The EMBO Journal Vol.18 No.19 pp.5423–5433, 1999

Resuming translation on tmRNA: a unique mode of determining a reading frame

Kelly P.Williams[1,2,3], Kimberly A.Martindale[1] and David P.Bartel[2,3]

[1]Department of Biology, Indiana University, 1001 East 3rd Street, Bloomington, IN 47405-3700 and [2]Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, MA 02142, USA

[3]Corresponding authors
e-mail: kwilliam@bio.indiana.edu and dbartel@wi.mit.edu

The bacterial ribosome switches from an mRNA lacking an in-frame stop codon and resumes translation on a specialized RNA known as tmRNA, SsrA or 10Sa RNA. We find that the ribosome can reach and use the extreme 3′ terminal codon of the defective mRNA prior to switching. The first triplet to be translated in tmRNA (the resume codon) is determined at two levels: distant elements in tmRNA restrict resume codon choice to a narrow window and local upstream elements provide precision. Insights from a randomization–selection experiment secure the alignment of tmRNA sequences from diverse species. The triplet UA(A/G) (normally recognized as a stop codon by release factor-1) is strongly conserved two nucleotides upstream of the resume codon. The central adenosine of this triplet is essential for tmRNA activity. The reading frame of tmRNA is determined differently from all other known reading frames in that the first translated codon is not specified by a particular tRNA anticodon.

Keywords: protein degradation/reading frame/ribosome/10Sa RNA/translational initiation

Introduction

Correct registration of translational reading frames is essential for accurate gene expression. During initiation of translation, the start codon is precisely registered in the ribosomal P site by the anticodon of a special initiator tRNA species. Other elements, such as the Shine–Dalgarno interaction, can limit initiation site choice to a relatively small region and thereby facilitate the utilization of certain other triplets that differ by one base from the standard AUG initiation codon. But even for such non-standard start codons, the anticodon of the initiator tRNA is responsible for precise registration of the reading frame.

Certain mRNAs can program frameshift events that are likewise mediated by tRNA anticodons. Typically, the P-site tRNA dissociates from the codon it first encountered in the ribosomal A site and reassociates with a suitably similar mRNA triplet (Farabaugh, 1996; Atkins *et al.*, 1999). Signals in mRNA that tend to slow the ribosome at the time of the shift, such as stop or rare codons in the A site prior to shift, a stable stem–loop or pseudoknot downstream of the shift site or an upstream equivalent of the Shine–Dalgarno sequence, are known to facilitate such re-registration. Most re-registration events are single-nucleotide shifts. Stepping well outside this paradigm is the extremely long shift that occurs on the bacteriophage T4 gene 60 mRNA (Huang *et al.*, 1988; Weiss *et al.*, 1990). This 50-nt bypass shares features of the short-range frameshifts; it depends on identical triplets at either end of the bypassed gap that exchange at the P-site anticodon, and requires a stop codon engaged in the A site prior to the shift event, with a stable stem–loop immediately downstream. It also has a unique requirement for a particular amino acid sequence in the nascent peptide.

In the most radical of ‘frameshifts’, the bacterial ribosome switches in mid-translation from a reading frame of one mRNA molecule to that of the specialized tmRNA molecule (Figure 1A). The switch can occur from the end of virtually any ‘broken’ mRNA (lacking an in-frame stop codon, perhaps due to an untimely encounter with a ribonuclease) or even from internal codons if cognate tRNA levels are very low (Tu *et al.*, 1995; Keiler *et al.*, 1996; Roche and Sauer, 1999). Translation resumes at one particular triplet (the ‘resume codon’) in tmRNA, adding a distinctive C-terminal peptide tag onto the incomplete protein product of the initiating mRNA. In addition to this mRNA-like function, tmRNA also has tRNA-like properties. It forms a half-tRNA structure consisting of the perfect equivalents of an acceptor stem with a CCA tail and a T stem–loop with the characteristic modified bases (Figure 1B); it bears the simple identity elements of alanine tRNA and is a substrate for the alanine tRNA synthetase (Komine *et al.*, 1994; Ushida *et al.*, 1994; Felden *et al.*, 1998). The polypeptide product of this *trans*-translation process contains one particular alanyl residue encoded in neither the initiating mRNA nor tmRNA, but located between the two encoded portions (Tu *et al.*, 1995); this alanyl appears to be the same residue that charged the tmRNA as it entered the ribosome (Nameki *et al.*, 1999b). tmRNA solves two problems that broken mRNA might cause for bacteria: (i) it provides the stop codon that the broken mRNA could not, and so can free otherwise stalled ribosomes; and (ii) the C-terminal peptide tag is a signal for degradation of the entire tagged protein (Keiler *et al.*, 1996). Given the intriguing mechanistic aspects of *trans*-translation, an understanding of tmRNA action is expected to provide new insight into ribosome function.

Here, we examine the manner in which the tag reading frame is determined. Features distant from the reading frame in tmRNA set a small window for the resume codon, while local upstream determinants precisely position the codon. The resume codon is defined prior to its pairing with cognate tRNA; thus, the reading frame of tmRNA is determined differently from all other known reading frames in that the first translated codon is not specified

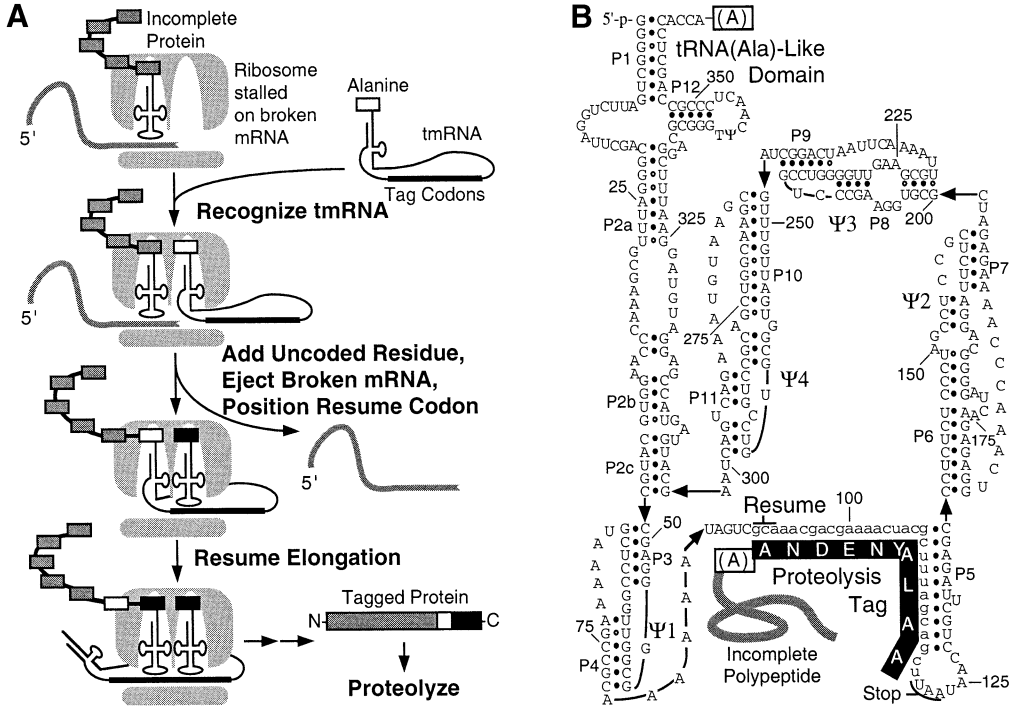

Fig. 1. *Trans*-translation by tmRNA. (**A**) *Trans*-translation mechanism (Keiler *et al.*, 1996). (**B**) tmRNA secondary structure (Williams and Bartel, 1996). P1–12, paired element; Ψ1–4, pseudoknot.

by a particular tRNA anticodon. An adenosine residue 4 nt upstream of the resume codon is essential for *trans*-translation. In natural tmRNA sequences this adenosine is in the center of a triplet that is almost always either UAA or UAG, triplets normally recognized as stop codons by release factor-1 (RF-1).

Results

Tagging of run-off translation products

To understand the mechanism of *trans*-translation, it is useful to characterize the precise state of the ribosome that initiates the process. The question of how closely the ribosome can approach the end of a broken mRNA was addressed *in vivo*. The ribosome is known to protect ~15 nt of mRNA on the downstream side of the codon in the P site (Steitz, 1975; Hartz *et al.*, 1988), and any of these positions could in principle be used by the ribosome to detect the impending 3′ end of a broken mRNA and signal entry of tmRNA.

Tagging was analyzed *in vivo* in an *ssrA*$^-$ *Escherichia coli* strain (i.e. with a disrupted chromosomal tmRNA gene), so that a suitable variant tmRNA gene on a low-copy plasmid could be tested (Figure 2A). To direct most of the cell's tagging activity to a single protein species, a discrete mRNA containing no in-frame stop codon was overproduced (Keiler *et al.*, 1996). Reporter protein that escaped tagging, and the larger tagged form, were detected as distinct bands in a protein gel; alternatively, the ensemble of reporter products was characterized using mass spectrometry.

It was first of interest to determine how far toward the end of a broken message the ribosome can proceed *in vivo* in the absence of tmRNA. Broken mRNA was overproduced in cells lacking tmRNA and the reporter protein was purified. Mass spectrometry revealed complexity in the reporter profile (Figure 2B, dark trace), even for this simple situation, which was not apparent in a protein gel (Figure 3B, lane 1). The profile is consistent with translation proceeding to and using the last possible codon in broken mRNA. Previous mapping of the 3′ termini generated by the transcriptional terminator employed here predicts that the Phe15 codon (Figure 2B, bold type) would be the last complete codon in ~88% of the terminated transcripts (Bertrand *et al.*, 1977). The major peak observed in the reporter profile matched the mass expected from translation up to and including Phe15 (with removal of the N-terminal formylmethionine residue). An additional 5% of the terminated transcripts should have the Phe16 codon intact, and a small peak matched the corresponding reporter species. Use of the last possible codon of broken mRNA in the absence of tmRNA constitutes a 'run-off translation' phenomenon comparable to run-off transcription from linear DNA templates.

A minor series of reporter species with lower molecular weight was also observed. These can be unambiguously identified as the ladder of reporter species trailing back residue-by-residue from the C-terminus of the major species, until the six-histidine unit used for purification was disrupted. Because each of these were substrates for tagging when tmRNA was present (below), they are probably neither breakdown products of the major protein species nor prematurely released from the ribosome. They could, in principle, represent products stalled at the time of sampling on internal codons of intact mRNA. We suspect instead that they result from nuclease action on the primary transcript (and consider it fortunate that such degradation was not more prevalent); i.e. they represent true run-off products from further truncated mRNAs.

The wild-type tmRNA is unsuitable for studying tagging

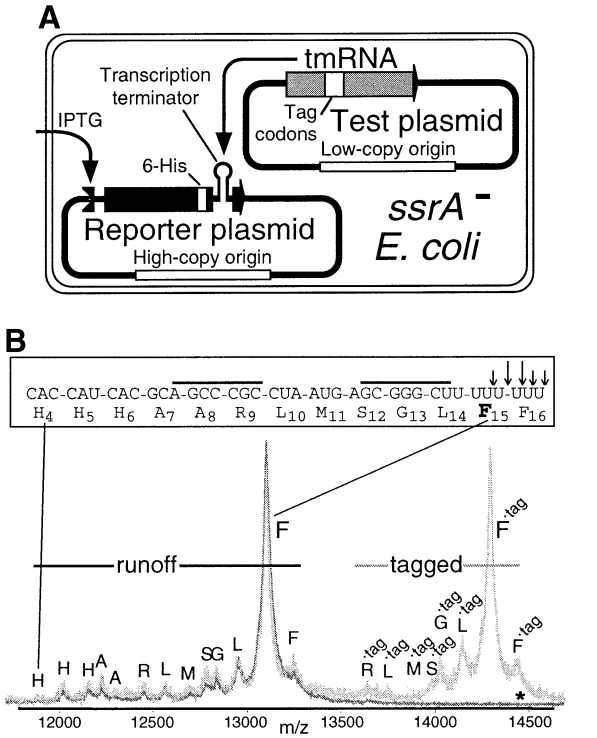

Fig. 2. Induction and analysis of *trans*-translation. (**A**) High-level tagging system. An inducible promoter and a gene-internal terminator allow overexpression of a reporter mRNA lacking an in-frame stop codon, leading to *trans*-translational tagging of the reporter protein when tmRNA is active. The His_6 unit in the reporter allows affinity purification on a nickel resin for further analysis. For the experiments of (B) and Figure 3, the reporter protein was cytoplasmic and DD versions of tmRNA (with the last two tag codons altered from Ala to Asp) were employed so that tagged reporter protein would be resistant to cytoplasmic tail-specific proteases. For the experiments of Figure 6, the reporter protein was periplasmic, the tmRNAs coded for proteolysis-inducing tags, and the major periplasmic tail-specific protease (Tsp) was inactivated by mutation. (**B**) Tagging of would-be run-off translation products. The 3′ sequence of the reporter mRNA is shown, with the stem of the *E.coli trp* attenuator terminator overlined and codons numbered from the first in the His_6 unit. Quantitation of *in vivo* termination site usage (arrows) (Bertrand *et al.*, 1977) predicts that the last intact codons will be L_{14}, F_{15} and F_{16} after 7, 88 and 5% of termination events, respectively. Mass spectrometry profiles are for reporter protein from cells lacking tmRNA (dark trace) or producing DD tmRNA (light trace). Peaks are marked with the last amino acid coded by the broken mRNA printed vertically and tag sequences printed at an angle; in this panel the sequence (A)NDENYALDD is abbreviated as 'tag'. An asterisk (*) marks the mass of the (undetected) protein expected from failure of transcription terminator.

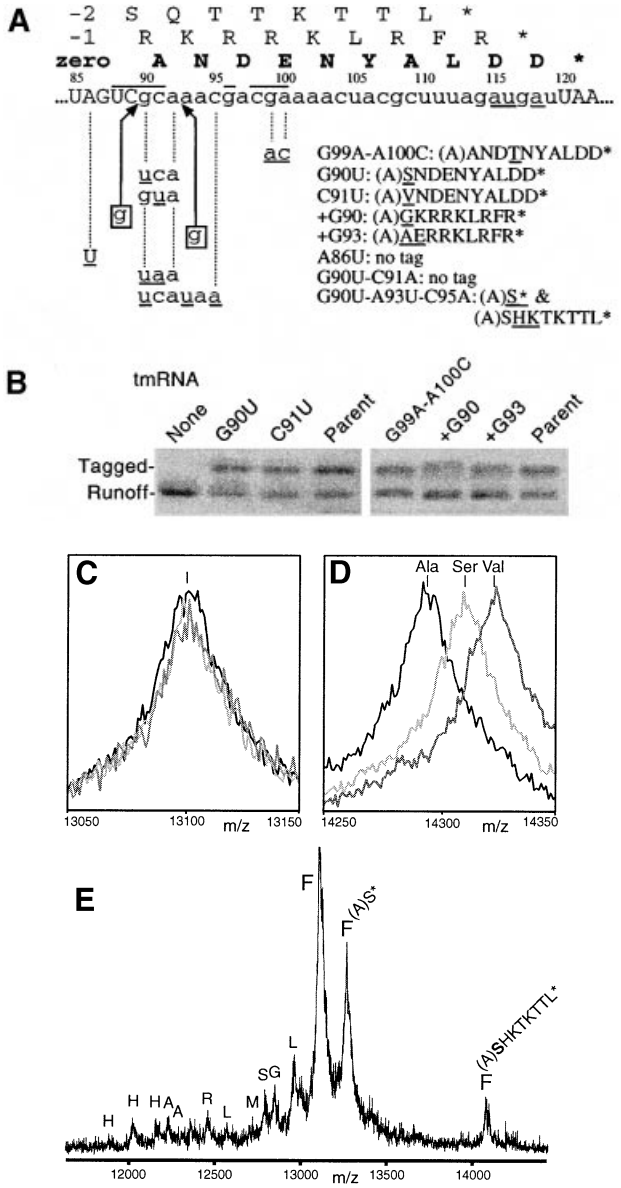

Fig. 3. Directed mutants of the tmRNA resume region. (**A**) The parental tmRNA (DD) for the set of mutants has the differences from wild type underlined; the zero reading frame used in *trans*-translation and those that would result from –1 or –2 frameshifts are shown. Altered bases are underlined and inserted bases are boxed. The tags deduced from mass spectrometry are given, with amino acids altered by mutation underlined. (**B**) Coomassie Blue-stained gel of nickel-purified reporter protein from two separate experiments. Scanning and quantitation gave the following fractional tagging of reporter protein: (left panel) G90U, 0.53; C91U, 0.49; Parental, 0.55; (right panel) G99A-A100C, 0.49; +G90, 0.40; +G93, 0.42; Parental, 0.51. (**C**) Major untagged reporter species. Portion of mass spectrometry traces for parental DD (dark trace), G90U (light trace) and C91U (medium trace) tmRNAs. Tick indicates expected peak position. (**D**) Major tagged reporter species, showing a different portion of the traces used in (C). Ticks mark expected positions of peaks with the indicated residue inserted at the resume codon. (**E**) Reporter mass spectrometry profile for tmRNA mutant G90U-A93U-C95A. The area under the peak corresponding to the expected *trans*-translation product F(A)S* is 4.5 times that for the presumed product F(A)**S**HKTKTTL* of –2 frameshifting upon incorporation of the Ser in bold type.

in this system because the tag it encodes causes degradation of the reporter. Instead, we employed a mutant tmRNA (termed DD) that is active in *trans*-translation but does not cause proteolysis of the tagged proteins (Roche and Sauer, 1999). This was arranged by altering the coding in the tag reading frame; hydrophobic C-terminal tag residues that favor proteolysis were replaced with charged residues (Keiler *et al.*, 1996). Broken mRNA was overproduced in the presence of DD mutant tmRNA, and the reporter protein was purified. When the reporter was examined by gel electrophoresis and Coomassie Blue staining (Figure 3B, 'Parent' lane), two bands of approximately equal intensity appeared, one co-migrating with run-off product and the other consistent with tagging. These artificial conditions of overproducing a pre-broken mRNA revealed that the tagging activity of this mutant tmRNA is significantly reduced relative to the wild type, which leaves little or no run-off reporter (data not shown). When the reporter ensemble produced by DD tmRNA was

examined by mass spectrometry (light trace in Figure 2B), half the profile coincided with that from cells lacking tmRNA, and the other half mirrored the first half, but with the peaks shifted by precisely the mass of the tag. Thus in cells bearing tmRNA, the reporter species observed in the absence of tmRNA were each competent for tagging. Certainly the predominant tagged reporter species and possibly all tagged reporter species initiated *trans*-translation from the last possible codon in broken mRNA, suggesting a simple formula for a signal of tmRNA entry into the ribosome: mRNA nucleotides missing in the ribosomal A site.

Upstream sequence provides the strongest determinants for the resume codon

Can rules for recognition of the resume codon in tmRNA be formulated? The *in vivo* tagging assay allowed us to test features of tmRNA for involvement in resume codon recognition. Site-directed mutants of the resume codon region of tmRNA were constructed, all in the context of the stabilizing DD mutation (Figure 3A). Their tagging activities were compared with that of the otherwise wild-type DD mutant by the assays described above.

A likely determinant of the resume codon position is the resume codon sequence itself. We made two point mutants of the resume codon (G90U and C91U, Figure 3A). Surprisingly, these both performed nearly as well in tagging as the DD parent, as judged by protein–gel analysis of purified reporter (Figure 3B). This result suggested that some if not all determinants of the resume codon lie outside of the codon itself; moreover, it provided an opportunity to compare the mechanism by which the resume codon is registered in the ribosome with the corresponding mechanism for a start codon. Start codons, even variants of the standard AUG that would otherwise encode other amino acids, are recognized uniquely by initiator tRNA. The mutant resume codons might similarly be determined uniquely by $tRNA^{Ala}$, the species that decodes the wild-type resume codon. Mass determination of the tagged reporters revealed instead that the mutant resume codons are read according to the genetic code; replacing the GCA (Ala) wild-type resume codon with UCA (Ser) or GUA (Val) shifted the mass of the tagged reporter by the corresponding 16 or 28 Da, respectively (Figure 3C and D). Thus, no particular incoming tRNA species is required to recognize the resume codon.

Since alteration of the resume codon did not significantly affect tmRNA activity, a search for determinants outside of the resume codon itself was initiated by inserting a nucleotide on one side or the other of the resume codon (+G90 and +G93, Figure 3A). Such an insertion strategy could distinguish effects from sequences upstream or downstream of the resume codon; domination by downstream sequence should lead to retention of the normal reading frame for C-terminal tag residues, while domination by upstream sequence should lead to utilization of the –1 frame. A balanced competition between upstream and downstream elements might have a mixed result or inactivate the tmRNA altogether. Both insertion mutants tagged nearly as well as the parental tmRNA (Figure 3B). Mass spectrometry revealed that they had used the –1 frame. In other words, sequence upstream of the resume codon dominated in the determination of the reading frame, so much so that a different resume codon (Gly) was chosen in the case where a nucleotide was inserted between the upstream sequence and the wild-type resume codon (+90G). This is not to say that sequence downstream of the resume codon normally makes no contribution to selection of the resume codon, but that any such contribution is outweighed by that from upstream sequence.

Two additional resume region variants were designed to test the lower limits of tag size. These were based on the active G90U (Ser) resume codon mutant described above. One variant, which additionally altered the second tag codon to a stop codon (G90U-A93U-C95A), added the expected two-residue tag (uncoded Ala + Ser) (Figure 3E). At a lower yield, an additional tagging product was observed whose mass matches that expected from slippage 2 nt upstream from the resume codon on the tmRNA by the P site $tRNA^{Ser}$, onto a different Ser codon, followed by translation to the first stop codon in the new frame. Two-nucleotide programmed frameshifts are very rare (Gesteland and Atkins, 1996); the observation here may reflect a special instability of the resume codon positioned in the A site.

The second variant of the G90U mutant added a single base change, C91A, creating a stop codon at the resume codon itself. This mutant breaks with the general tolerance by *trans*-translation of mutation in the tag reading frame. One might expect the addition of the uncoded alanine residue alone by this tmRNA mutant, but none was detected in mass spectrometry of the reporter (data not shown). An additional C-terminal alanine residue might have created a proteolysis signal, precluding detection of single-alanine tagged reporter in the preceding experiment. To test its stability, the major run-off product with a single additional C-terminal Ala residue was encoded traditionally, on a single mRNA with a stop codon, in a reporter plasmid and found to be quite stable upon overproduction (data not shown). Thus, the potential single-A tagged product of the mutant tmRNA could have been detected if it had been produced at a reasonable level.

Randomization–selection in the resume region

To survey the region comprehensively for determinants of the resume codon, randomization–selection was undertaken (Figure 4). A genetic system was developed to allow selection of active sequences from the planned pool of tmRNA variants, employing a plasmid that encodes a kanamycin-resistance gene under the control of the Arc transcriptional repressor of phage P22. The Arc mRNA was 'broken' using a terminator of transcription, so that cells lacking tmRNA would produce stable run-off Arc and be sensitive to kanamycin. The repressor is degraded in cells bearing active tmRNA that produces a proteolysis-inducing tag, allowing expression of the kanamycin-resistance gene. To ensure selection only of tmRNAs that could direct the ribosome to the original tag reading frame, a parental tmRNA variant was constructed that generated codons for charged (proteolysis-preventing) residues at the ends of the other two reading frames (Figure 4A). A block of 21 nt centered at the resume codon was completely randomized (with each base present at each position at a frequency near 0.25) to prepare a pool of tmRNA variant plasmids. Thus, the first four codons of the tag reading

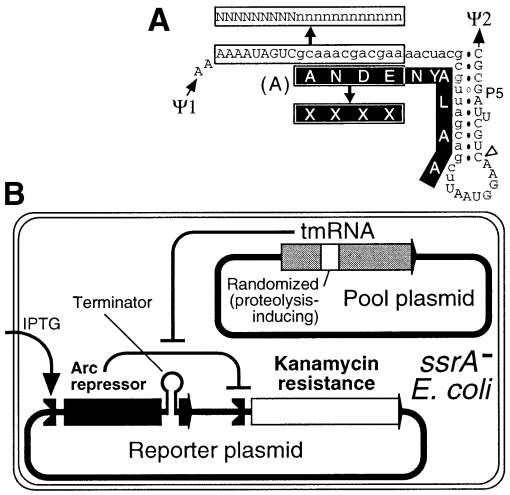

Fig. 4. Randomization–selection of resume region. (**A**) Parent for randomization, showing the region between pseudoknots Ψ1 and Ψ2. Changes relative to wild-type tmRNA (outlined bases, triangle) do not alter normal tag reading frame, but cause the other two reading frames to end with codons for charged residues, such that their encoded tags do not induce proteolysis. For the tmRNA variant pool, 21 positions including four tag codons were completely randomized (boxes). (**B**) Genetic system used for selection. Active tmRNA causes degradation of repressor of antibiotic-resistance gene.

frame were randomized and selection was therefore not only for recognition of a resume codon in the original frame, but also for tag peptide sequences that would not diminish the proteolysis-inducing function. However, the hydrophobic C-terminal tail of the tag, a major determinant of proteolysis, was unaffected by the randomization.

The fraction of plasmids passing the genetic screen was high (4×10^{-3}). tmRNA activity of selectants was initially assessed by growth on a set of plates with varying levels of isopropyl-β-D-galactopyranoside (IPTG) (controlling the broken mRNA for Arc) and kanamycin. This genetic test revealed that all of 23 tested selectants had tmRNA activity that was clearly lower than that of the parent. Apparently, tmRNA variants with the full parental level of activity were much less abundant in the pool than the moderately active variants that were selected. Activity defects were less apparent in an assay where broken mRNA was not purposely expressed. Four selectants were tested by competitive growth against cells lacking tmRNA (K.A.Martindale, E.Roche, R.T.Sauer, D.P.Bartel and K.P.Williams, manuscript in preparation). Each mixed culture was grown for multiple cycles of approach to saturation followed by dilution, conditions where the wild-type tmRNA provides a substantial growth advantage. Three of the selectants produced mixed culture takeover as rapidly as the wild-type tmRNA, and the fourth produced takeover that was substantial but distinctly lower than the others (Table I, column C). Thus, all tested tmRNAs produce some benefit during normal growth, most as strong as that of the wild type.

Selectant plasmids were sequenced (Table I) and consensus information was derived (Figure 5A) after mass spectrometry data presented in the following section mapped each resume codon to the same position as in the wild type. Seven different amino acid specificities for the resume codon were observed, providing further evidence that the identity of the incoming tRNA is relatively unimportant. As expected from the high frequency of selection, few positions showed conservation. Significant exceptions were G90, the first position of the resume codon (occurring in 20 out of 23 selectants) and A86, four positions upstream (21 out of 23 selectants). The probability that a given base would occur at a randomized position at least 20 times in a random sample of size 23 is $<10^{-9}$. A mutant of G90 had already been constructed and shown to be nearly as active as wild type (Figure 3B); apparently, other features provided by the otherwise fully wild-type resume region can compensate for the G90 function that has been revealed by selection. A single point mutation of A86, converting it to a U, was constructed in the DD context (Figure 3A) and assayed as in Figure 2; no tagging was observed, not even addition of the single uncoded Ala that might be expected from interfering with resume codon recognition (data not shown). This base change is unlikely to have global effects on tmRNA structure. In contrast to G90, A86 was essential in the context of the wild-type resume region, and we conclude that the selection has revealed an extremely important determinant of *trans*-translation.

Another benefit of the consensus information from the selection experiment is that it provides a basis for aligning the resume regions of the tmRNAs that have been found in nature (Table II). Only the *Escherichia coli* resume codon has been mapped directly by sequencing of tagged proteins. In other species, the downstream ends of the tag reading frames could be readily identified by the conserved character of the encoded amino acids, but identification of the upstream ends had been hampered by lack of understanding of resume codon determinants. The sequences available for 55 species from the tmRNA website (Williams, 1999) all match features that emerged from randomization–selection, having an A residue with a G residue 4 nt downstream, with the appropriate positioning relative to the tag reading frame and within a limited distance from the upstream first pseudoknot. Confidence is inspired in the resulting alignment by the substantial increase in consensus information it brought out relative to the moderately active selectants (Figure 5B). Thus, we can now predict the tags produced throughout phylogeny and begin phylogenetic analysis of other features in the resume region. We note in particular that the critical A identified by selection lies in the center of what is almost always either a UAG or a UAA triplet (four exceptions replace the U with an A or C).

Tagging activity of selectants

The cytoplasmic reporter protein employed in the tagging assays of Figures 2 and 3 is unsuitable for testing tmRNAs that encode proteolysis-inducing tags (as do the selectants); reporter bearing the wild-type tag is not observed, being subject to multiple cytoplasmic proteases. These proteases can be inactivated by mutation, but cell physiology is severely compromised. In contrast, Tsp (or Prc), the major periplasmic protease that recognizes the tmRNA-directed tag, can be inactivated by mutation with less severe physiological consequences under standard growth conditions (Silber and Sauer, 1994). We analyzed proteolysis-inducing tmRNAs by overproducing a broken mRNA that encodes a reporter protein (a derivative of cytochrome b562) that is directed to the periplasm in *ssrA*$^-$ *tsp*$^-$ *E.coli*

Table I. Selectant sequences and tagging summary

tmRNA	Sequence in randomized region[a]	Tag produced[b]	T[c]	C[d]
None[e]			0	
WT[e,f]	AAAAUAGUCgcaaacgacgaa	(A) A N D E N Y A L A A	100	+++
WT	(as above, but with downstream changes as in selectants)		101	
54[g]	UUAGUAACUggcaacuguccg	(A) G N C P N Y A L A A	82	+++
55[f]	UAAAUAUUCaacaaccauccc	(A) N N H P N Y A L A A	94	
56	GUCAUAAAGuucacccccaag	(A)/F T P K N Y A L A A	96	
57[e]	UAAUUAAUGucaaaagcaaca	(A) S K A T N Y A L A A	91	
58[e,g]	AAAGCAACGgcuacugccugu	(A) A T/A C/N Y A L A A	75	
59	CUAAUGAAUggccucauuucc	(A) G L I S N Y A L A A	88	
60	UUGAUAGUAgccgaccuagaa	(A) A D L E N Y A L A A	57	
61	AGACUAGCUgcaacuacugcc	(A)/A T T A/N Y A L A A	71	
62	GACUGAGAAgacguaacgcuc	(A)/D V/T L N Y A L A A	84	
63	ACUGCAGUUggcagauaugag	(A) G R Y E N Y A L A A	51	
64	AGAAUAAUGgcuuauaagucg	(A)/A Y K S N Y A L A A	51	+++
65	CUACAUACACguaacgcucca	(A) V/T L Q T T/A L A A	85	
66	AUGAUAUCCggcauucgggcc	(A) G I R/A N Y A L A A	69	
67	GUUACAGUCgacaagagcuau	(A) D K S Y N Y A L A A	96	
68	ACAUUAGUUgcaaccuccuug	(A)/A T/S L N Y A L A A	81	
69[f,g]	CGACGAUCUgacaagucaugu	(A) D K S C N Y A L A A	71	
70	AUGACAUACgucggucgaagc	(A) V G R S N Y A L A A	24	+++
71	GAAUUAGACgggcuccuccca	(A) G L L P N Y A L A A	90	
72	ACCUUAAUUgguguaccaacc	(A) G V P T N Y A L A A	9	++
73[f]	GUCACGCCGgacaacucugaa	(A) D N S E N Y A L A A	87	
74	UUUAGAGUUgcaacccuaucc	(A)/A T L S N Y A L A A	88	
75	CCUGUACAUggccaagucaug	(A) G Q V M N Y A L A A	31	
76[e]	CCAACACUAggcguauuaaug	(A)/G V/L M N/Y A L A A	93	

[a]Aligned by resume codon (first triplet in lower case) as determined by tagging assay. The shift for clone 65 is due to an extra C found inserted 6 nt downstream of the randomized block.
[b]Slashes mark proteolytic cleavage sites deduced from mass spectrometry of reporter localized in Tsp$^-$ periplasm, with tagging by *trans*-translation and in some cases by hard-coding.
[c]Comparison of mass spectrometry peaks indicative of tagging (see Materials and methods).
[d]Effectiveness in growth competition against cells producing no tmRNA.
[e]Primary tagging product produced by hard-coding, confirming proteolysis site mapping.
[f]Also tested in DD-tagging form with cytoplasmic reporter, confirming resume codon mapping.
[g]Cys-containing products exceeded expected mass by 305 Da, returning to expected mass on incubation with dithiothreitol, indicative of disulfide linkage to glutathione moiety.

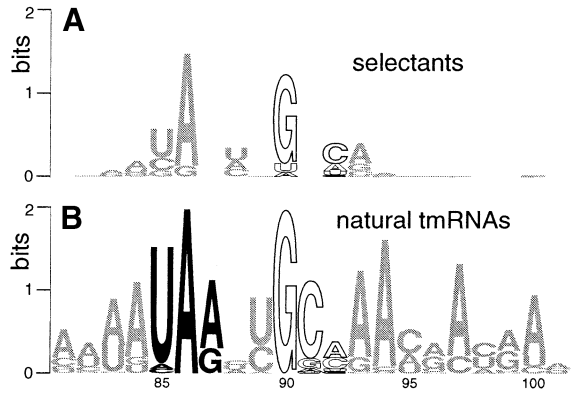

Fig. 5. Information content (Schneider and Stephens, 1990) of the randomized region, for sequences from (**A**) selectants or (**B**) natural tmRNAs.

(Keiler *et al*., 1996). With cells lacking tmRNA, this periplasmic reporter has a mass spectrometry profile that corresponds well with that observed using the cytoplasmic reporter, dominated by the run-off product from the last codon (Phe15) of the major primary transcript (Figures 6A and 2C). The profile of tagged periplasmic reporter produced by the DD mutant tmRNA also corresponded well with that observed for the cytoplasmic reporter; again the major peak was from proteins tagged after use of the Phe15 codon (data not shown).

Wild-type tmRNA produces a rather different profile in that the main peak of the profile corresponds to the last use of the Gly13 rather than the Phe15 codon (Figure 6A). We do not believe that the wild-type tmRNA recognizes the tagging substrate in a fundamentally different way to the DD mutant. We suspect that the difference is due to periplasmic proteases other than Tsp, which recognize the wild-type tag but not the DD tag and, furthermore, are sensitive to the sequence at the reporter–tag junction, such that reporter tagged after Phe15 is significantly less stable than that tagged after Gly13. Other notable peaks in the reporter profile from wild-type tmRNA might correspond to run-off at Ala8 or Met11 codons, but compared with a normal run-off profile their peak heights are exaggerated relative to their neighbors. Instead, we interpret them as representatives of tagging products that have been cleaved by proteases after Ala8 or Met11 to form metastable degradation products. Ala8 and Met11 cleavage products were also observed when the wild type-tagged periplasmic reporter was encoded traditionally on a single mRNA with a stop codon (data not shown).

This assay was applied to the selectant tmRNAs (Figure 6B and C; Table I). The masses of tagging products allowed the identification of the resume codon for each selectant; all selectant resume codons mapped to the same position as did the wild type. The new tag residues encoded by the randomized tmRNA codons frequently affected susceptibility to proteolysis, in a few cases altering

Table II. Resume region alignment and predicted tags for natural tmRNA sequences

Aae	GCUAAAACAgcucccgaagcu	(A)APEAE----------LALAA*
Tma	AGAAUAAGUgccaacgaaccu	(A)ANEP----------VAVAA**
Tth	GGCAUAACUgccaacaccaac	(A)ANTN----------YALAA**
Dra	CAUUUAACUggcaaccagaac	(A)GNQN----------YALAA*
Pgi	AGUUUAAUUggcgaaaauaac	(A)GENN----------YALAA*
Cte	CCAAUAGAUgcagacgauuau	(A)ADDYS----------YAMAA*
Cpn	ACAAUAAAUgccgaaccuaag	(A)AEPKAECEIISLFDSVE------ERLAA*
Ctr	ACAAUAAAUgccgaaccuaag	(A)AEPKAECEIISFADLED------LRVAA*
Cmu	ACAAUAAAUgccgaaccuaag	(A)AEPKAECEIISFADLND------LRVAA*
Nmu	AAAGUAAAUgcgaauaacauc	(A)ANNIVKFARKD----------ALVAA*
Sco	AAAGUAACUgcgaacaacauc	(A)ANNIVPFARKA----------APVAA*
Scy	AAAGUAACUgcgaauaacauc	(A)ANNIVSFKR----------VAIAA*
Gth	CAAAUAAAUgcaagcaauaua	(A)ASNIVSFSSKR----------LVSFA*
Ppu	UAAAUAAAUgcagaaaauaau	(A)AENNIIAFSR----------KLAVA*
Twe	CUAAUAAAUgcuaauaauaua	(A)ANNIIPFIFKAVKTKKEAMA---LNFAV*
Osi	AUAAUAAAUgcuaauaauuua	(A)ANNLISSVFKSLSTKQNSLN---LSFAV*
Cpa	AAUUUAAACgcaacuaauauu	(A)ATNIVRFNRK----------AAFAV**
Tpa	UUUAUAACUgccaauucugac	(A)ANSDSFD----------YALAA*
Bbu	AAAAUAAAUgcaaaaauauau	(A)AKNNNFTSSN----------LVMAA*
Ngo	CAAAUAGUCgcaaacgacgaa	(A)ANDET----------YALAA*
Nme	CAAAUAGUCgcaaacgacgaa	(A)ANDET----------YALAA*
Aeu	GCAAUAACUgcuaacgacgaa	(A)ANDER----------YAL-A*
Bpe	ACUACAAACgccaacgacgag	(A)ANDER----------LALAA*
Bbr	ACUACAAACgccaacgacgag	(A)ANDER----------FALAA*
Xfa	CUUAUAGUUgccaacgaagac	(A)ANEDN----------FAVAA*
Dno	CUUAUAGUUgcaaacgacgac	(A)ANDDN----------YALAA*
Ftu	AUAAUAACUggcaacaaaaaa	(A)GNKKANRVAANDSNFAA------VAKAA*
Lpn	AAUAUAAAUgcaaacgaugaa	(A)ANDENFAGG----------EAIAA**
Pae	CUUAUAGUUgccaacgacgac	(A)ANDDN----------YALAA*
Mhy	GUAAUAGUCgcaaacgacgaa	(A)ANDEN----------YALAA*
Pha	AAAGUAAUCgcaaacgacgau	(A)ANDDN----------YSLAA*
Spu	GUUAUAGUUgcaaacgacgau	(A)ANDDN----------YALAA*
Vch	AAAAUAGUCgcaaacgacgaa	(A)ANDEN----------YALAA*
Asa	AAAAUAGUCgcaaacgacgaa	(A)ANDEN----------YALAA**
Eco	AAAAUAGUCgcaaacgacgaa	(A)ANDEN----------YALAA**
Sty	AAAAUAGUCgcaaacgacgaa	(A)ANDET----------YALAA**
Ype	AAAAUAGUUgcaaacgacgaa	(A)ANDEN----------YALAA**
Aac	AAAAUAGUCgcaaacgacgaa	(A)ANDEQ----------YALAA**
Hin	AAAAUAGUCgcaaacgacgaa	(A)ANDEQ----------YALAA**
Dde	AAAGUAAUUgccaacaacgau	(A)ANNDYD----------YAYAA**
Hpy	AAAAUAACUguaaacaacaca	(A)VNNTDYAPA----------YAKAA*
Hpy	AAAAUAACUguaaacaacgca	(A)VNNADYAPA----------YAKAA*
Cje	AAAUUAAACgcaaacaacguu	(A)ANNVKFAPA----------YAKAA*
Mle	AAUAUAAGCgccgauucauau	(A)ADSYQRD----------YALAA*
Mtu	CAAAUAAGCgccgauucacau	(A)ADSHQRD----------YALAA*
Bme	AUAAUAACUggcaaaucuaac	(A)GKSNNN----------FALAA***
Bsu	AAUAUAACUggcaaaacuaac	(A)GKTNSFNQN----------VALAA**
Sau	AUAAUAACUggcaaaucaaac	(A)GKSNNN----------FAVAA*
Efa	AAUAUAACUgcuaaaaacgaa	(A)AKNENNS----------FALAA*
Spy	AAUAUAACUgcaaaaaauaca	(A)AKNTNS----------YALAA*
Smu	AAUAUAACUgcaaaaaauaca	(A)AKNTNS----------YAVAA*
Sgo	AAUAUAACUgcaaaaaauaau	(A)AKNNTS----------YALAA*
Spn	AAUAUAACUgcaaaaaauaac	(A)AKNNTS----------YALAA*
Mca	UAAAAAAACgcaaauaaaaac	(A)ANKNEETFEMPAFMMNNASAGA-NFMFA**
Uur	AUUUUAAAUgcagaaaauaaa	(A)AENKKSSEVELNPAFMASATNANYAFAY*
Mge	UCAAUAACCgacaaagaaaau	(A)DKENNEVLVEPNLIINQQASV--NFAFA*
Mge	UCAAUAACCgacaaagaaaau	(A)DKENNEVLVDPNLIINQQASV--NFAFA*
Mpn	AUAAUAACCgacaaaaauaac	(A)DKNNDEVLVDPMLIANQQASI--NYAFA*

Species names abbreviated as in the tmRNA Website (Williams, 1999).

the ratio of surviving reporter tagged after Gly13 to that tagged after Phe15, and also creating new cleavage sites in or near the altered tag positions. The new cleavage sites conferred by these tags were confirmed for some selectants by constructing plasmids that overproduce the tagging products (Table I) traditionally, from single unbroken mRNAs with stop codons. The three selectants that encode Cys residues produced reporters whose masses exceeded the expected values by 305 Da, suggesting a disulfide linkage to a glutathione moiety (Berger *et al.*, 1998). This interpretation was supported by the loss of the extra 305 Da for all Cys-containing protein species after treatment with dithiothreitol (DTT).

Apparent tagging activities ranged from near wild type to barely detectable (Table I, column T). It should be noted that the DD tmRNA profile indicated that this periplasmic reporter tagging assay was less sensitive to deficient activity than the cytoplasmic reporter assay; DD-tagged periplasmic protein was much more abundant than the run-off (data not shown), whereas with the cytoplasmic

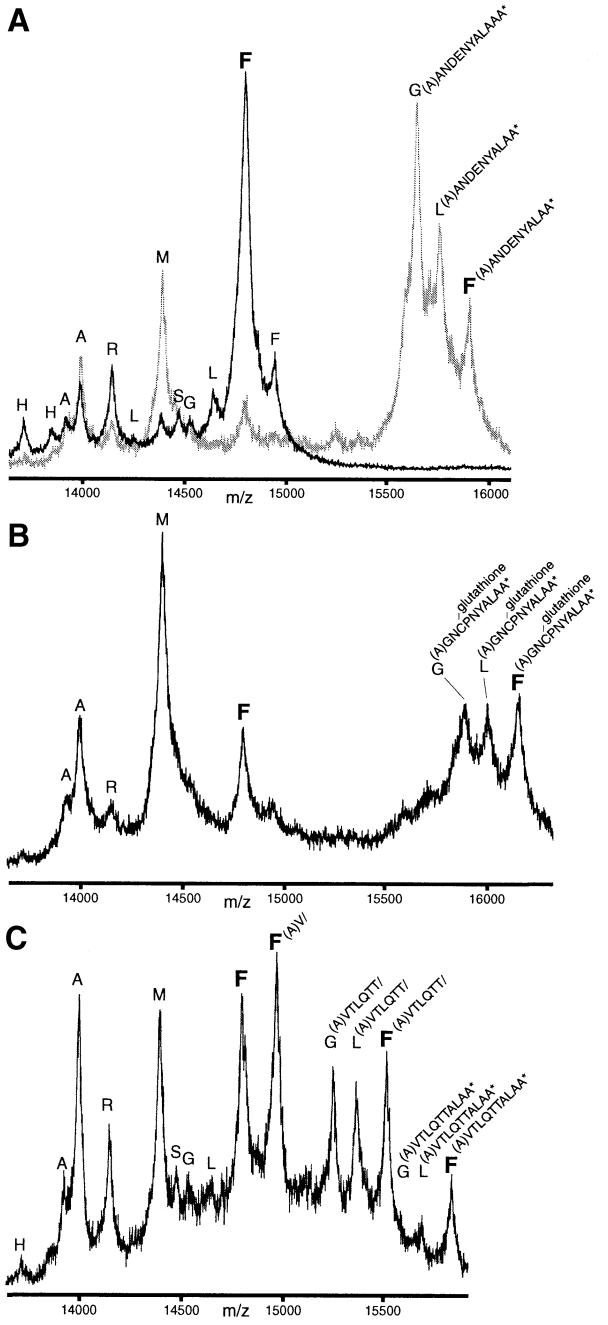

Fig. 6. Mass spectrometry of periplasmic reporter from *tsp*⁻ *E.coli*. Profiles for (**A**) no tmRNA (dark trace) and WT tmRNA (light trace), (**B**) selectant 54 and (**C**) selectant 65. The three peaks at the right of (B) were shifted to the left by 305 Da upon treatment with dithiothreitol. Slashes after tag sequences denote post-translational proteolytic products.

reporter system, tagged and run-off protein levels were equivalent (Figure 2B). This apparent increase in DD tmRNA activity may be correlated with lower periplasmic reporter protein levels or with modulation of the translating ribosome due to periplasmic transport (Powers and Walter, 1997).

Discussion

The recent appreciation of tmRNA structure and function has opened several intriguing avenues of research. The unusual mechanistic aspects of *trans*-translation are expected to reveal new insight into normal ribosome function. Our work focused on the question of how the resume codon of tmRNA is registered in the ribosomal A site. This line of research also revealed the state of the ribosome at the time that *trans*-translation begins and provided a new phylogenetic perspective, since the alignment of resume codons from natural tmRNA sequences now has an experimental basis.

New phylogenetic perspective

Resume codon alignment implicitly predicts the encoded tags (Table II). Tag reading frames range from nine to 28 codons in length. C-terminal hydrophobic/aromatic character is conserved; Ala accounts for 54% of the residues observed at the last five tag residues, and Leu, Tyr, Val and Phe account for another 37%. This hydrophobicity at the extreme C-terminus is known to be a major component of the signal for the several proteases that act on tagged proteins. At the other end of the tag reading frame, the resume codon is Ala in 81% of the tags, and Asn, Asp and Glu make up 71% of residues from the second to fourth tag codons. Conservation at this end may also contribute to recognition by proteases, but results from our randomization–selection experiment suggest a negative role: to prevent cleavage by these proteases within the tag itself, perhaps thereby favoring cleavage in the body of the tagged protein. Cleavage sites within the tag were readily generated upon randomizing only four codons at the upstream end of the tag reading frame. Such cleavage within the tag would defeat the proteolysis-inducing function of tagging if it rendered the product stable. Natural tag sequences may be evolutionarily constrained to prevent cleavage within themselves.

The alignment of tmRNA resume codons also allows phylogenetic evaluation of previous proposals concerning the resume region. A model for tmRNA secondary structure worked out by phylogenetic sequence comparison left the resume codon in the longest unpaired region of the molecule (Figure 1B; Williams and Bartel, 1996). While chemical probing studies of *E.coli* tmRNA supported all the pairings established by phylogenetic comparison, they indicated one additional pairing involving the resume codon itself (Felden *et al.*, 1997; Hickerson *et al.*, 1998) that might be expected to contribute to *trans*-translation. Mutations disrupting as many as two of its four proposed base pairs did not diminish tagging activity, suggesting that the pairing is not necessary for tmRNA function (Figure 3A and B). The alignment of natural resume codons reveals that this pairing is not well conserved (Table II). Likewise, phylogenetic analysis does not support either of two specific proposals for base-pairing interactions between the tmRNA resume codon region and the decoding region of the small subunit ribosomal RNA (Muto *et al.*, 1998). Some other rRNA–tmRNA interaction analogous to the Shine–Dalgarno interaction is possible; with our firm alignment of resume codons, careful phylogenetic consideration can now be part of the search.

The substrate for tmRNA action

Our data indicate that the ribosome runs to the last possible codon (Phe15) on a broken mRNA during run-

off translation and in many cases during *trans*-translation, suggesting that a signal for entry of tmRNA may be as simple as nucleotides missing in the ribosomal A site. This conclusion is especially clear when using the DD mutant tmRNA, which produces a tag that does not induce proteolysis. Analysis of the state of the ribosome acted upon by the wild-type tmRNA has been more difficult due to susceptibility of tagged reporter protein to multiple proteases. The wild-type tmRNA was assayed using a periplasmic reporter protein in cells lacking Tsp, the major periplasmic protease recognizing the wild-type tag. At face value, the mass spectrometry profile for this sample suggests that tagging occurs primarily after use of the antepenultimate (Gly13) codon, in contrast to the DD tmRNA result. A precedent for the use of mRNA-internal codons exists (Roche and Sauer, 1999). However, we consider it unlikely that the DD mutation would drastically affect substrate selection, and prefer the alternative explanation that the Phe15/WT tag-reporter is less stable than the Gly13/WT-tag form in the Tsp^- periplasm. No degradation products were observed for the traditionally encoded (on a single mRNA with a stop codon) Gly13/WT-tag protein in Tsp^- periplasm, while the Ala8 and Met11 metastable cleavage products were observed for the traditionally encoded Phe15/WT-tag protein (data not shown). Effects on proteolysis by sequence near the tag junction were also evident in reporter proteins tagged by selectant tmRNAs (Table I). Candidate periplasmic proteases for the residual activity observed in the absence of Tsp are DegP (HtrA) and DegQ (HhoA); overexpression of DegQ can complement defective *tsp* or *degP* alleles (Bass *et al.*, 1996; Kolmar *et al.*, 1996; Waller and Sauer, 1996).

Finding the resume codon

It is significant that all 23 selectants direct the ribosome to resume translation at the same position as in the wild type. A priori, other triplets could have been chosen as the resume codon; a total of seven triplets in the zero frame had been fully randomized. (Our system did not allow selection of tmRNAs directing the ribosome to either of the non-zero frames.) That none of the other six possible new resume codon sites were chosen clearly indicates that elements outside of the region we had randomized, remaining unidentified, act to restrict resume codon choice to a small window (probably no larger than 5 nt). These outside elements cannot be fully responsible for precise positioning however. One of our selectants (#65) had accrued a nucleotide insertion in the tag reading frame, which effectively placed an extra nucleotide between the upstream pseudoknot and the resume codon that was chosen (Table I). Likewise, our directed single-nucleotide insertion mutants demonstrate that sequence downstream from the first position of the resume codon is not essential for precise positioning (Figure 3A). These observations point to the segment between the first pseudoknot and the resume codon as the locus of features responsible for precise resume codon selection. Thus resume codon determinants in tmRNA are hierarchical; distant elements restrict resume codon choice to a narrow window, but local upstream features provide precision. Related information comes from a study where a single nucleotide was inserted or deleted downstream of the first pseudoknot with little effect on *trans*-translation (Nameki *et al.*, 1999b). The authors concluded that the first pseudoknot is not a determinant of the resume codon, which is clearly true for precise positioning, but this pseudoknot is still a good candidate for an element that could help restrict resume codon choice to a narrow window.

The resume-positioning segment (between the first pseudoknot and the resume codon) has the sequence AAAAAAUAGUCG in *E.coli*; its monotonic upstream portion seems less likely to act with precision than its downstream portion. It contains the base (A86, underlined) that we found to be invariant among natural tmRNAs and essential for *trans*-translation in the context of the wild-type resume region sequence. A86 was also present in 21 of our 23 selectants; it will be of interest to determine which features of the other two selectants permit A86 to vary. Conservation was also observed downstream of the resume codon, but since this region is translated, it is not clear whether conservation there reflects requirements at the level of RNA action or at the level of encoded tag function. The periodic pattern, with higher conservation at the first two bases in each codon than at the last base, suggests that coding may be the overriding function of this region.

How does the upstream segment act to position the resume codon? Entry of the tRNA-like domain into the P site by transpeptidation is likely to play a major general role in positioning. Contact between the upstream segment and the tRNA-like domain may automatically position the resume codon in the A site. Analysis of potential base-pairing co-variation has not yet revealed any intramolecular pairing for the resume region. Alternatively, the region may position the resume codon through interaction with extrinsic factors. Although phylogenetic analysis (above) rules out particular proposals that have been made for Shine–Dalgarno-like rRNA–tmRNA interactions, redoubled search efforts may eventually reveal such interaction. A possible clue to function comes from the nature of the most highly conserved positions in this region, which form a UAR sequence 2 nt upstream of the resume codon. UAA and UAG are recognized by RF-1 as stop codons during normal translation (Ryden *et al.*, 1986; Tate and Mannering, 1996), and the result may indicate that RF-1 assists in registering the resume codon in the ribosome. For RF-1 to function in this way in *trans*-translation, its normal peptidyl-tRNA hydrolase function would need to be suppressed; furthermore, the mechanism that would eventually translocate the resume codon into the A site is unclear. It would also be necessary to explain why the UAR is not perfectly conserved in nature (the U is altered in four out of 55 species) and why our selectants vary even more widely; one selectant (Table I, line 73) matches none of the UAR bases, yet tags quite well.

Practically any tRNA species (and C-terminal amino acid) may reside in the P site just prior to *trans*-translation; sequencing of 24 different tmRNA-tagged forms of an overproduced protein showed essentially no preference for any particular codon as the last to be used from the broken mRNA (Tu *et al.*, 1995). Then an entire peptidyl elongation cycle intervenes before the resume codon is utilized, as the uncoded Ala residue is added between the sequences coded by the broken mRNA and the tmRNA.

This uncoded elongation cycle can be viewed as a way to clear out the arbitrary peptidyl-tRNA left in the P site by the broken mRNA, so that the ribosome presents a uniform tRNA/C-terminal amino acid context in the P site (i.e. tmRNA/Ala) at the time that the resume codon is engaged. Such clearing of the context left by the broken mRNA may be important for reliable use of the resume codon.

With this uniformity of P-site filling by the uncoded elongation cycle, it might have been expected that the subsequent step of A-site filling would need to be uniform. Instead, altering the resume codon to different sense codons did not significantly diminish its efficacy and different tRNA species were used, following only the normal rules of the genetic code that apply to A site decoding. We conclude that the site of the resume codon is determined before the cognate tRNA is engaged. This contrasts fundamentally with the manner in which the first codon of the reading frame is recognized in normal mRNA, where a dedicated initiator tRNA species is employed in the P site. The difference may stem from the fact that the resume codon is utilized in the A site.

We can glean additional mechanistic insight from some of our other results. As expected, two single-base variants of active tmRNAs, A86U and G90U-C91A, disallow proper utilization of the resume codon, one by altering a positioning determinant and the other by converting the codon from sense to stop. These variants could, however, have been expected to allow the addition of the uncoded Ala to the reporter protein; instead, both were completely inactive. This may indicate that the resume codon engages its cognate tRNA before the tRNA-like function of tmRNA (addition of the uncoded Ala) can be fulfilled. Combining this idea with the conclusion that the resume codon is determined before engaging the tRNA would suggest that such determination is a very early step in *trans*-translation, conceivably occurring outside the ribosome.

The unexpected frameshifted tagging product observed for the mutant with a stop codon following the resume codon (G90U-A93U-C95A, Figure 3E) could, in principle, have resulted from direct initial use of the –2 Ser codon as the resume codon. However, both the resume-to-stop mutant G90U-C91A and the parental G90U mutant failed to make use of the identical –2 Ser codon. We therefore believe that $tRNA^{Ser}$ initially recognized the zero-frame Ser resume codon, then, faced with an in-frame stop codon, an appreciable fraction of ribosomes underwent –2 slippage to the nearest codon available for repairing. Frameshift by two nucleotides is only rarely observed, suggesting that the resume codon may be less stably positioned in the P site than a typical mRNA codon. Further experiments will be necessary to confirm these interpretations.

Detailed mechanistic analysis of *trans*-translation will require the development of a clean *in vitro* system, so that the complexes along the pathway can be separately prepared. Encouragingly, *in vitro trans*-translation has been reported (Himeno *et al.*, 1997; Roche and Sauer, 1999) and a protein required for tmRNA function (SmpB) has been identified (Karzai *et al.*, 1999). It will be of great interest to elucidate additional protein factors required and how they might interact to promote the special events of *trans*-translation.

Materials and methods

In vivo tagging assays

Escherichia coli was doubly transformed, with a low-copy (p15A replicon) plasmid bearing a variant of the *E.coli* tmRNA gene under control of the native promoter and terminator, and with a high-copy (ColE1 replicon without Rom) plasmid encoding under control of the IPTG-inducible *tac* promoter either cytoplasmic phage lambda repressor N-terminal fragment or periplasmic cytochrome b562, followed by the M2 epitope, a His_6 unit and the *trp* attenuator (Keiler *et al.*, 1996), or C-terminal variants of these placing a stop codon upstream of the terminator. Doubly transformed cells were grown at 37°C in Luria–Bertani (LB) broth with 60 μg/ml ampicillin, 50 μg/ml tetracycline to an A_{600} of 0.3–0.4, then IPTG was added to 0.75 mM and growth was continued for 1 h. The cytoplasmic reporter was used in *ssrA*$^-$ *E.coli* X91 (Keiler *et al.*, 1996) in a 50 ml culture; the periplasmic reporter was used in *ssrA*$^-$ *tsp*$^-$ *E.coli* X94, constructed by P1 transduction from KS1000 (Silber and Sauer, 1994) into X91, in an 800 ml culture. Harvested cells were lysed in 6 M guanidinium, 0.1 M phosphate, 0.01 M Tris pH 8, 0.01 M imidazole and reporter protein was purified in batch on Ni-NTA–agarose (Qiagen, 1 ml resin/l culture). Reporter protein was dialyzed into 0.01 M Tris pH 7.5, 50 mM NaCl, desalted on a C18 Sep-Pak cartridge (Waters), resuspended in 50 μl H_2O and analyzed by Coomassie Blue staining of a Laemmli gel or by MALDI-TOF mass spectrometry with myoglobin as an internal standard. Although mass spectrometry is notoriously non-quantitative when comparing unrelated molecules, we contend that neighboring proteins in these profiles are similar enough in character that peak areas should be rough indicators of relative abundance. For the periplasmic reporter, the fraction of the sum of peak heights for all tagged products and for Ala8 and Met11 (peaks due mainly to cleavage of tagged products) to the same sum plus Phe15 run-off was computed, and then scaled, after setting the values for no tmRNA (0.27) and for wild-type tmRNA (0.95) to 0 and 100, respectively. This is an imperfect index of the extent of tagging (T), since some degraded products of tagging may escape detection.

Genetic selection/screen for active tmRNA

The selective plasmid p6A was constructed in the vector used for the reporter plasmids above, with the phage P22 Arc repressor gene under control of the *tac* promoter and lacking a stop codon but tailed directly by a transcriptional terminator; the kanamycin-resistance gene from pACYC177 was also included, under control of the Arc-sensitive phage P22 promoter Pant. tmRNAs were tested as clones in the low-copy vector by transforming together with p6A into *E.coli* X91. Without tmRNA activity, the run-off repressor is stable and active, preventing KmR expression; active tmRNA causes tagging and proteolysis of the repressor, so that KmR is expressed. Standard selective plates were LB agar with 60 μg/ml ampicillin, 50 μg/ml tetracycline, 100 μg/ml kanamycin and 100 μM IPTG. Activities of tmRNA could also be compared roughly using a series of plates with varied kanamycin and IPTG levels.

To ensure selection only of tmRNAs that could direct the ribosome to the original tag reading frame, a mutant tmRNA was constructed that generated codons for charged residues at the ends of the other two reading frames and tested positive for tmRNA activity in the genetic screen. Variants that shifted tagging to each of the other two reading frames in this mutant, and in the wild-type tmRNA, were constructed by inserting or deleting a single adenosine residue within a block of four adenosines in the center of the tag reading frame. These variants testing undesired reading frames were all negative in the genetic screen except for the deletion variant of the wild-type tmRNA, which was weakly positive. The mutant was therefore improved over the wild type for this genetic purpose and was used as the parent for pool construction.

Randomization–selection

The oligonucleotides GTTGgcctcgtaaaaagccgcaaN_{21}aactacgcgttagcagctt-aaggtaactgcttagcgc and GAGGgcgctaagcagttaccttaagctgctaacgcgtagttN_{21}t-tgcggctttttacgaggc (complementary regions in lower case, N_{21}: randomized resume region, Figure 4) were hybridized at equimolar concentration and ligated into the low-copy tmRNA vector linearized with *Bsa*I to form the appropriate 5′ overhangs. The ligation mix was desalted and electroporated into *E.coli* X91 that had been previously transformed with p6A. The fraction of plasmids passing the genetic screen was measured directly by plating transformants without kanamycin selection, then replica plating 2451 colonies onto selective plates, which yielded 18 survivors. The tmRNA plasmid from each survivor was retested by

retransforming into fresh selective cells after inactivating the copurified selective plasmid with restriction enzymes, and 10 retested positive. This test indicated that a high background operates in the genetic system as implemented, but also that the fraction of plasmids producing tmRNA activity is high enough (4×10^{-3}) that it is not overwhelmed by the background. Thirteen additional selectants retesting positive were included for sequencing and assay.

Sequence alignments and information content

Resume region sequences were available for 55 species from 10 phyla at the tmRNA website <http://sunflower.bio.indiana.edu/~kwilliam/tmRNA/home.html>. The downstream portions of their tag reading frames were readily identified because the encoded amino acids were conserved in sequence and hydrophobic character. Resume codons in these reading frames were predicted using the rule suggested from our randomization–selection experiment, that a G should occur at the first position of the resume codon with an A 4 nt upstream. Segments from the first pseudoknot to the tag stop codon were examined. All had at least one fit to the in-frame AnnnG rule, but many had multiple possible fits. Eliminating these left 23 sequences from six phyla with unambiguous matches (the *E.coli* sequence was also included in this group because its resume codon is precisely known). To overcome phylogenetic sampling bias, a phylogenetic tree based on 16S rRNA sequences was generated using the Subtree program at the Ribosomal Database Project <http://rdpwww.life.uiuc.edu/index2.html>, and shared branches in the tree were split to produce a weighting factor for each species. The sequences were aligned according to the AnnnG rule match, without gaps, and positions where a base was conserved with a frequency of at least 0.9 (upper case) or at least 0.75 (lower case) produced the consensus AUAAnuGcnaAnnanna. One best match to this consensus was found for each of the remaining 32 sequences, producing the alignment shown in Table II. A sequence logo (Schneider and Stephens, 1990) that presents the information content of aligned sequences was prepared at the web page <http://www.bio.cam.ac.uk/seqlogo> of S.Brenner using the weighted aligned list; one generated without weighting did not differ substantially.

Acknowledgements

We thank Robert Sauer and Eric Roche for useful discussions and access to key materials and to results prior to publication. We thank Jim Drummond and Eric Roche for helpful comments on the manuscript.

References

Atkins,J.F., Böck,A., Matsufuji,S. and Gesteland,R.F. (1999) Dynamics of the genetic code. In Gesteland,R.F., Cech,T.R. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 637–673.

Bass,S., Gu,Q. and Christen,A. (1996) Multicopy suppressors of *prc* mutant *Escherichia coli* include two HtrA (DegP) protease homologs (HhoAB), DksA and a truncated R1pA. *J. Bacteriol.*, **178**, 1154–1161.

Berger,S.J., Claude,A.C. and Melançon,P. (1998) Analysis of recombinant human ADP-ribosylation factors by reversed-phase high-performance liquid chromatography and electrospray mass spectrometry. *Anal. Biochem.*, **264**, 53–65.

Bertrand,K., Korn,L.J., Lee,F. and Yanofsky,C. (1977) The attenuator of the tryptophan operon of *Escherichia coli*. Heterogeneous 3′-OH termini *in vivo* and deletion mapping of functions. *J. Mol. Biol.*, **117**, 227–247.

Farabaugh,P.J. (1996) Programmed translational frameshifting. *Annu. Rev. Genet.*, **30**, 507–528.

Felden,B., Himeno,H., Muto,A., McCutcheon,J.P., Atkins,J.F. and Gesteland,R.F. (1997) Probing the structure of the *Escherichia coli* 10Sa RNA (tmRNA). *RNA*, **3**, 89–103.

Felden,B., Hanawa,K., Atkins,J.F., Himeno,H., Muto,A., Gesteland,R.F., McCloskey,J.A. and Crain,P.F. (1998) Presence and location of modified nucleotides in *Escherichia coli* tmRNA: structural mimicry with tRNA acceptor branches. *EMBO J.*, **17**, 3188–3196.

Gesteland,R.F. and Atkins,J.F. (1996) Recoding: dynamic reprogramming of translation. *Annu. Rev. Biochem.*, **65**, 741–768.

Hartz,D., McPheeters,D.S., Traut,R. and Gold,L. (1988) Extension inhibition analysis of translation initiation complexes. *Methods Enzymol.*, **164**, 419–425.

Hickerson,R.P., Watkins-Sims,C.D., Burrows,C.J., Atkins,J.F., Gesteland,R.F. and Felden,B. (1998) A nickel complex cleaves uridine in folded RNA structures: application to *E.coli* tmRNA and related engineered molecules. *J. Mol. Biol.*, **279**, 577–587.

Himeno,H., Sato,M., Tadaki,T., Fukushima,M., Ushida,C. and Muto,A. (1997) *In vitro trans* translation mediated by alanine-charged 10Sa RNA. *J. Mol. Biol.*, **268**, 803–808.

Huang,W.M., Ao,S.Z., Casjens,S., Orlandi,R., Zeikus,R., Weiss,R., Winge,D. and Fang,M. (1988) A persistent untranslated sequence within bacteriophage T4 DNA topoisomerase gene 60. *Science*, **239**, 1005–1012.

Karzai,A.W., Susskind,M.M. and Sauer,R.T. (1999) SmpB, a unique RNA-binding protein essential for the peptide-tagging activity of SsrA (tmRNA). *EMBO J.*, **18**, 3793–3799.

Keiler,K.C., Waller,P.R. and Sauer,R.T. (1996) Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science*, **271**, 990–993.

Kolmar,H., Waller,P.R. and Sauer,R.T. (1996) The DegP and DegQ periplasmic endoproteases of *Escherichia coli*: specificity for cleavage sites and substrate conformation. *J. Bacteriol.*, **178**, 5925–5929.

Komine,Y., Kitabatake,M., Yokogawa,T., Nishikawa,K. and Inokuchi,H. (1994) A tRNA-like structure is present in 10Sa RNA, a small stable RNA from *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **91**, 9223–9227.

Muto,A., Ushida,C. and Himeno,H. (1998) A bacterial RNA that functions as both a tRNA and an mRNA. *Trends Biochem. Sci.*, **23**, 25–29.

Nameki,N., Felden,B., Atkins,J.F., Gesteland,R.F., Himeno,H. and Muto,A. (1999a) Functional and structural analysis of a pseudoknot upstream of the tag-encoded sequence in *E.coli* tmRNA. *J. Mol. Biol.*, **286**, 733–744.

Nameki,N., Tadaki,T., Muto,A. and Himeno,H. (1999b) Amino acid acceptor identity switch of *Escherichia coli* tmRNA from alanine to histidine *in vitro*. *J. Mol. Biol.*, **289**, 1–7.

Powers,T. and Walter,P. (1997) Co-translational protein targeting catalyzed by the *Escherichia coli* signal recognition particle and its receptor. *EMBO J.*, **16**, 4880–4886.

Roche,E.D. and Sauer,R.T. (1999) SsrA-mediated peptide tagging caused by rare codons and tRNA scarcity. *EMBO J.*, **18**, 4579–4589.

Ryden,M., Murphy,J., Martin,R., Isaksson,L. and Gallant,J. (1986) Mapping and complementation studies of the gene for release factor 1. *J. Bacteriol.*, **168**, 1066–1069.

Schneider,T.D. and Stephens,R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, **18**, 6097–6100.

Silber,K.R. and Sauer,R.T. (1994) Deletion of the *prc* (*tsp*) gene provides evidence for additional tail-specific proteolytic activity in *Escherichia coli* K-12. *Mol. Gen. Genet.*, **242**, 237–240.

Steitz,J.A. (1975) Ribosome recognition of initiator regions in the RNA bacteriophage genome. In Zinder,N. (ed.), *RNA Phages*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 319–352.

Tate,W.P. and Mannering,S.A. (1996) Three, four or more: the translational stop signal at length. *Mol. Microbiol.*, **21**, 213–219.

Tu,G.F., Reid,G.E., Zhang,J.G., Moritz,R.L. and Simpson,R.J. (1995) C-terminal extension of truncated recombinant proteins in *Escherichia coli* with a 10Sa RNA decapeptide. *J. Biol. Chem.*, **270**, 9322–9326.

Ushida,C., Himeno,H., Watanabe,T. and Muto,A. (1994) tRNA-like structures in 10Sa RNAs of *Mycoplasma capricolum* and *Bacillus subtilis*. *Nucleic Acids Res.*, **22**, 3392–3396.

Waller,P.R. and Sauer,R.T. (1996) Characterization of *degQ* and *degS*, *Escherichia coli* genes encoding homologs of the DegP protease. *J. Bacteriol.*, **178**, 1146–1153.

Weiss,R.B., Huang,W.M. and Dunn,D.M. (1990) A nascent peptide is required for ribosomal bypass of the coding gap in bacteriophage T4 gene 60. *Cell*, **62**, 117–126.

Williams,K.P. (1999) The tmRNA website. *Nucleic Acids Res.*, **27**, 165–166.

Williams,K.P. and Bartel,D.P. (1996) Phylogenetic analysis of tmRNA secondary structure. *RNA*, **2**, 1306–1310.

Received July 6, 1999; revised and accepted August 6, 1999