

Presence and location of modified nucleotides in *Escherichia coli* tmRNA: structural mimicry with tRNA acceptor branches

Brice Felden¹, Kyoko Hanawa²,
John F. Atkins³, Hyouta Himeno²,
Akira Muto², Raymond F. Gesteland^{1,3},
James A. McCloskey^{4,5} and Pamela F. Crain^{5,6}

¹Howard Hughes Medical Institute, University of Utah, Salt Lake City, UT 84112, USA, ²Department of Biology, Faculty of Science, Hirosaki University, Hirosaki 036, Japan and Departments of ³Human Genetics, ⁴Biochemistry and ⁵Medicinal Chemistry, University of Utah, Salt Lake City, UT 84112, USA

⁶Corresponding author
e-mail: pamelacrain@rna.pharm.utah.edu

Escherichia coli tmRNA functions uniquely as both tRNA and mRNA and possesses structural elements similar to canonical tRNAs. To test whether this mimicry extends to post-transcriptional modification, the technique of combined liquid chromatography/electrospray ionization mass spectrometry (LC/ESIMS) and sequence data were used to determine the molecular masses of all oligonucleotides produced by RNase T₁ hydrolysis with a mean error of 0.1 Da. Thus, this allowed for the detection, chemical characterization and sequence placement of modified nucleotides which produced a change in mass. Also, chemical modifications were used to locate mass-silent modifications. The native *E. coli* tmRNA contains two modified nucleosides, 5-methyluridine and pseudouridine. Both modifications are located within the proposed tRNA-like domain, in a seven-nucleotide loop mimicking the conserved sequence of T loops in canonical tRNAs. Although tmRNA acceptor branches (acceptor stem and T stem-loop) utilize different architectural rules than those of canonical tRNAs, their conformations in solution may be very similar. A comparative structural and functional analysis of unmodified tmRNA made by *in vitro* transcription and native *E. coli* tmRNA suggests that one or both of these post-transcriptional modifications may be required for optimal stability of the acceptor branch which is needed for efficient aminoacylation.

Keywords: electrospray mass spectrometry/5-methyluridine/modified nucleotides/pseudouridine/tmRNA

Introduction

tmRNA (10Sa RNA) is unique since it functions both as tRNA and as mRNA (e.g. Himeno *et al.*, 1997; Williams and Bartel, 1998; Zweib *et al.*, 1998). It is charged with alanine (Komine *et al.*, 1994; Ushida *et al.*, 1994) and also has a short reading frame coding for 10 amino acids (Tu *et al.*, 1995). This 11 amino acid tag (alanine plus the 10 encoded amino acids) is added to polypeptides trans-

lated from mRNAs lacking a termination codon, and the added 11 amino acid C-terminal tag makes the protein a target for specific proteolysis (Keiler *et al.*, 1996). Figure 1 summarizes the current mechanistic view of the function of *Escherichia coli* tmRNA: the elimination of aberrant proteins derived from truncated mRNAs and the recycling of stalled ribosomes. Recent structural analyses based on phylogenetic (Felden *et al.*, 1996b; Williams and Bartel, 1996) and probing (Felden *et al.*, 1997) data have led to a secondary structure with similarities to canonical tRNAs. The tRNA-like mimicry immediately raises the question of whether post-transcriptional nucleoside modifications, which are a ubiquitous feature of tRNAs (McCloskey and Crain, 1998), are present in *E. coli* tmRNA.

An early report raised the possibility that a study of this question might prove productive. Ray and Apirion (1979) detected pseudouridine (Ψ) at an unidentified position in an 11-mer obtained from RNase T₁ digestion of '10S' RNA (thought to be 550 to 609 nt) from *E. coli*. This RNA was subsequently shown to consist of RNase P RNA and what is now called tmRNA (Jain *et al.*, 1982). Here we analyze the 363 nt *E. coli* tmRNA using a method for the detection and sequence location of post-transcriptionally modified nucleotides in RNA. It is based on the accurate measurement of the molecular mass of oligonucleotides produced by selective enzymatic cleavage, as well as chemical probing for detection of Ψ , a mass-silent modification. An outline of the protocol is shown in Figure 2. This allows analysis of all the RNA fragments produced by specific enzymatic cleavage (e.g. RNase T₁ digestion) directly by liquid chromatography/electrospray ionization mass spectrometry (LC/ESIMS), without any intervening purification (to remove enzyme and salts) or fractionation of the oligonucleotide mixture. This makes the procedure easy (one step), reliable and applicable to any RNAs ≤ 363 nt in length. Two modified nucleosides, Ψ (at two positions) and 5-methyluridine (m^5U), have been identified and located. Compositions of the sequences of the RNase T₁ fragments have all been characterized by their molecular masses, indicating that there are apparently no other modifications that cause a mass change. The biological relevance of the presence of these nucleoside modifications in *E. coli* tmRNA is discussed.

Results

Native and engineered E. coli tmRNAs

tmRNA is present in *E. coli* cells in low abundance (~1000 copies/cell; Lee *et al.*, 1978), and we anticipated difficulty in obtaining a sufficient amount of pure RNA for *de novo* analysis. Accordingly, we used an overproduction system that yielded a 100-fold greater amount of properly processed *E. coli* tmRNA (from a plasmid containing the

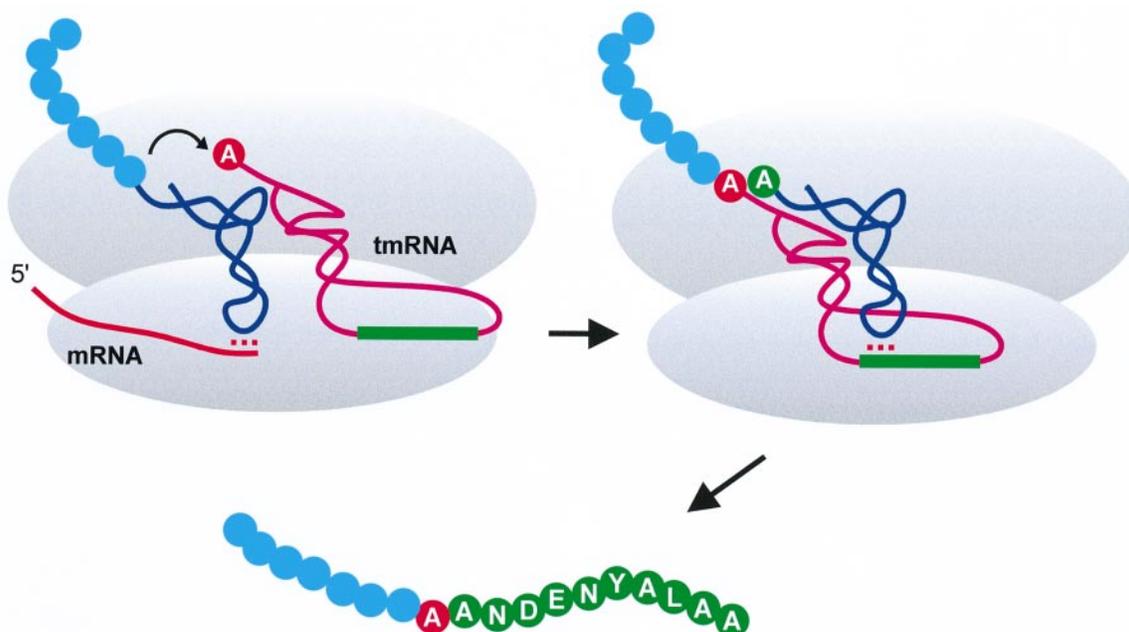


Fig. 1. Schematic illustration of the proposed biological role of tmRNA in *E. coli*. The mechanism allows the degradation of aberrant proteins and ribosome recycling via trans-translation (Keiler *et al.*, 1996).

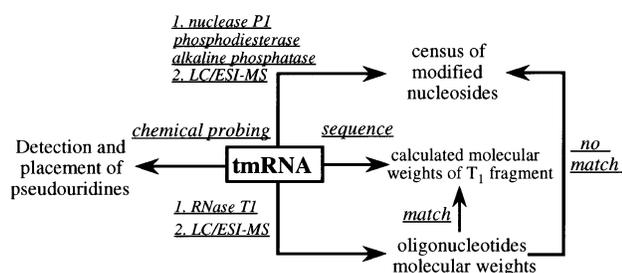


Fig. 2. Protocol used for the detection and localization of post-transcriptionally modified nucleotides in *E. coli* tmRNA.

inducible *E. coli* tmRNA gene expressed from a T7 promoter) to generate sufficient material for preliminary structural studies. Once the experimental conditions were optimized with the overexpressed tmRNA, a native tmRNA was analyzed. A T7 transcript was also generated *in vitro* for use as an unmodified negative control, and to devise a gradient for reversed-phase separation of oligonucleotides from an RNase T₁ digest of the 363-mer.

Total nucleoside analysis of *E. coli* tmRNAs by LC/ESIMS

The three tmRNAs (overexpressed, native and T7 transcript) were quantitatively hydrolyzed to their nucleoside constituents using nuclease P1, phosphodiesterase I and alkaline phosphatase (Crain, 1990). The hydrolysate from the overexpressed tmRNA was initially screened by LC/ESIMS for the presence of any modified nucleosides. In both native and overexpressed RNAs, two were found: Ψ and m^5U . Their structure assignments (Pomerantz and McCloskey, 1990) were established from their electrospray ionization (ESI) mass spectra (Figure 3), and their relative chromatographic retention times (chromatogram not shown, but similar to the one shown in Figure 4).

Mass spectra of nucleosides consist of two main ion types. One represents the protonated nucleoside molecule

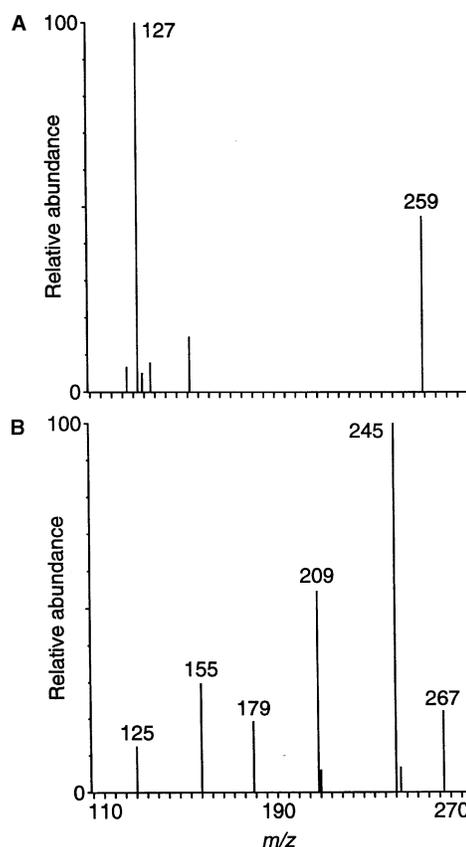


Fig. 3. ESI mass spectra of m^5U (A) and Ψ (B), acquired from a total nucleoside digest of overexpressed *E. coli* tmRNA.

(MH⁺), and another corresponds to the protonated free base of the nucleoside (BH), formed by cleavage of the N-C glycosidic bond with transfer of one hydrogen from the sugar to the base (B) (Pomerantz and McCloskey, 1990). These ions have a mass-to-charge ratio (m/z) of

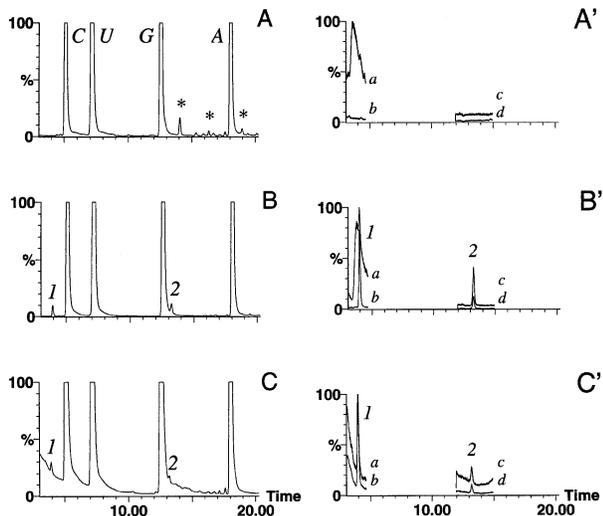


Fig. 4. Detection of Ψ and m^5U in overexpressed and native *E. coli* tmRNAs and the corresponding synthetic tmRNA (negative control). Left panels are the traces from UV detection at 254 nm for the synthetic tmRNA (A), the overexpressed tmRNA (B) and native tmRNA (C). Peaks in (A) marked with asterisks are nucleosides from traces of DNA template in the digest. Nucleoside identities are Ψ (1) and m^5U (2). Right panels show traces from selected ion recordings of ions characteristic of Ψ (a, m/z 209 and b, m/z 245) and m^5U (c, m/z 127 and d, m/z 259) in synthetic tmRNA (A'), the overexpressed tmRNA (B') and native tmRNA (C').

259 and 127, respectively, for m^5U (Figure 3A). However, Ψ contains a C–C glycosidic bond, preventing formation of the BH ion. Instead, a prominent ion (m/z 209) is produced by concerted loss of two water molecules from the sugar moiety (Figure 3B). This ion loses formaldehyde (the 5' substituent) to give m/z 179. Other ions in the spectrum of Ψ consist of the uracil moiety and fragments derived from the ribose: m/z 125 (B + 14) and m/z 155 (B + 44) (McCloskey, 1974). The m/z 267 ion is the Na^+ adduct of M.

For detection of Ψ and m^5U in the native tmRNA preparation, where the amount of material was limited, a selected ion-recording analysis was performed. Four ions, two each specific for Ψ (m/z 245 and 209) and m^5U (m/z 259 and 127), along with the UV channel, were monitored throughout the LC/ESI-MS analysis of the T7 transcript, the overexpressed tmRNA and native tmRNA (Figure 4). As expected, neither Ψ nor m^5U is present in the T7 transcript (Figure 4A and A'); however, trace amounts of the DNA template are evident (asterisks in Figure 4A). Both the overexpressed (Figure 4B and B') and native (Figure 4C and C') tmRNAs contain Ψ (peak 1) and m^5U (peak 2), evident from the coincident tracks for two diagnostic ions and a UV peak, eluting at the expected time for each nucleoside. Based on the lack of similar sized peaks in the UV detection chromatogram, we conclude that no other UV-absorbing modified nucleosides are present in tmRNA.

Analysis of *E. coli* tmRNA RNase T₁ digests by LC/ESI-MS

Oligonucleotides containing a nucleoside modification that yields a mass shift are recognized from mass increments (e.g. 14 Da for one methyl group) above the measured relative molecular mass (M_r) values for the corresponding

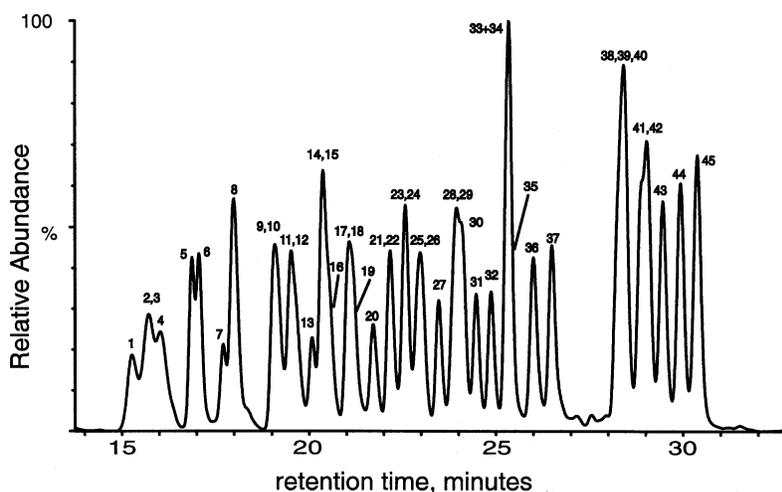
unmodified oligonucleotide. Consequently the latter will not be observed unless partial modification occurs, or there are multiple occurrences of the sequence (or isomers) within the RNA. To locate the m^5U , we used RNase T₁ (selectively cleaves after Gp) to totally digest all three tmRNAs. The primary sequence of *E. coli* tmRNA (Komine *et al.*, 1994) contains 102 RNase T₁ sites, resulting in 103 oligonucleotide fragments: 30 Gps, 17 dinucleotides (3 sequences), 18 trinucleotides (8 sequences), thirteen 4-mers (11 sequences), seven 5-mers, six 6-mers, three 7-mers, one 8-mer, two 9-mers, and one each 10-mer through 15-mer. Each of the oligonucleotide mixtures was analyzed directly by LC/ESI-MS without prior purification. The resulting chromatogram from the C-18 reversed phase fractionation, monitored by UV absorbance, is shown in Figure 5 for trimer and larger oligonucleotides from the RNase T₁ digest of overexpressed tmRNA (other chromatograms not shown). Oligonucleotide identities were established from ESI mass spectra acquired continuously every 2 s throughout the analyses, and summed across each peak. Oligonucleotides are numbered by their elution order in Figure 5. The elution order for the oligonucleotides is predominantly based on hydrophobicity, with A-rich oligonucleotides eluting later than pyrimidine-rich ones.

Determination of molecular masses and base compositions of oligonucleotides from *E. coli* tmRNA by LC/ESI-MS

By accurate determination of mass, we demonstrate that a composition can be assigned to each oligonucleotide within the mixtures derived from RNase T₁ digestion of native and overexpressed *E. coli* tmRNAs and the corresponding synthetic RNA. The compositions and sequence locations, and calculated and measured M_r values, are listed in Table I for all 7-mer and longer oligonucleotides. There are multiple occurrences of the same sequences or compositions at the hexamer and shorter lengths, so they are not included, although they were assigned (Figure 5). All compositions could be assigned either directly from the mass spectra if they were unique, or from integrated peak heights for sequence isomers. This demonstrates that a complex mixture of RNA oligonucleotides can be accurately analyzed by the method reported here. Mean mass errors were 0.1–0.2 Da (range 0–0.4 Da) for the three sets of T₁ oligonucleotides. The greatest error, 0.4 Da for the M_r 4808.9 oligonucleotide from the T7 transcript, for example, corresponds to a mass measurement error of $\leq 0.008\%$, which can routinely be obtained in the quadrupole mass analyzer utilized for these analyses.

Sequence placement of modified nucleotides from *E. coli* tmRNAs

Based on the *E. coli* tmRNA sequence (Komine *et al.*, 1994), the predicted M_r values of all of the oligonucleotides generated from RNase T₁ digestion were calculated, along with the corresponding multiply charged ions of the type $(M - nH)^{n-}$ characteristic of ESI (Fenn *et al.*, 1989). All of the predicted oligonucleotides were identified in the T7 transcript (negative control), and in the overexpressed and native tmRNA (Table I) with one exception. Indeed, the ions m/z 1154.3, 865.5 and 692.2, calculated for the –3,



1. UGp	10. UUUGp	19. CAAGp	28. CCUCCGp	37. UAAAAAGp
2. CUGp	11. UCCGp	20. AAUGp	29. CUCCACCA-OH	38. AAAACUACGp
3. UCGp	12. CCUGp	21. CCUCGp	30. CUUAGp	39. AAACCAAGp
4. CCGp	13. UUAGp	22. CCCUGp	31. AUUCUGp	40. UUCAACUCCCGp + CH ₃
5. UAGp	14. CAUGp	23. UACCGp	32. CAAACGp	41. CAAAAAUAGp
6. CAGp	15. CUAGp	24. CUUAGp	33. AUCAAAGp	42. CUUAAUAACCUg
7. AUGp	16. CCAg	25. AUUCGp	34. ACUAAGp	43. CCCUCUCCCUAGp
8. ACGp	17. AUCCg	26. AUUUGp	35. CUCUAGp	44. UCAAACCAAAAAGp
9. AAGp	18. ACUGp	27. UAAAGp	36. AAUUCGp	45. UAAAACUAAUCAGp

Fig. 5. Chromatogram (detection at 254 nm) from LC/ESIMS analysis of an RNase T₁ digest of 50 pmol of overexpressed *E. coli* tmRNA. Peak numbers in the chromatogram correspond to the oligonucleotide identities shown below.

Table I. Mol. wt of 7-mers and longer from RNase T₁ digestion of native and overexpressed *E. coli* tmRNAs and the corresponding synthetic RNA

Oligonucleotides ^a	Position	Calculated M_r	Measured M_r		
			Native tmRNA	Overexpressed tmRNA	T7 transcript
CUCUAGp (35)	157–163	2221.3	2221.0	2221.2	2221.3
AAUUUCGp (36)	326–332	2245.3	2245.3	2245.1	2245.5
UAAAAAGp (37)	68–74	2315.4	2315.4	2315.3	2315.5
CUCCACCA-OH (29)	356–363	2428.8	2428.5	2428.5	428.6
AAACCAAGp (39)	32–40	2924.8	2924.6	2924.9	2924.9
AAAACUACGp (38)	100–108	2925.8	2925.8	2925.8	2926.0
CAAAAAUAGp (41)	78–87	3279.0	3279.1	3278.9	3279.3
UUCAACUCCGp	341–351	3466.1	n.d.	n.d.	3466.5
UUCAACUCCCGp + CH ₃ (40)	341–351	3480.1	3480.1	3479.9	n.d.
CUUAAUAACCUg (42)	118–129	3820.3	3820.4	3820.3	3820.4
UCAAACCAAAAAGp (44)	182–194	4194.6	4194.4	4194.3	4194.8
CCCUCUCCCUAGp (43)	137–150	4358.6	4358.5	4358.4	4358.9
UAAAACUAAUCAGp (45)	229–243	4808.9	4809.0	4808.6	4809.3

^aPeak numbers are those shown in Figure 5.
n.d., not detected.

−4 and −5 charge states, respectively, of the oligonucleotide ₃₄₁UUCAACUCCCG₃₅₁ (calculated M_r 3466.1), were not found in overexpressed and native tmRNAs, although this oligonucleotide is present in the T7 transcript (Table I). In addition, an oligonucleotide whose ESI mass spectrum was consistent with that of a species with a calculated M_r of 3480.1 ± 0.2 , was identified in overexpressed and native tmRNAs. The ESI mass spectrum of the corresponding oligonucleotide from the T₁ digest of native tmRNA is shown in Figure 6. The ions m/z 695.1, 869.1 and 1159.0 all track together (Figure 6, inset), and are assigned to an oligonucleotide of M_r 3480.1, eluting in the right shoulder of the UV peak centered at ~28.7 min. Calculated mass versus oligonucleotide composition tables (Pomerantz *et al.*, 1993) list only one candidate composition within

± 0.6 Da of the measured molecular mass: C₅U₃A₂Gp + CH₃ (M_r 3480.1). This result suggests that the m⁵U present in both native and overexpressed tmRNAs (Figure 4) is located within the oligonucleotide ₃₄₁UUCAACUCCCG₃₅₁. No other oligonucleotides were found to contain a methyl group.

Another modified nucleoside, Ψ, is also present in both native and overexpressed tmRNAs (Figures 3 and 4). Pseudouridine residues in oligonucleotides cannot be identified by LC/ESIMS because conversion of U to Ψ does not result in a mass change. Sequence location of Ψ was, therefore, performed using a specific chemical treatment of the overexpressed and native tmRNAs followed by reverse transcription (Bakin and Ofengand, 1993); see Figure 7A for details. Using five different DNA primers,

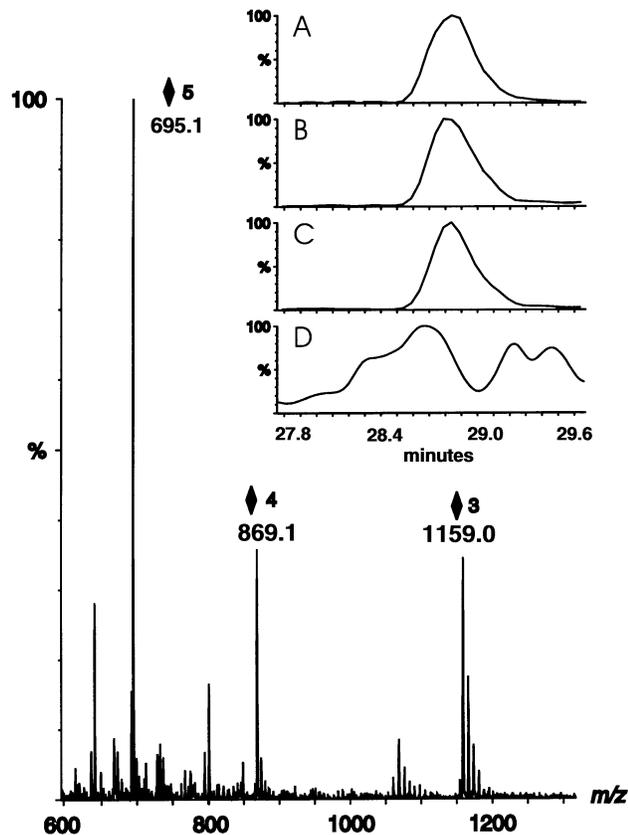


Fig. 6. ESI mass spectrum of ${}_{341}\text{UUCAACUCCCG}_{351}$ + methyl (M_r 3480.2) from an RNase T₁ digest of native *E. coli* tmRNA. The number adjacent to each character indicates the number of charges on the ion; for example, $\blacklozenge 5$ indicates the ion $(M - 5H)^{5+}$. Inset: A limited portion of the LC/ESIMS analysis showing the elution position of the methylated oligonucleotide indicated by the tracks of its characteristic ions (A) m/z 695.1, (B) m/z 869.1 and (C) m/z 1159.0. UV detection at 254 nm is shown in trace (D).

the entire sequence of the native and overexpressed tmRNAs was examined to locate Ψ residue(s).

As shown in Figure 7B, a strong stop was observed at positions one base 3' to U₃₄₂ in both the native and overexpressed *E. coli* tmRNAs treated with *N*-cyclohexyl-*N'*-(4-methylmorpholinium)-ethylcarbodiimide *p*-tosylate (CMC). A band is also visible at the site itself as a result of 'stuttering' (Denman *et al.*, 1988). These strong stops were not seen in the two natural tmRNAs following reverse transcription in the absence of CMC treatment, or in treated or untreated synthetic tmRNA (Figure 7B). Interestingly, a weak stop was also observed at position one base 3' to U₃₄₇ only for the native tmRNA (difficult to see in Figure 7B, and revealed only after quantitation of the bands). The quantitation of radioactivity in the bands corresponding to stops at Ψ_{342} , in both native and overexpressed tmRNAs, indicates that the overexpressed tmRNA is 12% less modified than the native tmRNA. When tmRNA is overexpressed, the modifying enzymes may be saturated. Compared with Ψ_{342} in the native tmRNA, position Ψ_{347} is 18-fold less modified. The entire tmRNA sequence was examined for the presence of other putative Ψ , but no other sites were found. A careful comparative quantitative analysis was performed because several scattered uridines in both native and synthetic

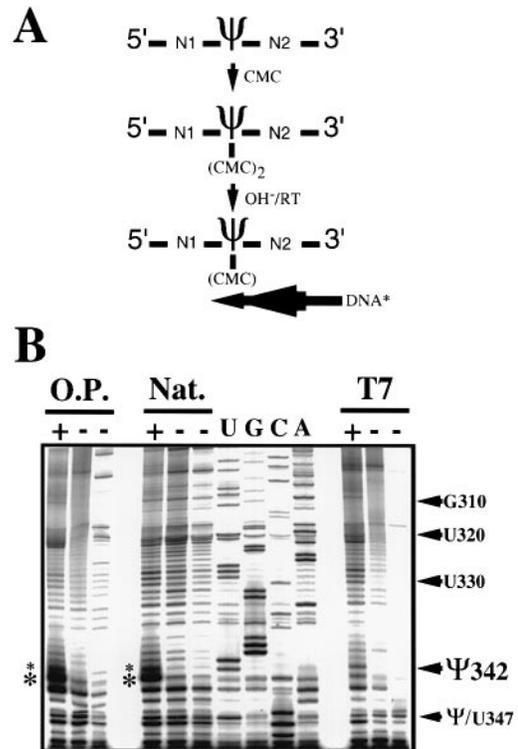


Fig. 7. (A) Schematic diagram illustrating the detection of Ψ residues in RNAs by a CMC-alkali treatment followed by reverse transcription (Ho and Gilham, 1971; Bakin and Ofengand, 1993), and its application to both native and overexpressed tmRNAs as compared with a T7 transcript (negative control). (B) Autoradiograph of 14% PAGE of cleavage products of 5'-labeled DNA primer. Reverse transcription using a primer corresponding to residues G₃₅₁-A₃₆₃. Samples were treated with (+) or without (-) CMC for 20 min at 37°C, exposed to pH 10.4 for 3 h at 37°C, and reverse transcribed. A, C, U and G are sequencing lanes using overexpressed tmRNA. Sequencing tracks are numbered every 10 residues. Notice that after chemical treatment, a major band appears one base downstream from Ψ sites [big arrow in (A) corresponding to the big star in (B)] as well as a minor one at the site itself [small arrow in (A) corresponding to the small star in (B)] as a result of 'stuttering' (Denman *et al.*, 1988).

RNAs were found reactive following CMC treatment. These results allow the following partial sequence to be assigned: ${}_{341}\text{U}\Psi\text{CAAC}\Psi\text{CCCG}_{351}$.

This same oligonucleotide, ${}_{341}\text{U}\Psi\text{CAAC}\Psi\text{CCCG}_{351}$, is the one found to contain a methyl group by LC/ESIMS analysis of RNase T₁ digests of *E. coli* tmRNA. Of a total of three uridines present in this oligonucleotide, two (U₃₄₂ and U₃₄₇) are modified to Ψ . Consequently, the m⁵U can be assigned to the only uridine remaining, U₃₄₁. Remarkably, all of the post-transcriptional modifications of the overexpressed and native *E. coli* tmRNAs have, therefore, been assigned to the following sequence: ${}_{341}\text{m}^5\text{U}\Psi\text{CAAC}\Psi\text{CCCG}_{351}$, and are located within a 7 nt segment, out of 363 nt.

Discussion

Utility of LC/ESIMS for the detection of modified bases in RNAs

The accurate measurement of molecular masses of RNase T₁ oligonucleotides from any RNA allows nucleotide compositions to be determined and compared with the

unmodified RNA or its gene (Kowalak *et al.*, 1993; Pomerantz *et al.*, 1993). To locate the m⁵U found in native and overexpressed *E. coli* tmRNAs from LC/ESIMS analyses of total nucleoside digests, we adapted an earlier protocol originally devised for mapping the modified nucleosides in rRNA (Kowalak *et al.*, 1993). In the original protocol, the RNase T₁ digest was fractionated by anion exchange chromatography prior to mass analysis, because the number of oligonucleotides generated was too large for direct analysis. In addition, LC/ESIMS analysis of oligonucleotide mixtures was not a sensitive nor routine procedure at that time. Encouraged by the recent description of a favorable solvent system for LC/ESIMS analysis of oligonucleotides (Apffel *et al.*, 1997), we eliminated the anion exchange fractionation and analyzed the RNase T₁ digests directly by LC/ESIMS without any purification. The extent of the sequence successfully analyzed was unexpected given both the complexity and the large number of oligonucleotides derived from RNase T₁ digestion of a 363 nt-long RNA.

The overall protocol outlined in Figure 2 can be, in principle, applied to the mapping of any structural modification of RNA producing a mass change, for RNAs ≤363 nt in length, depending on the extent of modification. Indeed, a T₁ digest of the RNA, still containing the enzyme and buffers, can be directly analyzed by LC/ESIMS without prior clean-up. Mass accuracies using a linear quadrupole mass analyzer, as in the present study, are routinely within 0.01–0.05%, sufficient to distinguish C from U (1 Da difference) in most cases. An estimated 10 pmol of RNase T₁ digest of native tmRNA was consumed in the LC/ESIMS of the oligonucleotides, with average absolute errors of 0.1–0.2 Da (Table I).

Matrix-assisted laser desorption–ionization time-of-flight (MALDI-TOF) mass spectrometry has recently been proposed for mass analysis of enzymatically derived oligonucleotides from RNA (Hahner *et al.*, 1997), and was illustrated for a synthetic 20-mer. In contrast with the present study, extensive sample clean-up was required. There is, likewise, no certainty that, as is suggested, the method will be generally applicable to modified RNAs. Moreover, MALDI-TOF is incompatible with chromatographic sample introduction, thus limiting its application to relatively simple mixtures. Finally, mass accuracy with the TOF analyzer is less than that obtained with linear quadrupole instruments, and limits its application to ‘quite small oligonucleotides’ (Hahner *et al.*, 1997).

A perfect structural analog of a T-loop in *E. coli* tmRNA?

Figure 8 shows the location of the post-transcriptionally modified nucleotides in a refined model of the *E. coli* tmRNA secondary structure (Hickerson *et al.*, 1998). Remarkably, all three modified nucleosides are located in a seven-nucleotide loop (U₃₄₁–U₃₄₇), within the proposed tRNA-like domain of tmRNA (Komine *et al.*, 1994). No other post-transcriptional modifications were found in either the mRNA-like domain or in the connecting segment as described (Felden *et al.*, 1996b). The present study suggests that the nucleotides m⁵U₃₄₁Ψ₃₄₂C₃₄₃A₃₄₄ in *E. coli* tmRNA mimic the universally conserved sequence m⁵U₅₄Ψ₅₅C₅₆R₅₇ (R purine) in T-loops of canonical tRNAs. The solution conformation of loop m⁵U₃₄₁–Ψ /

U₃₄₇ might, therefore, be very similar to that of a T-loop of canonical tRNAs, with a reverse-Hoogsteen base pair T₃₄₁–A₃₄₅ as suggested (Felden *et al.*, 1996a), and in agreement with our recent chemical and enzymatic probing data (Felden *et al.*, 1997).

In canonical tRNAs, Ψ₅₅ and C₅₆ interact via two tertiary interactions with G₁₈ and G₁₉ from the D-loop, respectively, to stabilize their overall three-dimensional architectures (Suddath *et al.*, 1974). Perhaps the conserved residues G₁₃ and G₁₄ (the numbering is as for *E. coli*, but varies according to species) in all known tmRNA sequences (Williams and Bartel, 1996; Felden *et al.*, 1997) are the structural analogs of G₁₈ and G₁₉ in tRNAs, but additional experiments including a mutational analysis of these residues would be helpful for conclusive answers. The structural mimicry of tmRNAs with canonical tRNAs may be extended to several nucleotides from H5, especially those in the stem–loop U₁₀–A₂₀, perhaps mimicking both the positioning and conformation of a D stem–loop (see Figure 8 for details). In both native and overexpressed tmRNAs, however, loop U₁₂–C₁₈ does not appear to contain any dihydrouridines.

Two enzymes from *E. coli*, tRNA (m⁵U₅₄)-methyltransferase (RUMT; Kealey and Santi, 1994) and tRNA psi 55 synthase (Nurse *et al.*, 1995), usually specific for the modification of canonical tRNAs, are good candidates for recognizing and modifying U₃₄₁ and U₃₄₂, respectively, in an RNA deviating considerably in structure and length as with *E. coli* tmRNA. For the RUMT, a seven-base T-loop and five-base-pair stem is a good substrate for modification of tRNAs (Gu and Santi, 1991), suggesting that the overall conformation of tmRNA does not perturb (e.g. by steric hindrance) the efficient recognition of the stem–loop G₃₃₆–C₃₅₂ by the enzyme.

Pseudouridine is not reported in any sequenced tRNA at position 60 (Sprinzl *et al.*, 1998), which corresponds to Ψ₃₄₇ in tmRNA. *E. coli* pseudouridine synthase I (Kammen *et al.*, 1988) is a potential candidate for partial (~5%; see above) modification of U₃₄₇ in only the native tmRNA. Indeed, this enzyme may recognize the T stem–loop of tmRNA as it normally does for the canonical tRNA anticodon stem–loop. Seven nucleotides in both T and anticodon loops, as well as weaker tertiary interactions in tmRNA as compared with canonical tRNAs, could be responsible for the recognition of this unusual RNA substrate by the enzyme.

In canonical tRNAs, there are only about 40 nt between the last nucleotide of the 5'-strand of the acceptor stem (nt 7) and the first nucleotide of the 5'-strand of the T-stem (nt 49 according to the numbering of tRNAs), whereas in *E. coli* tmRNA there are 329 nt. Consequently, tmRNA has built up its acceptor branch on different architectural rules than those of canonical tRNAs, but it can still be recognized by at least four enzymes specific to canonical tRNAs, including alanyl-tRNA synthetase and RNase P from *E. coli* (Komine *et al.*, 1994; Ushida *et al.*, 1994). This suggests that the conformations of tRNA and tmRNA acceptor branches (acceptor stem and T stem–loop) are very similar, if not identical. Alternatively, if the modifications are mediated by tmRNA-specific enzymes, then structural mimicry with tRNA may not be absolute.

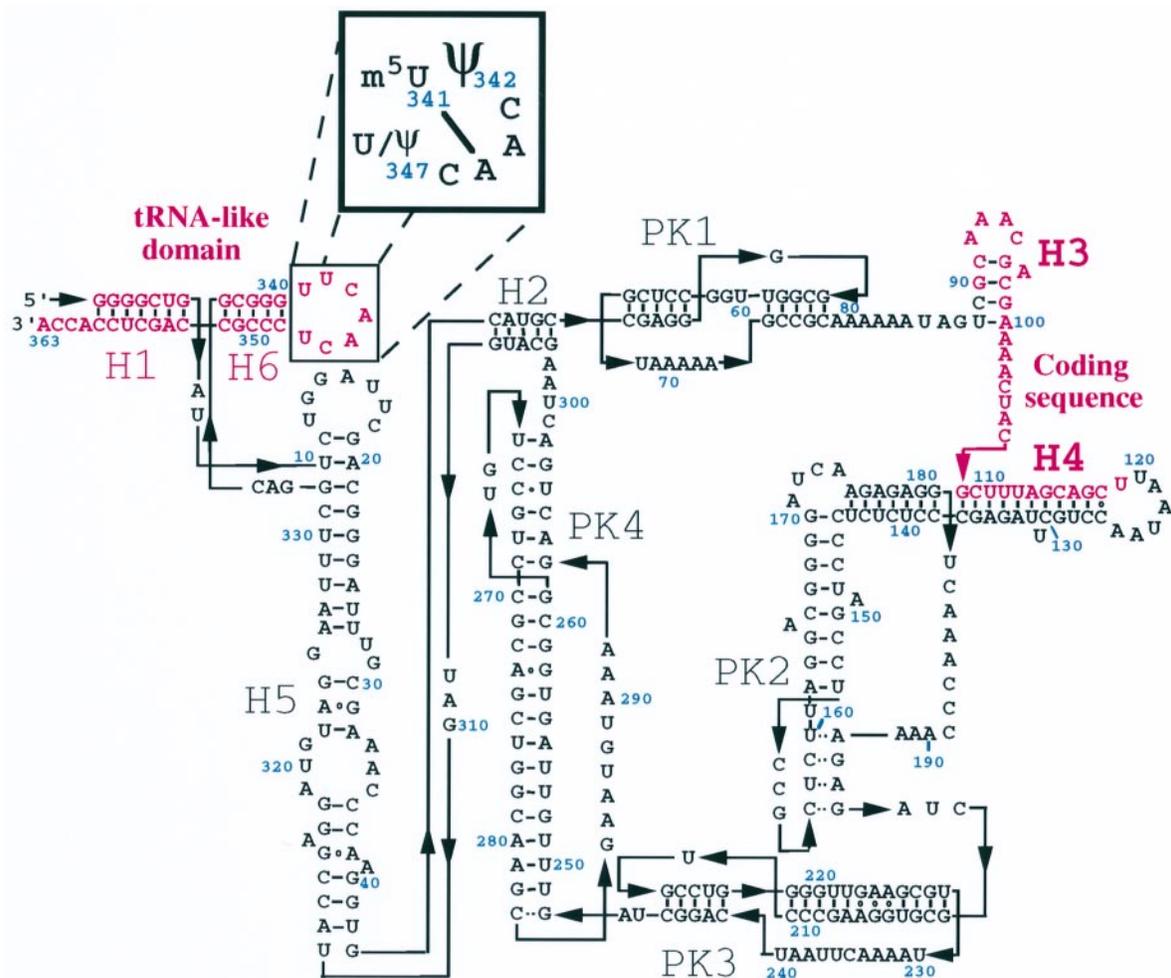


Fig. 8. Experimentally and phylogenetically supported secondary structure model for the *E. coli* tmRNA (derived from Felden *et al.*, 1996b, 1997; Williams and Bartel, 1996), showing the presence and location of the modified nucleotides determined in the present study. The two functional units of the molecule are shown in red.

Structural and functional implications of the presence of post-transcriptional modifications in *E. coli* tmRNA

Are these post-transcriptional modifications present for specific structural and/or functional purpose(s) or are they present inadvertently? In bacterial tRNAs, these two modifications are universally conserved. m^5U_{54} increases the stability of tRNAs (Davanloo *et al.*, 1979), decreases translation errors and increases ribosomal A-site binding, although the lack of m^5U_{54} in tRNA only marginally reduces (4%) the growth rate (reviewed in Björk, 1992). In tRNAs, Ψ_{55} participates in the specific geometry of T-loops, and probably plays a role in stabilizing the D-loop–T-loop tertiary interactions. Since Tamura *et al.* (1991) have reported that the kinetics parameters of alanylation of a T7 transcript corresponding to the sequence of the *E. coli* tRNA^{Ala} are very similar to that of a fully modified *E. coli* tRNA^{Ala}, the direct implication of the modified nucleotides in tmRNA aminoacylation is unlikely.

Comparative UV melting curves between synthetic and overexpressed *E. coli* tmRNAs indicate that there is no difference in t_m between the two molecules (not shown), suggesting that the post-transcriptional modifications have only a minor, if any, contribution to the overall stability of the RNA. To investigate further the putative structural

role of these modified nucleotides in *E. coli* tmRNA, the solution conformations of a synthetic versus an overexpressed tmRNA have been probed using dimethylsulfate and CMC (B.Felden *et al.*, unpublished data). The results obtained suggest that the modified bases stabilize the overall conformation of the acceptor branch at 37°C.

To address whether modified nucleotides have any role in the alanylation of *E. coli* tmRNA, comparative charging experiments between a synthetic and an overexpressed tmRNA were performed with *E. coli* alanyl-tRNA synthetase (highly purified enzyme). Interestingly, the synthetic tmRNA is reproducibly aminoacylated to a lower extent (2–15%, depending on the preparation) than the overexpressed tmRNA (10–30%, also depending on the preparation). This lower charging efficiency for a T7 tmRNA transcript as compared with a native tmRNA has already been reported (Komine *et al.*, 1994). The calculation of the amount of T7 tmRNA charged has been corrected, since it contains only about half (48%) of the correct CCA end (calculated from the LC/ESIMS analysis of a T₁ digest of synthetic tmRNA; not shown). However, comparative charging experiments (aminoacylation plateaus only) performed at different temperatures, 4, 25 and 37°C, have shown that the aminoacylation profiles are similar at 4°C for a T7 and an overexpressed tmRNA. At

both 25 and 37°C, however, the overexpressed tmRNA is charged more efficiently than the T7 tmRNA (not shown). This experiment suggests that the modified nucleotides are required to stabilize the conformation of the acceptor stem at 37°C which is needed for efficient alanylation. Indeed, the major determinants for *E. coli* tRNA^{Ala} alanine acceptance are a wobble G₃-U₇₀ pair within the acceptor stem and the discriminator base (Hou and Schimmel, 1988; McClain and Foss, 1988). A weak structural perturbation of the tmRNA conformation close to the G₃-U₇₀ pair, as evidenced by our structural probing of the synthetic tmRNA at 37°C (B. Felden *et al.*, unpublished data), may have unpredictable indirect consequences concerning the alanine acceptance of the molecule [see Ramos and Varani (1997) for the different hypotheses concerning the contribution of the G₃-U₇₀ pair in tRNA^{Ala} identity].

In tmRNA, the acceptor stem is made by the pairing of both the extreme 5' and 3'-ends. The post-transcriptional modifications may play an indirect role in alanylation in stabilizing the acceptor branch which is constructed with unusual architectural rules as compared with tRNAs. Additional experiments, including the progressive insertion of these post-transcriptional modifications into a synthetic tmRNA using purified enzymes, and following the alanylation properties of the molecule (as well as the verification by LC/ESIMS that the post-transcriptional modifications are added to the synthetic tmRNA), should be helpful in developing a better understanding of the structure-function relationships of this fascinating RNA.

Materials and methods

Nucleic acid preparation

tmRNAs were either extracted from cells (both native and overexpressed tmRNAs) or prepared by *in vitro* transcription. Native tmRNA was purified from *E. coli* cells (strain W3110). The nucleic acid fraction was phenol-extracted from mid-log phase cells followed by ethanol precipitation. After the second phenol extraction and ethanol precipitation, the resulting fraction was subjected to differential isopropanol precipitations to roughly remove the DNA, followed by incubation with FPLC grade DNase I (Pharmacia; Piscataway, NJ). Native tmRNA was further purified by 5% denaturing PAGE. Two closely spaced bands were always observed on the gels, and they could not be further separated by either electrophoresis or HPLC. These two RNAs correspond mainly to the native tmRNA successfully analyzed in the present study, and to the RNase P RNA (Jain *et al.*, 1982), based on the detection of characteristic RNase T₁ oligonucleotides from the latter RNA (data not shown).

The overexpressed tmRNA was purified as described (Felden *et al.*, 1997). This overproduction system yielded ≥100-fold amount of tmRNA, properly processed as shown by its ability to be aminoacylated *in vitro* with alanine. RT-PCR using vent polymerase (New England Biolabs; Beverly, MA) was performed directly on a small amount of this tmRNA to generate a template for construction of an unmodified T7 transcript of the *E. coli* tmRNA to serve as a negative control. Two DNA templates containing a T7 polymerase promoter for *in vitro* transcription (5'-TGGTGGAGCTGGCGGGAG-3' and 5'-GGGGCTGATCTGGATT-CGA-3') were synthesized on an Applied Biosystems model 394 synthesizer using the phosphoramidite method. These DNAs were used to amplify by RT-PCR the purified *E. coli* tmRNA to prepare a DNA suitable for T7 transcription. After transcription using a RiboMAX™ kit (Promega; Madison, WI) phenol extraction followed by a DNase I treatment was performed, and unincorporated NTPs and DNA fragments were removed using Microcon 50 microconcentrators (Amicon; Beverly, MA). Both size and sequence of the T7 transcript were verified and compared with the tmRNA purified from cells. Spectrophotometric measurements were made to determine the concentration of the RNA.

Enzymatic hydrolysis of RNAs

Native and overexpressed *E. coli* tmRNAs (0.3 nmol), along with a T7 transcript, were hydrolyzed to nucleosides using nuclease P1, phosphodi-

esterase I and alkaline phosphatase (Crain, 1990). The overexpressed and synthetic tmRNAs (50 pmol each) and the native tmRNA preparation were each also digested with 1000 U RNase T₁ (Ambion; Austin, TX) for 30 min at 37°C (Kowalak *et al.*, 1993). The amount of native tmRNA in the isolate that was analyzed by LC/MS of its RNase T₁ digest was estimated to be ~10 pmols based on the ratio of its mass spectral peaks to the corresponding ones from the other two tmRNAs.

Combined LC/ESIMS of nucleosides from *E. coli* tmRNAs

Hydrolysates (nuclease P1) of tmRNA were injected directly onto an LC-18S column (250×2.1 mm) with 20×2.1 mm Supelgard LC-18 precolumn cartridge (Supelco; Bellefonte, PA). The solvent system consisted of 5 mM ammonium acetate, pH 5.3 (buffer A) and acetonitrile + H₂O (40/60, v/v), developed from 100% A to 100% B using a multilinear gradient (Pomerantz and McCloskey, 1990). UV spectra were acquired over the range 190–320 nm.

The liquid chromatograph (Hewlett-Packard 1090) was interfaced directly to a Fisons Quattro II mass spectrometer equipped with an ESI source (Micromass; Beverly, MA), and controlled using the Micromass MassLynx v.2.22 data system. The chromatographic effluent (300 μl/min) was conducted into the ion source without prior splitting. The source was maintained at 180°C. Capillary and lens voltages were 3.20 and 0.15 kV, respectively. Positive ions were detected. One-sec scans (including 0.1 s delay) were acquired over a range of *m/z* 103–500 throughout the entire 50-min separation for the initial screening analyses. The cone voltage was ramped from 35–5 V over the mass range scanned.

For high sensitivity detection of Ψ and m⁵U, selected ion recording was performed for *m/z* 209 and 245 (Ψ) between 3.0 and 4.7 min and for *m/z* 127 and 259 (m⁵U) between 12.0 and 15.0 min. The dwell time was 0.08 s for all ions. The cone voltage was set to 40 V for generating the fragment ions (*m/z* 209 and 127) and to 25 V for detecting the MH⁺ ions (*m/z* 245 and 259).

Combined LC/ESIMS of oligonucleotides from RNase T₁ digests of *E. coli* tmRNAs

Hydrolysates of tmRNAs were injected directly onto a 150×1 mm PhaseSep ODS2 column (ChromTech; Apple Valley, MN) with 15×1 mm Optiguard C-18 precolumn cartridge (Optimize Technologies; Oregon City, OR). The solvent system consisted of 0.8 M 1,1,1,3,3,3-hexafluoro-2-propanol (J.T. Baker; Phillipsburg, NJ), adjusted to pH 7.0 with triethylamine, half of which was diluted 1:1 with HPLC grade water (buffer A) and the other half 1:1 with methanol (buffer B) (Apffel *et al.*, 1997). The column was eluted using a linear gradient of 0–100% B in A + B over 45 min, at a flow rate of 40 μl/min.

The chromatographic effluent (40 μl/min) was conducted into the ion source without prior splitting. The source was maintained at 145°C. Capillary and lens voltages were –3.20 and 0.30 kV, respectively; the cone voltage was 52 V. Two-second scans (including 0.1 s delay) were acquired over a scan range of *m/z* 500–1600. Negative ions were detected.

Determination of oligonucleotide composition from molecular mass

Electrospray mass spectra were converted to *M_r* values using utility subroutines in the Micromass MassLynx v.2.22 software suite. Calculation of expected RNase T₁ fragments from the *E. coli* tmRNA sequence, and of multiply charged ions predicted from ESI of these fragments, were accomplished using software written in-house. A look-up table consisting of all RNA compositions containing up to two Gs and six methyl groups, with *M_r* values ≤3200, sorted by increasing mass has also been calculated.

CMC modification

The procedure was performed essentially as described (Bakin and Ofengand, 1993). For sequence localization of Ψ residues in tmRNA, five synthetic DNA oligonucleotides (5'-TGGTGGAGCTGGC-3', 5'-TTACATTCCGTTGCCAGC-3', 5'-CAGGCAGGCTTCCACGC-3', 5'-GAGAGAGGGCTCTAAGCA-3', 5'-TTTTTACGAGGCCAACCG-3'), complementary to residues G₃₅₁-A₃₆₃, G₂₇₄-A₂₉₁, G₂₀₀-G₂₁₇, U₁₂₈-C₁₄₅ and C₅₆-A₇₃, respectively, were synthesized and 5'-labeled using [γ -³²P]ATP at 3200 Ci/mol (Dupont NEN; Wilmington, DE) and phage T4 polynucleotide kinase (New England Biolabs; Beverly, MA) prior to initiation of reverse transcription. The whole tmRNA sequence was screened with these five primers by reverse transcription using AMV reverse transcriptase (Promega; Madison, WI).

Quantitation

The relative amount of U residues converted to Ψ was analyzed by phosphorimager using 'ImageQuant' (Molecular Dynamics; Sunnyvale,

CA). For the native, overexpressed and T7 tmRNAs as well as for the control lanes, the amount of radioactivity in the bands one base downstream from a putative Ψ site (A) was normalized to the total amount of radioactivity in the lane (B). This value ($A' = A/B$) was then subtracted from the value obtained with the corresponding controls ($A'' = A' - A'_C$, where A'_C is the corresponding control). A'' was finally subtracted from the corresponding value obtained with the T7 tmRNA, to give the relative amount of Ψ modification at this site ($A''' = A'' - A'''_{T7}$).

Acknowledgements

Software routines used to generate mass lists and to aid interpretation of electrospray ionization mass spectra of oligonucleotides from RNase T₁ digests were written by S.C.Pomerantz. The plasmid encoding the *his*-tagged alaRS was a generous gift from Professor P.Schimmel, Massachusetts Institute of Technology. R.F.G. is an Investigator of the Howard Hughes Medical Institute. This work was also supported by NIH grants GM29812 (J.A.M.) and GM48152 (J.F.A.).

References

- Apffel,A., Chakel,J.A., Fischer,S., Lichtenwalter,K. and Hancock,W.S. (1997) Analysis of oligonucleotides by HPLC-electrospray ionization mass spectrometry. *Anal. Chem.*, **69**, 1320–1325.
- Bakin,A. and Ofengand,J. (1993) Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center: Analysis by the application of a new sequencing technique. *Biochemistry*, **32**, 9754–9762.
- Björk,G.R. (1992) The role of modified nucleosides in tRNA interactions. In Hatfield,D.L., Lee,B.J. and Pirtle,R.M. (eds), *Transfer RNA in Protein Synthesis*. CRC Press, Boca Raton, FL, pp. 23–85.
- Crain,P.F. (1990) Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry. *Methods Enzymol.*, **193**, 782–790.
- Davanloo,P., Sprinzl,M., Watanabe,K., Albani,M. and Kersten,H. (1979) Role of ribothymidine in the thermal stability of transfer RNA as monitored by proton magnetic resonance. *Nucleic Acids Res.*, **6**, 1571–1581.
- Denman,R., Colgan,J., Nurse,K. and Ofengand,J. (1988) Crosslinking of the anticodon of the P site bound tRNA to C-1400 of *E.coli* 16S RNA does not require the participation of the 50S subunit. *Nucleic Acids Res.*, **16**, 165–178.
- Felden,B., Atkins,J.F. and Gesteland,R.F. (1996a) tRNA and mRNA both in the same molecule. *Nature Struct. Biol.*, **3**, 494.
- Felden,B., Himeno,H., Muto,A., Atkins,J.F. and Gesteland,R.F. (1996b) Structural organization of *Escherichia coli* tmRNA. *Biochimie*, **78**, 979–983.
- 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**, 89–104.
- Fenn,J.B., Mann,M., Meng,C.K., Wong,S.F. and Whitehouse,C.M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science*, **246**, 64–71.
- Gu,X.R. and Santi,D.V. (1991) The T-arm of tRNA is a substrate for tRNA (m⁵U54)-methyltransferase. *Biochemistry*, **30**, 2999–3002.
- Hahner,S., Lüdemann,H.-C., Kirpekar,F., Nordhoff,E., Roepstorff,P., Galla,H.-J. and Hillenkamp,F. (1997) Matrix-assisted laser desorption/ionization mass spectrometry (MALDI) of endonuclease digests of RNA. *Nucleic Acids Res.*, **25**, 1957–1964.
- Hickerson,R.P., Watkins-Sims,C.D., Burrows,C.J., Atkins,J.F., Gesteland,R.F. and Felden,B. (1998) A nickel complex cleaves uridines in folded RNA structures: application to *E.coli* tmRNA and related engineered molecules. *J. Mol. Biol.*, in press.
- 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.
- Ho,N.W.Y. and Gilham,P.T. (1971) Reaction of pseudouridine and inosine with *N*-cyclohexyl-*N*-beta-(4-methylmorpholinium)ethylcarbodiimide. *Biochemistry*, **10**, 3651–3657.
- Hou,Y.-M. and Schimmel,P. (1988) A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature*, **333**, 140–145.
- Jain,S.K., Gurevitz,M. and Apirion,D. (1982) A small RNA that complements mutants in the RNA processing enzyme ribonuclease P. *J. Mol. Biol.*, **162**, 515–533.
- Kammen,H.O., Marvel,C.C., Hardy,L. and Penhoet,E.E. (1988)

- Purification, structure, and properties of *Escherichia coli* tRNA pseudouridine synthase I. *J. Biol. Chem.*, **263**, 2255–2263.
- Kealey,J.T. and Santi,D.V. (1994) High level expression and rapid purification of tRNA (m⁵U54)-methyltransferase. *Protein Expr. Purif.*, **5**, 149–152.
- Keiler,K.C., Waller,P.R.H. and Sauer,R.T. (1996) Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science*, **271**, 990–993.
- 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.
- Kowalak,J.A., Pomerantz,S.C., Crain,P.F. and McCloskey,J.A. (1993) A novel method for the determination of post-transcriptional modification in RNA by mass spectrometry. *Nucleic Acids Res.*, **21**, 4577–4585.
- Lee,S.Y., Bailey,S.C. and Apirion,D. (1978) Small stable RNAs from *Escherichia coli*: evidence for the existence of new molecules and for a new ribonucleoprotein particle containing 6S RNA. *J. Bacteriol.*, **133**, 1015–1023.
- McClain,W.H. and Foss,K. (1988) Changing the specificity of a tRNA by introducing a G-U wobble pair near the 3' acceptor end. *Science*, **240**, 793–796.
- McCloskey, J.A. (1974) Mass spectrometry. In Ts'o,P.O.P.(ed.), *Basic Principles in Nucleic Acid Chemistry*, Vol. I. Academic Press, New York, pp. 209–309.
- McCloskey,J.A. and Crain,P.F. (1998) The RNA modification database–1998. *Nucleic Acids Res.*, **26**, 196–197.
- Nurse,K., Wrzesinski,J., Bakin,A., Lane,B.G. and Ofengand,J. (1995) Purification, cloning, and properties of the tRNA psi 55 synthase from *Escherichia coli*. *RNA*, **1**, 102–112.
- Pomerantz,S.C. and McCloskey,J.A. (1990) Analysis of RNA hydrolyzates by LC/MS. *Methods Enzymol.*, **193**, 796–824.
- Pomerantz,S.C., Kowalak,J.A. and McCloskey,J.A. (1993) Determination of oligonucleotide composition from mass spectrometrically measured molecular weights. *J. Am. Soc. Mass Spectrom.*, **4**, 204–2090.
- Ramos,A. and Varani,G. (1997) Structure of the acceptor stem of *Escherichia coli* tRNA^{Ala}: role of the G3-U70 base pair in synthetase recognition. *Nucleic Acids Res.*, **25**, 2083–2090.
- Ray,B.K. and Apirion,D. (1979) Characterization of 10S RNA: a new stable RNA molecule from *E. coli*. *Mol. Gen. Genet.*, **174**, 25–32.
- Sprinzl,M., Horn,C., Brown,M., Ioudovitch,A. and Steinberg,S. (1998) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.*, **26**, 148–153.
- Suddath,F.L., Quigley,G.J., McPherson,A., Sneden,D., Kim,J.J., Kim,S.H. and Rich,A. (1974) Three-dimensional structure of yeast phenylalanine transfer RNA at 3.0 angstroms resolution. *Nature*, **248**, 20–24.
- Tamura,K., Asahara,H., Himeno,H., Hasegawa,T. and Shimizu,M. (1991) Identity elements of *Escherichia coli* tRNA^{Ala}. *J. Mol. Recogn.*, **4**, 129–132.
- 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 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.
- Williams,K.P. and Bartel,D.P. (1996) Phylogenetic analysis of tmRNA secondary structures. *RNA*, **2**, 1306–1310.
- Williams,K.P. and Bartel,D.P. (1998) The tmRNA website. *Nucleic Acids Res.*, **26**, 163–165.
- Zweib,C., Larsen,N. and Wower,J. (1998) The tmRNA database (tmRNA). *Nucleic Acids Res.*, **26**, 166–167.

Received January 15, 1998; revised March 16, 1998;
accepted March 26, 1998