954–960 Nucleic Acids Research, 1998, Vol. 26, No. 4 © 1998 Oxford University Press

Translational termination in *Escherichia coli*: three bases following the stop codon crosslink to release factor 2 and affect the decoding efficiency of UGA-containing signals

Elizabeth S. Poole, Louise L. Major, Sally A. Mannering and Warren P. Tate*

Department of Biochemistry and Centre for Gene Research, University of Otago, PO Box 56, Dunedin, New Zealand

Received October 31, 1997; Revised and Accepted December 17, 1997

ABSTRACT

The observations that the *Escherichia coli* release factor 2 (RF2) crosslinks with the base following the stop codon (+4 N), and that the identity of this base strongly influences the decoding efficiency of stop signals, stimulated us to determine whether there was a more extended termination signal for RF2 recognition. Analysis of the 3′ contexts of the 1248 genes in the *E.coli* genome terminating with UGA showed a strong bias for U in the +4 position and a general bias for A and against C in most positions to +10, consistent with the concept of an extended sequence element. Site-directed crosslinking occurred to RF2 from a thio-U sited at the +4, +5 and +6 bases following the UGA stop codon but not beyond (+7 to +10). Varying the +4 to +6 bases modulated the strength of the crosslink from the +1 invariant U to RF2. A strong selection bias for particular bases in the +4 to +6 positions of certain *E.coli* UGANNN termination sites correlated in some cases with crosslinking efficiency to RF2 and *in vivo* termination signal strength. These data suggest that RF2 may recognise at least a hexanucleotide UGA-containing sequence and that particular base combinations within this sequence influence termination signal decoding efficiency.

INTRODUCTION

The Class I polypeptide chain release factors (RFs) are proteins involved in decoding the translational termination signal (1). A tRNA analogue model for how RFs function was proposed after evidence that these proteins spanned the decoding site of the small subunit of the ribosome and the peptidyltransferase centre of the large subunit like a tRNA (2). Subsequently, it was proposed that a region of the RF mimics the tRNA anticodon and is homologous to regions of EF-G domain IV (3–5). Whereas tRNA recognition of the translational ‘elongation signals’ (the sense codons) involves only three bases in the mRNA because of codon:anticodon interaction, the recognition of translational termination signals by the RF proteins clearly must occur by a different mechanism and could involve a larger sequence element than three bases (6,7).

Over three decades, comprehensive studies on the suppression of nonsense codons have supported the concept that the termination signal may be >3 nt (8–21). Elucidating the elements important for termination has proven difficult and the development of new approaches to address this problem was needed.

For this purpose, a database, TransTerm, containing sequences at translational termination sites was established (22) and has been developed and updated continually (23). We have used this database to analyse sequences around stop codons in genes from a wide range of organisms and have shown that there is a significant bias in the surrounding codon context on both sides of the termination codon (24). The most striking bias was in the position following the codon (+4 base) and a number of experimental approaches with different organisms have been applied to provide strong evidence that the base in this position is a key determinant of the efficiency by which the signal is decoded (16,25–27). *In vitro* experiments with *E.coli* termination complexes using a zero-length crosslinking moiety, thio-U, as the +4 base of the mRNA showed crosslinking to the decoding factor, implying close proximity of the RF to this position in the mRNA (28).

The statistical analyses of translational termination sites raised the possibility that there may be further contacts between the decoding RF and other bases in the 3′ context of the stop codon, and an extended sequence element might form part of the molecular signature of the termination signal. In this study, we have addressed this question using termination complexes with *E.coli* RF2 and small designed mRNAs containing thio-U residues (29) at various positions 3′ to the stop codon. Subsequent photoactivation induces crosslinks from the thio-U and allows a ‘snapshot’ of where decoding RF molecules are ‘fixed’ in close proximity to the bases. In addition, we have determined *in vivo* whether selected contexts from the TransTerm analyses influence the strength of the termination signal when placed in competition with the +1 frameshift event at the RF2 frameshift site.

MATERIALS AND METHODS

Materials

Deoxyoligonucleotides were made using an Applied Biosystems 380B DNA synthesizer or purchased from Macromolecular Resources, Colorado State University. [α-^{32}P]GTP (3000 Ci/mmol)

*To whom correspondence should be addressed. Tel: +64 3 479 7839; Fax: +64 3 479 7866; Email: warren.tate@stonebow.otago.ac.nz

and Hybond transfer membranes were supplied from Amersham. A RiboMAX™ Large Scale RNA Production System (Promega) kit was used for *in vitro* transcription reactions. $tRNA^{Ala-2}$ was supplied by Subriden. UV irradiation was carried out using a 15 watt National Matsushita black light (FL15 BL) in a Griffin UV light box through a glass filter that was impermeable to light <300 nm. Ribonuclease T1 was purchased from Boehringer Mannheim. The pMAL™-c2 plasmid and Maltose Binding Protein (MBP) antibody, restriction enzymes and buffers, and T4 DNA ligase were purchased from New England Biolabs. A Promega Wizard™ Miniprep DNA purification kit was used to prepare plasmid DNA and cloned DNA was sequenced using a 373A AB1 Sequencer. Plasmids were electroporated into bacterial cells using an Electro Cell Manipulator© 600 (BTX). Schleicher and Schuell supplied the nitrocellulose transfer membranes. Other chemical reagents were purchased from Sigma. Bio-Rad supplied the Mini-PROTEAN II electrophoresis cell and the Mini Trans-Blot electrophoretic transfer cell that were used for gel electrophoresis and protein transfer respectively, as well as a GS-670 imaging densitometer used for laser densitometry of the proteins from the *in vivo* bacterial experiments. An LKB ULTROSCAN XL laser densitometer was used for the analysis of the bands produced from autoradiography of the crosslinked proteins.

Media and bacterial strains

Bacteria containing plasmids conferring ampicillin resistance were grown in Luria Broth with 100 µg/ml ampicillin. Protein expression was induced from the P_{tac} promotor with IPTG at 1 mM. Primary cloning was carried out in the *E.coli* strain TG1, then extracted plasmids were electroporated into the *E.coli* strain FJU112 [D(lac pro) gyrA ara recA56^Tn/*10*, F′lacI^{Q1}] (30) for analysis of fusion proteins. This strain contains wild-type ribosomes and no suppressor tRNAs which could compete with the termination or frameshifting events in the translational termination/frameshift assay.

Computer sequence analysis

Statistical analyses of nucleotide sequences were performed on the complete genome of *E.coli* (GenBank Accession No. U00096) after entry into the TransTerm database (23; http://biochem.otago.ac.nz:800/Transterm/homepage.html) as described previously (31).

For UGA termination signals χ^2 values of the non-randomness at each nucleotide position following the stop codon were calculated. Expected values were predicted by mononucleotide frequencies observed in the 100 nt of non-coding region 3′ to the UGA stop codon (Fig. 1A). If a base occurred at the termination site at a greater frequency than expected, then the bias was presented as a positive value. If it occurred at a lower frequency, then the bias was presented as a negative value. In such an analysis, each χ^2 value was multiplied by the sign of z where (observed – expected) = z (Fig. 1B).

Deviation in the use of hexanucleotide termination signals (Fig. 4A) was calculated as follows: deviation = (observed – expected)/expected. For a given hexanucleotide stop signal, expected values were calculated from the occurrence of the hexanucleotide signal in the non-coding region of all genes. For example, the signal UGACUG occurred 72 times in all non-coding

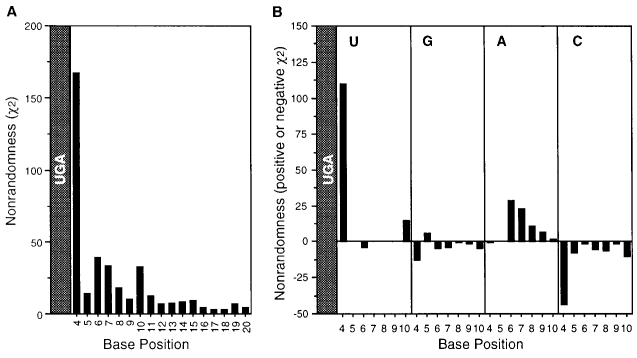

Figure 1. Statistical analysis of the 3′ context of UGA-containing *E.coli* termination signals. (**A**) The non-randomness (χ^2) was determined for each of the 17 positions 3′ to the 1248 genes ending in UGA from the complete *E.coli* genome. (**B**) The bias of each base in the +4 to +10 positions following UGA stop codons. An occurrence at a higher frequency than expected is expressed as a positive value, and an occurrence at a lower frequency than expected is expressed as a negative value.

regions. Since there were a total of 12179 'stop signals' in the non-coding regions, UGACUG represented 0.14%. Therefore, at termination sites, of the 4160 hexanucleotide stop signals included in this analysis 0.14% or 25 would be expected to be UGACUG. This contrasts with the observed occurrence of 6. There is a general bias against UGAC and UGAA termination signals in the *E.coli* genome. Therefore, in determining the deviation for the stop signals UGACUN and UGAAGN the difference between the actual deviation for each individual signal within the set and the average deviation of all the signals in each set was calculated (Fig. 5A).

In vitro transcriptions

The method of Stade *et al.* (32) was used to prepare 4-thio-UTP from 4-thio-UDP (Sigma). Transcription reactions were carried out as described (28). In brief, the single-stranded DNA templates (~0.2 nmol) were annealed to an equimolar amount of T7 polymerase primer in transcription buffer by heating (65°C for 3 min) then cooling to room temperature. Transcription reactions (100 µl vol) containing the annealed primer:templates in transcription buffer supplemented with 7.5 mM ATP and CTP, 0.25 mM GTP, 0.52 mM 4-thio-UTP, 26 µCi [α-^{32}P]GTP and 3000 U T7 RNA polymerase were incubated at 37°C overnight. The mRNA transcripts were purified on a 15% polyacrylamide [38:2 acrylamide:bis(acrylamide)]/ 7 M urea/ 0.1% SDS gel then located by autoradiography. The bands were excised, extracted with phenol/SDS buffer, collected by ethanol precipitation and the yields calculated from Cerenkov radiation of the samples.

Ribosomal complex formation and site-directed crosslinking

Ribosomal complexes were formed by incubating 25 pmol 70S ribosomes with 1000 pmol of [^{32}P]mRNA, 75 pmol $tRNA^{Ala-2}$ (anticodon VGC) and 140 pmol RF2 in 50 µl buffer containing 20 mM Tris–HCl pH 7.4, 100 mM NH_4Cl, 20 mM $MgCl_2$ and 6% (v/v) ethanol at 37°C for 30 min. Crosslinks were formed by dispensing the complexes onto Parafilm resting on ice then UV irradiating from a distance of 10 cm for 30 min.

Analysis of crosslinked complexes by gel separation

Aliquots (10 µl) of crosslinked complexes were digested with 100 U of ribonuclease T1 at 37°C for 30 min. Crosslinked samples, and those treated with ribonuclease, were resolved on a 12% polyacrylamide [29:1 acrylamide:bis(acrylamide)] mini gel system prior to transfer to a nitrocellulose membrane as described (25). The crosslinked proteins were visualised by autoradiography and the intensity of the crosslink bands from at least three separate experiments were measured by densitometry. An average of these measurements was used to calculate the relative crosslink efficiency of RF2–mRNA specific bands at different contexts.

Plasmid construction and analysis of expressed fusion proteins

The formation and analysis of the plasmid constructs has been described in detail previously (31). In brief, for the current study complementary redundant deoxyoligonucleotides spanning the RF2 frameshift window and containing UGANUA, UGAAGN and UGACUN stop signal context series were annealed and directionally cloned into the pMal™ expression plasmid. Following introduction of the plasmids into the *E.coli* strain TG1, recombinant clones were identified by sequencing. Plasmid DNA expressing fusion proteins from the RF2 frameshift window were introduced into *E.coli* strain FJU112, then fusion protein expression induced and the products analysed immunologically following PAGE and western blotting using the MBP as described (25). Relative proportions of frameshift and termination products were determined by laser densitometry. Data from at least three separate experiments that used multiple clones of the same construct were used to calculate the relative termination efficiency. Release factor selection rates were calculated using a modification of the equation derived by Pedersen and Curran (16) as described in Poole *et al.* (25).

RESULTS

Analysis of bases 3′ to UGA stop codons

Recently, the sequence of the complete *E.coli* genome has become available, enabling the contexts of all the genes to be extracted for entry into our TransTerm database (23). Of particular interest to our current research programme are those contexts downstream from UGA stop codons (1248 genes of 4268 coding sequences of the complete *E.coli* genome). UGA-containing signals are decoded only by RF2, the factor used for a comprehensive site-directed crosslinking study from defined sites in the mRNA as described below. Analysis of the sequences in the TransTerm database allowed predictions of how 3′ contexts from stop codons might influence translational termination, and then these predictions were tested experimentally.

The non-randomness of nucleotides in positions +4 to +20 following UGA stop codons is shown in Figure 1A. Consistent with previous analyses for all stop codons in a more restricted dataset (24), the most significant deviation was for the base immediately following the stop codon. However, there were biases (as indicated from the χ^2 values) which were significantly higher in the positions to +10 that were not evident in the more distant positions.

Were these biases indicative of a preference for a particular base at each position (as might be expected if there was an extended sequence element recognised by RF2)? To determine this, the occurrence of a particular base in each position following UGA at termination sites was compared with its occurrence following UGA in non-coding regions. The analysis considered the first 100 positions downstream from the stop codon of each gene, or fewer positions if a second gene began within this region. Of the 4268 genes, 4160 were included in the analysis as they contained no base ambiguities and sufficient intergenic sequence. If a base occurred at the termination site at a greater frequency than expected, then the bias was presented as a positive value, if it occurred at a lower frequency, then the bias was presented as a negative value (Fig. 1B). The most striking differences between termination sites and the non-coding sites was the strong bias for U and against C in the +4 position. For the other positions extending to +10, there was a general bias for A in the +6 to +10 positions and a more modest bias against C.

We have subdivided the dataset further into four base (UGAN) signals and analysed the following bases in the same manner. The general bias for A and against C was still evident, although there was some variation among the +4 base contexts (not shown). This analysis revealed why A apparently was not significant in the +5 position (as shown in Fig. 1B). For two contexts (+4 A or C), A was strongly favoured. However, this was contrasted by A being relatively rare in the other two contexts (+4 U or G).

The critical question is whether these observed statistical biases are important for the decoding mechanism during translational termination? We know already that the strong nucleotide bias at the +4 position of all stop signals profoundly reflects termination efficiency in both bacteria and mammals (16,25,26), and that this base is in close proximity to the decoding RF (28). In the current study, in order to define the extent of the RF recognition signature, we have used a site-directed crosslinking strategy to determine RF contact between the +4 to +10 positions, together with an *in vivo* assay that allows measurement of termination signal strength.

RF2 crosslinks to bases 3′ of the UGA stop codon

For analysis of the extent of RF2 proximity with bases 3′ to the stop signal, seven ^{32}P-labelled mRNA analogues were transcribed *in vitro* from designed long oligonucleotide templates. The oligonucleotides were designed to reflect the statistical bias for A in the 3′ context positions. Each mRNA analogue contained the zero-length crosslinking reagent thio-U in the +1 position of the stop signal and a second thio-U sited in one of the positions +5 to +10 substituting for A. The statistical analysis has shown that U occurs at the expected frequency in most positions from +5 to +10 following UGAG, the signal used for these experiments. As a positive control, we have used the thio-U to substitute for G in the +4 position, the position shown recently to support crosslinking to RF2 (28). A thio-U for U substitution in the essential +1 position of the stop codon in mRNA analogues has been demonstrated previously not to affect RF binding in termination complexes (29). In the current experiments termination complexes were formed using *E.coli* ribosomes, with the stop codon of the mRNA positioned in the A-site by a P-site bound $tRNA^{Ala-2}$ in the presence of RF2. Irradiation with UV at wavelengths >300 nm specifically activates the thio-U residues to form crosslinks with molecules in the immediate vicinity.

Our previous study has shown that RF2 crosslinked to any mRNA fragment is retarded during electrophoresis (28). By subjecting the mRNA in the crosslinked complex to ribonuclease

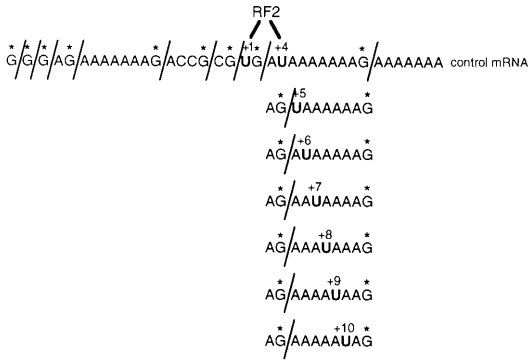

Figure 2. The products generated by ribonuclease (RNAse) T1 digestion of mRNA analogues containing thio-U. The control mRNA (+1 and +4 thio-U) is shown in full above the test mRNAs containing +1 and +5 to +10 (as shown) thio-U. The sites of RNAse T1 cleavage are shown by a diagonal bar. The asterisks denote the radiolabelled G. The solid bars represent crosslinks to RF2 from the control mRNA.

T1 digestion, which specifically cuts on the 3′ side of G nucleotides, a fragment of a defined length is attached to the RF. The mobility of the RF2–mRNA fragment then depends on the length of the attached mRNA fragment (and hence the position in the mRNA from where the crosslink originated). The strategy with the potential crosslinked mRNA fragments for the six mRNAs under study and the control is illustrated in Figure 2; diagonal bars represent the cleavage positions. A crosslink from the +1 position will result in a dinucleotide fragment attached to RF2, a crosslink from the +4 position results in the attachment of a decanucleotide fragment and a crosslink from any of the +5 to +10 positions leaves an octanucleotide fragment attached to RF2. The mRNAs were radiolabelled at the G nucleotides so that after ribonuclease T1 digestion and separation by PAGE, the crosslinked fragments could be identified by autoradiography. We used an antibody to RF2 to confirm that the mRNA fragment was attached to the factor.

The results obtained with the control experiment (where the mRNA analogue contains thio-U in the +1 and +4 positions) is shown in Figure 3A. A prominent crosslinked band was seen in complexes containing RF2 (lane 1, arrow) that was absent in complexes lacking RF2 (lane 3). Ribonuclease T1 digestion destroyed the complex and produced two bands. Previously, we have identified the lower prominent band at the position of native RF2 (46 kDa) (lane 2) as being derived from the +1 position crosslink and containing a dinucleotide crosslinked to RF2. The less intense RF-specific radiolabelled band in lane 2 that migrated between the band at the native RF2 position (+1 band) and the position of the undigested RF2–mRNA complex is derived from the +4 position crosslink and contains a decanucleotide fragment crosslinked to RF2 (28). This band was absent when the mRNA contained only a thio-U in the +1 position of the stop codon (not shown). We have used previously RNA fingerprinting to confirm that these bands arose from crosslinks from the +1 and +4 positions in the RNA (28). The two bands were immunologically positive for RF2 and absent in complexes lacking RF2 (lanes 3 and 4).

Typical RF2–mRNA crosslink patterns for mRNA analogues containing thio-U from the +4 to the +10 position after ribonuclease T1 digestion are shown in Figure 3B. An undigested RF2–mRNA crosslinked complex was included (lane 1). All crosslinked complexes showed a significant band at the position of native RF2 corresponding to a crosslink from the +1 position (lanes 2–8). The higher RF2-specific crosslinked band that originated from the downstream position was evident in complexes containing a +5 and +6 thio-U (lanes 3 and 4) as well as +4 (lane 2). In each case, they reacted with an RF2-specific antibody consistent with the migration of RF2 being retarded by the attached octanucleotide fragments. In contrast, this band was absent from RF2–mRNA complexes with a thio-U sited in the +7 through to the +10 positions of the mRNA (lanes 5–8). A band migrating at the position of the undigested RF2–mRNA complex was sometimes present (asterisk, lanes 2–8) but it did not react with the RF2-specific antibody. A +1 thio-U-containing mRNA analogue has been shown to crosslink to ribosomal protein S1 in

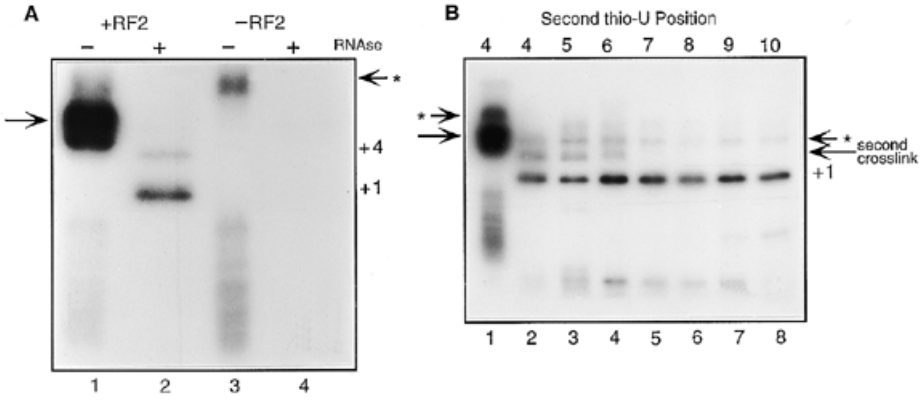

Figure 3. PAGE separation of crosslinked complexes. (**A**) Analysis of the fragments from mRNA containing UGAU. Lanes 1 and 2, and lanes 3 and 4 show the crosslinks formed in the presence and absence of RF2, respectively, both before (lanes 1 and 3) and after (lanes 2 and 4) ribonuclease (RNAse) T1 digestion. The large arrow shows the position of the RF2–mRNA crosslinked species prior to digestion. The upper crosslinked species (denoted with an asterisk) in lanes 1 and 3 represent probable crosslinks from the mRNA to ribosomal protein S1. (**B**) Analysis of the fragments from mRNAs containing a thio-U in the +1 position and a second thio-U in any of the +4 to +10 positions as shown. All crosslinks were formed in the presence of RF2. An undigested crosslinked complex is shown in lane 1 (large arrow) and the complexes shown in lanes 2–8 have all been digested with RNAse T1. The upper crosslinked species (denoted with an asterisk) represents probable crosslinks from the mRNA to ribosomal protein S1.

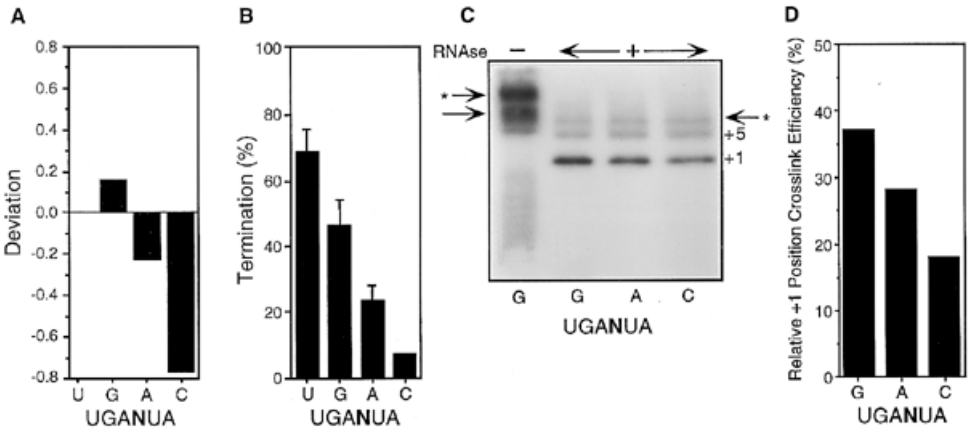

Figure 4. Analysis of UGA**N**UA hexanucleotide contexts. (**A**) Deviations of the observed from the expected occurrences for each N base in this context. Bars above the line mean that the base occurs more frequently than expected and those below the line that the base occurs less frequently than expected. (**B**) The *in vivo* translational termination efficiency at UGA**N**UA contexts. The data was derived from laser densitometry of the expressed protein products as described in the Materials and Methods. (**C**) Analysis of fragments from mRNAs containing UGA**N**UA stop signals (N ≠ U). All crosslinks were formed in the presence of RF2. The left lane shows the RF2–mRNA crosslinked complex prior to RNAse T1 digestion. The upper crosslinked species in all lanes (asterisk) represents probable crosslinks from the mRNA to ribosomal protein S1. (**D**) Relative efficiency of the +1 crosslink in mRNAs containing UGA**N**UA contexts (N ≠ U). Bands from at least three separate experiments were analysed by laser densitometry and the efficiency of the +1 crosslinks is expressed as a percentage of the total combined crosslinks of the +1 and +5 positions.

the presence or absence of RF and migrates to a similar position (29) and is the most likely identity of this radiolabelled band.

Stop signal context affects termination efficiency

The crosslinking experiments demonstrated that the RF is in close proximity with the +1 and the +4 to +6 positions of the mRNA. Termination efficiency might be related to how well the RF contacts each signal at these positions. UGACUA and UGAGUA are examples of hexanucleotide contexts that we have shown previously to have stop signal efficiencies that are low and high respectively (25). As mRNA proximity to RF2 extends to the +6 base, we investigated whether variations in the +4, +5 and +6 bases in these two contexts would affect the crosslink from the invariant +1 U position of UGA.

Firstly, reversing the order of the +5 and +6 bases from UGAGUA to UGAGAU increased RF2 crosslinking efficiency from the +1 position by 37% but a change from UGACUA to UGACAU slightly reduced efficiency (Table 1). These results suggest that the 3′ context can modulate the closeness or orientation of the first position of the stop codon with respect to RF-2. Second, a detailed comparison of the characteristics of UGACUA, the weak signal, with UGAGUA and the other two members of the series, UGAUUA and UGAAUA, have been examined by a number of criteria. Figure 4A shows a statistical analysis of the relative frequency of these signals at termination sites over that expected after comparison with the non-coding regions. In the +4 position of the UGANUA context, G is selected for, U is at the expected frequency and A and particularly C are selected against. Figure 4B shows the influence of the +4 base on *in vivo* termination efficiency at these UGANUA signals in *E.coli* (31), as measured in a competition assay where termination competes with a frameshift event at the RF2 recoding site. The order of efficiency of the signal was U>G>A>C for the +4 base in these hexanucleotides. The RF selection rate (16) for UGAUUA was 28-fold more efficient than for UGACUA. Is this order of signal strength at UGANUA contexts reflected in how efficiently RF2 can be crosslinked with the +1 and +5 thio-U bases of the same sequences? Figure 4C shows an example of the crosslink patterns for UGAGUA, UGAAUA and UGACUA contexts after ribonuclease T1 digestion. The UGAUUA context was not included as the adjacent +4 and +5 thio-Us could not be separated by the ribonuclease T1 digestion strategy and would confound the analysis. The efficiency of crosslinking from the +5 base showed little difference among the three contexts. In contrast, RF2 contact with the +1 base showed significant variation, with the strength of the crosslink decreasing in the order of +4 G>A>C. A histogram of the relative efficiency of +1 crosslink formation where the +4 base is G, A or C is shown in Figure 4D. Strikingly, this order of crosslink efficiency correlates with both the frequency of occurrence of the signals at termination sites and with the *in vivo* termination efficiency directed by the +4 base of the signals (Fig. 4A and B).

Table 1. Relative efficiency of RF2 crosslinks at each hexanucleotide signal

	UA	AU
UGAGNN	1.0	1.37
UGACNN	1.0	0.88

The efficiency is the average crosslink intensity (measured by densitometry) for each signal relative to that for UGAGUA and UGACUA, respectively. The fifth and sixth bases are shown at the head of the Table.

A statistical analysis was performed on termination sites from the complete *E.coli* genome with six bases (+1 to +6) constituting the signal. Five of the six bases were fixed and there was variation in the remaining position. The biases in the numbers of specific sequences in each dataset provide useful predictive data as to what might be more or less efficient termination signals. From this analysis we selected for examination several hexanucleotide series varying in the +6 position, including UGAAGN and UGACUN. These latter contexts show striking statistical differences between observed and

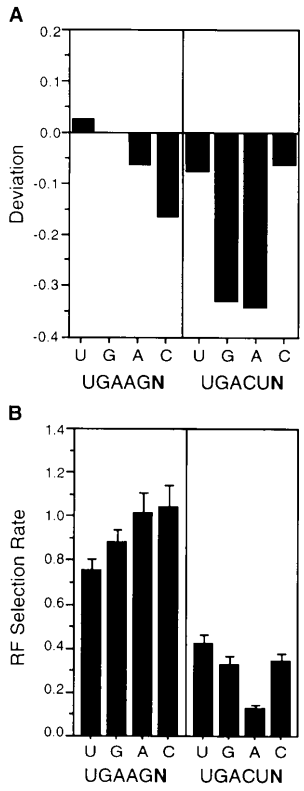

Figure 5. Analysis of UGAAG**N** and UGACU**N** hexanucleotide contexts. (**A**) Deviations of the observed from the expected occurrences for each N base in these contexts. Bars above the line mean that the base occurs more frequently than expected and those below the line that the base occurs less frequently than expected. (**B**) The *in vivo* RF selection rate at each context. The rate of RF selection relative to the rate of frameshifting was calculated from data derived from at least three separate experiments that used multiple clones of the same construct.

expected occurrences at termination sites for each of the four bases in the +6 position (Fig. 5A). Both contexts are already strongly selected against because of the +4 A and +4 C, respectively, which are known to affect termination efficiency for UGA-containing signals, and the data are normalised for this +4 base bias in each case. For the UGAAGN contexts, A and particularly C are selected against in the +6 position. The UGACUN contexts are rare at termination sites with a stronger bias against A and G in the +6 position than against U and C. The selection against A in these contexts was in contrast to the general bias for A as a +6 base when all UGA-containing termination signals are considered (Fig. 1B).

We used, as above, the frameshift site in the RF2 gene to measure *in vivo* the rate of RF selection at each UGACUN and UGAAGN termination signal in competition with the frameshifting event in *E.coli*. The RF selection rates for the UGAAGN and UGACUN contexts varied with the +6 base in the order C = A>G>U for the UGAAGN series, and U>G = C>>A for the UGACUN series. In these cases, in contrast to the UGANUA series, the data did not parallel the statistical selection biases. For example, although C and G were infrequently found at the +6 position of UGAAGN and UGACUN contexts, respectively, they did not weaken the signal (compare Fig. 5B with A).

DISCUSSION

Statistical biases extending far beyond the triplet were found in the bases downstream of the stop codon when the complete gene complement of *E.coli* was examined. If these sequence contexts reflect an important aspect of bacterial cell biology, then an obvious candidate is the termination phase of protein biosynthesis.

Already, extensive evidence has been collected for a role of the base following the stop codon (+4) in polypeptide chain termination (16,25–27), and for this base being in close proximity to the decoding RF in a termination complex (28). Do bases beyond the +4 base have a role in termination and perhaps form part of the recognition element for the decoding RFs? In the current work we have combined two experimental approaches to answer this question; *in vitro* site-directed crosslinking, and *in vivo* measurement of frameshifting efficiencies as a function of termination signal competitiveness. A critical experiment illustrated in Figure 3B has shown that the structural elements of RF2 are positioned close enough to the +5 and +6 bases of the mRNA for site-directed crosslinking to occur. In the previous study with the +4 base, we showed that the PAGE-retarded band released from the crosslinked product after RNAse T1 digestion (Fig. 3A) contained the +4 thio-U oligonucleotide (28). We interpret these new data to suggest that the contact the 'anticodon-like' recognition region of the RF makes with the mRNA may extend as far as the +6 base. The failure to crosslink from the +7 to +10 positions could reflect that elements of RF structure are no longer in close proximity to these bases, given that the thio-U is essentially a 'zero-length' crosslinking moiety. However, it is also possible that the microenvironment simply may be unfavourable for crosslinking compared with the +4 to the +6 positions. We probed the recognition process further by showing that changing bases in the +4, +5 and +6 positions affects how efficiently the +1 base crosslinks to the RF. That the sequences downstream of the stop codon can influence this event presumably reflects a changed orientation of the mRNA with respect to the critical elements of RF structure at the +1 base position.

The TransTerm database is particularly helpful to predict sequences which may be strong or poor termination signals when the length of the sequence is small (for example, four or even five bases) as relatively large numbers of sequences occur in each subset. Extending the analysis of termination signals in *E.coli* to a hexanucleotide sequence is at the limit of reliability for significance as the numbers of sequences in each dataset become relatively small. This is particularly so if the +4 base is a C. Nevertheless, we selected potentially suitable hexanucleotide candidates to test termination signal strength. Some sets showed variation in termination signal strength as a function of the +6 base (as shown in Fig. 5). These data suggested that the +6 base was influencing termination but not necessarily as the TransTerm analysis had predicted.

The ultimate aim of our continuing studies has been to define a consensus sequence element for the termination signal that is recognised by the RF, but this may be difficult particularly if each position cannot be defined independently of the others. For a series varying in the +4 base (the UGANUA series), there is a correlation not only between the frequency of occurrence of the four possible sequences at termination sites and their strengths when in competition with frameshifting at the RF2 frameshift site, but also with the efficiency of the crosslink between the decoding RF and the invariant +1 U of the stop signal (Fig. 4). Within this series,

UGACUA seems to be a context that is not recognised easily by RF2. Indeed, this rare UGA context is also found at the +1 frameshift recoding site of the RF2 gene itself (33). In this case, the stop signal is thought to contribute to a pause in translation that is believed to enhance ribosomal slippage over a run of uracils on the mRNA immediately 5′ to the stop signal (34,35).

Important features of the termination signal may not be so evident in a total dataset derived from all the gene sequences of an organism, whereas a special subset can be more informative. Previously, we have used the Codon Adaptation Index (CAI, a measure of sense codon usage correlating with expression) to show quite definite characteristics in the termination signal with respect to the +4 base in the most highly expressed subset (top 5–10%) (24). The key features gradually became less obvious as subsets of lower CAI were examined successively. For example, UAAU, experimentally shown to be the strongest four base stop signal, is by far the most prevalent signal in the highest CAI subset, whereas UGAC the weakest signal, is never found in this subset. This implies that highly expressed genes have a strict requirement for a strong signal which is decoded rapidly, whereas genes that are expressed in lower amounts can tolerate a wider range of signals decoded at different rates. While it is difficult to perform a statistical analysis of larger sequence elements in a CAI subset of ~400 genes because the expected numbers for each specific sequence are small, we are examining currently which hexanucleotide sequences occur more frequently within the top CAI subset, to select candidate sequences for further experimental testing.

Will it be possible to define a consensus sequence element for decoding RF recognition? We have defined the core of the termination signal as URRN (excluding UGGN and where for *E.coli* N:U/G>A/C) (28). An important consideration for a consensus sequence is the sequences 5′ to the stop codon, and a series of important experiments have been carried out that indicate at least two codons (i.e. the six bases –1 to –6) significantly influence termination efficiency (19,20). Here the coding potential of the 5′ sequence is most likely to be the critical feature, and these data may reflect interactions the decoding RF makes with the last tRNA and the last two amino acids. Currently, we are examining the sequences 5′ and 3′ of the stop codon together. It should be possible to define sequence elements which are recognised with high affinity by the decoding RF and decoded rapidly, and those which are recognised poorly and decoded slowly. These will mark the extremes for a consensus recognition element and sequences between these extremes are likely to be tolerated as termination signals by most genes which are under no particular expression demands or are not subjected to a competitive recoding event.

ACKNOWLEDGEMENTS

We would like to thank Dr Mark Dalphin for expert help with the TransTerm database. W.P.T. is an International Scholar of the Howard Hughes Medical Institute and the work described here has been supported by a grant from the Human Frontier Science Program (awarded to W.P.T. and Yoshikazu Nakamura) and the Marsden Fund of NZ.

REFERENCES

1 Tate,W.P., Dalphin,M.E., Pel,H.J. and Mannering,S.A. (1996) *Gen. Eng.*, **18**, 157–182.
2 Moffat,J.G. and Tate,W.P. (1994) *J. Biol. Chem.*, **269**, 18899–18903.
3 Ito,K., Ebihara,K., Uno,M. and Nakamura,Y. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 5443–5448.
4 Nakamura,Y., Ito,K. and Isaksson,L.A. (1996) *Cell*, **87**, 147–150.
5 Uno,M., Ito,K. and Nakamura,Y. (1996) *Biochimie*, **78**, 935–943.
6 Tate,W.P. and Mannering,S.A. (1996) *Mol. Microbiol.*, **21**, 213–219.
7 Tate,W.P., Poole,E.S., Dalphin,M.E., Major,L.L., Crawford,D.J.G. and Mannering,S.A. (1996) *Biochimie*, **78**, 945–952.
8 Salser,W. (1969) *Mol. Gen. Genet.*, **105**, 125–130.
9 Salser,W., Fluck,M. and Epstein,R. (1969) *Cold Spring Harbor Symp. Quant. Biol.*, **34**, 513–520.
10 Fluck,M.M., Salser,W. and Epstein,R.H. (1977) *Mol. Gen. Genet.*, **151**, 137–149.
11 Bossi,L. and Roth,J.R. (1980) *Nature*, **286**, 123–127.
12 Bossi,L. (1983) *J. Mol. Biol.*, **164**, 73–87.
13 Miller,J.H. and Albertini,A.M. (1983) *J. Mol. Biol.*, **164**, 59–71.
14 Smith,D. and Yarus,M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 4397–4401.
15 Buckingham,R.H., Sörensen,P., Pagel,F.T., Hijazi,K.A., Mims,B.H., Brechemier-Baey,D. and Murgola,E.J. (1990) *Biochim. Biophys. Acta*, **1050**, 259–262.
16 Pedersen,W.T. and Curran,J.F. (1991) *J. Mol. Biol.*, **219**, 231–241.
17 Kopelowitz,J., Hampe,C., Goldman,R., Reches,M. and Engelberg-Kulka,H. (1992) *J. Mol. Biol.*, **225**, 261–269.
18 Björnsson,A. and Isaksson,L.A. (1993) *J. Mol. Biol.*, **232**, 1017–1029.
19 Mottagui-Tabar,S., Björnsson,A. and Isaksson,L.A. (1994) *EMBO J.*, **13**, 249–257.
20 Björnsson,A., Mottagui-Tabar,S. and Isaksson,L.A. (1996) *EMBO J.*, **15**, 1696–1704.
21 Zhang,S., Ryden-Aulin,M. and Isaksson,L.A. (1996) *J. Mol. Biol.*, **261**, 98–107.
22 Brown,C.M., Dalphin,M.E., Stockwell,P.A. and Tate,W.P. (1993) *Nucleic Acids Res.*, **21**, 3119–3123.
23 Dalphin,M.E., Brown,C.M., Stockwell,P.A. and Tate,W.P. (1997) *Nucleic Acids Res.*, **25**, 246–247.
24 Tate,W.P., Poole,E.S., Horsfield,J.A., Mannering,S.A., Brown,C.M., Moffat,J.G., Dalphin,M.E., McCaughan,K.K., Major,L.L. and Wilson,D.N. (1995) *Biochem. Cell Biol.*, **73**, 1095–1103.
25 Poole,E.S., Brown,C.M. and Tate,W.P. (1995) *EMBO J.*, **14**, 151–158.
26 McCaughan,K.K., Brown,C.M., Dalphin,M.E., Berry,M.J. and Tate,W.P. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 5431–5435.
27 Bonetti,B., Fu,L., Moon,J. and Bedwell,D.M. (1995) *J. Mol. Biol.*, **251**, 334–345.
28 Poole,E.S., Brimacombe,R. and Tate,W.P. (1997) *RNA*, **3**, 974–982.
29 Tate,W., Greuer,B. and Brimacombe,R. (1990) *Nucleic Acids Res.*, **18**, 6537–6544.
30 Jørgensen,F. and Kurland,C.G. (1990) *J. Mol. Biol.*, **215**, 511–521.
31 Major,L.L., Poole,E.S., Dalphin,M.E., Mannering,S.A. and Tate,W.P. (1996) *Nucleic Acids Res.*, **24**, 2673–2678.
32 Stade,K., Rinke-Appel,J. and Brimacombe,R. (1989) *Nucleic Acids Res.*, **17**, 9889–9908.
33 Craigen,W.J., Cook,R.G., Tate,W.P. and Caskey,C.T. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3616–3620.
34 Weiss,R.B., Dunn,D.M., Atkins,J.F. and Gesteland,R.F. (1987) *Cold Spring Harbor Symp. Quant. Biol.*, **52**, 687–693.
35 Hatfield,D. and Oroszlan,S. (1990) *Trends Biochem. Sci.*, **15**, 186–190.