresponding position of pheL virtually abolished frameshifting. Thus, the context of the termination codon in TnaC makes frameshifting in this gene highly unlikely. More than a decade of elegant work has revealed the nature of tnaC expression. By contrast, many aspects of pheL expression, such as possible mRNA cleavage, remain to be studied and are outside the scope of the present work.

Non-standard translational events other than frameshifting

Inefficient translational termination context in the pheL gene can result not only in frameshifting but other non-standard translational events. Stop-codon readthrough, measured by pulse-chase analysis with the GST–pheL–MBP construct, occurs at an ~4% level in E. coli (data not shown). The GST–PheL termination product has different mobility with different constructs (i.e. in Figures 5 and 6). For example, in Figure 6D the first lane has products from the HLH5 construct in which the UUC codon preceding CCC UGA was changed to another phenylalanine codon, UUU. The band corresponding to the termination product migrates slower than in the next lane, which contains products from the construct in which the UUC codon was changed to a CUC leucine codon. The likely explanation for this phenomenon is that phenylalanine in the penultimate position, unlike leucine, hinders termination and promotes alternative translational events such as frameshifting and tmRNA tagging (71). A tmRNA-tagged termination product would be 10 amino acids longer and would migrate slightly slower than the untagged version. The duration of the pulse-chase experiment is only 1–2 min and at least part of any tagged product may not be degraded. Our attempts to purify a GST–PheL tmRNA-tagged product from either wild-type cells or tmRNA− cells expressing tmRNA encoding a 6-histidine
Frameshifting in $p$he$L$ decoding

Frameshifting at the CCC_UGA in decoding $E$. coli K12 $p$he$L$ occurs with 15% efficiency, whereas in wild-type $S$. enterica the comparable frameshifting efficiency is much lower, ~1.9% (Figure 7A). The reason for the difference between the $p$he$L$ frameshifting levels in $S$. enterica and $E$. coli has not been studied, but two aspects merit consideration. $E$. coli is 15% more negatively supercoiled than $S$. enterica with significant effects on gene expression (73). Increasing the level of negative supercoiling of the $S$. enterica promoter of the 4-gene tRNA operon that includes proM, the gene for the near-cognate proline tRNA$_{cma5UGG}$ (74), leads to elevated levels of this tRNA (75). Perhaps there is an elevated ratio of proM/proL-encoded tRNAs in $E$. coli compared with $S$. enterica and, if so, it would lead to more frameshifting; however, it is unknown whether the relative expression of proM/proL has been adapted to compensate for the different levels of supercoiling in different organisms. Another, even more relevant, consideration is a key difference between the release factors 2 of $S$. enterica and $E$. coli K12 that results in reduced termination efficiency of the latter (76,77). This is correlated with enhanced frameshifting at a codon 5’-adjacent to UGA (78).

Björnsson et al. (15) showed that phenylalanine in the penultimate position diminishes termination efficiency and suggested that hydrophobic and acidic amino acids, in general, have this effect. Indeed, when the $p$he$L$ penultimate sense codon, that for phenylalanine, was substituted with a tyrosine (hydrophobic) or aspartic acid (acidic) codon, the level of frameshifting was reduced by half, which was still much higher than the undetectable level when an arginine (basic) codon is present (Figure 6D and E). Nevertheless, changing the phenylalanine codon preceding CCC_UGA in $p$he$L$ to a tyrosine codon did result in a 2-fold reduction of frameshifting efficiency (Figure 6D and E), suggesting that the antitermination effect of the penultimate amino acid is not solely dependent on its hydrophobic properties.

Programmed autoregulatory frameshifting is utilized in $E$. coli release factor 2 mRNA decoding. The sequences features involved in this frameshifting incidentally cause some level of internal initiation, but the process has no functional implications (79). By contrast, the present analysis suggests that in $p$he$L$ decoding it is the frameshifting that is incidental to sequence features selected to influence attenuation via their effect on termination. The general reason for the extent of the nucleotide conservation 5’ of $p$he$L$ genes that terminate with UGA remains for future work to discern.

Searches for new cases where frameshifting is utilized for gene expression (80) should be (i) mindful that some programmed and high-level frameshifting may be incidental and (ii) that nascent peptide stimulators may be awaiting discovery even if the levels of frameshifting attained does not approach the 60% found here with a heterologous nascent peptide stimulator.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEGEMENTS

We are grateful to Norma Wills for her assistance and suggestions on experimental procedures, to Dr James Pittard for his crucial advice on de-repression conditions for phenylalanine operon and for providing the ABI360 strain and to the $E$. coli Genetic Stock Center for providing pLA2 and pINT-ts vectors.

FUNDING

Swedish Science Council and the Carl Trygger Foundation (to G.R.B.); National Institute of Health (grant number GM079523 to J.F.A.); Science Foundation Ireland (to P.V.B. and J.F.A.). Funding for open access charge: Science Foundation Ireland (to J.F.A.).

Conflict of interest statement. None declared.

REFERENCES


