

SURVEY AND SUMMARY

Comparative sequence analysis of tmRNA

Christian Zwieb*, Iwona Wower¹ and Jacek Wower¹

Department of Molecular Biology, The University of Texas Health Science Center at Tyler, 11937 US Highway 271, Tyler, TX 75708-3154, USA and ¹Department of Animal and Dairy Sciences, Cell and Molecular Biosciences Program, Auburn University, Auburn, AL 36849, USA

Received February 9, 1999; Revised and Accepted March 22, 1999

ABSTRACT

Minimal secondary structures of the bacterial and plastid tmRNAs were derived by comparative analyses of 50 aligned tmRNA sequences. The structures include 12 helices and four pseudoknots and are refinements of earlier versions, but include only those base pairs for which there is comparative evidence. Described are the conserved and variable features of the tmRNAs from a wide phylogenetic spectrum, the structural properties specific to the bacterial subgroups and preliminary 3-dimensional models from the pseudo-knotted regions.

INTRODUCTION

tmRNA, previously named 10S RNA (1,2), is a small stable RNA with properties of tRNA and mRNA combined in a single molecule. tmRNA was first identified in *Escherichia coli* and has been found in numerous bacteria, including the genomes of plastids. There is significant evidence for the involvement of tmRNA in translation where tmRNA is believed to rescue arrested ribosomes that have reached the 3'-end of an mRNA lacking a stop codon. In this situation, the ribosome appears to bind alanine-charged tmRNA and resumes translation by decoding a region near the center of the tmRNA molecule. The C-terminus of the mRNA-encoded polypeptide is tagged with a peptide also specified by tmRNA. Presumably, this tagging mechanism signals the proteolytic destruction of an abnormal potentially harmful protein (3–7). Furthermore, the tmRNA provides the missing stop codon for the ribosome to terminate translation. In addition to its participation in protein tagging, tmRNA may play a role in modulating the activity of DNA-binding proteins (8).

Particularly fascinating aspects of tmRNA research are the overall degree of tmRNA sequence conservation, the level of its relationship to tRNA and the potential interactions of tmRNA with other cellular components. Understanding the functional role of tmRNA in these elaborate, yet unknown, processes requires a detailed knowledge of its structure. In numerous other large and small RNAs, the method of comparative sequence analysis has proven to be useful to determine structure and thus gain insights into the potential functions of the respective molecules (9–11). Here, we apply this approach to the secondary

structure of tmRNA, for which many sequences from a wide phylogenetic spectrum of bacteria have become available recently (12). By using 50 tmRNA sequences, we refine the secondary structure models derived previously (13), include only those base pairs for which support from comparative analysis is available and thus present reliable minimal tmRNA secondary structure models. For *E.coli* tmRNA, despite some differences in the tRNA-like part, the agreement between the phylogenetically derived secondary structure and the structures derived by enzymatic and chemical probes (14,15) is relatively good.

tmRNA SEQUENCES

Table 1 shows the full species names of the 50 tmRNA sequences that were used in our analysis (catalogued on January 5, 1999). Abbreviated species names refer to the compilation of the sequences that are available from the tmRDB (12) at the internet address <http://psyche.uthct.edu/dbs/tmRDB/tmRDB.html>. If possible, referral to the primary sources of the sequences is provided at the tmRDB. Additional information about tmRNA function and the tmRNA sequences is available at the tmRNA web site (16). Table 1 also shows which methods were used in the determination of sequences and indicates that most tmRNAs (34 sequences) were extracted from completely sequenced genomes. Seven sequences were obtained by amplification with primers directed to the conserved tRNA-like portions of tmRNA and thus lack precision in regions near the termini of the tmRNA. The tmRNA sequence of *Clostridium acetobutylicum* is incomplete as it is lacking the 5'-portion of the tRNA-like region and the peptide tag-encoding part, but provides information about the 3'-portion. Of all known tmRNAs, only those from *Bacillus subtilis*, *Mycoplasma capricolum* and *E.coli* have been isolated and tested for their ability to participate in aminoacylation reactions and ribosome binding (17). Moreover, only the structure of the *E.coli* tmRNA has been studied in depth using chemical and enzymatic approaches (15,18).

All known tmRNAs are of bacterial origin and no tmRNA genes have been identified in the archaeobacteria. Within the bacterial domain, the tmRNA sequences belong to a wide variety of phylogenetic subgroups, with some species being deeply rooted (12). Genes for tmRNA have also been identified in the cyanelle genome of *Cyanophora paradoxa* and the chloroplast genomes of *Odontella sinensis*, *Porphyra purpurea*, *Guillardia theta* and *Thalassiosira weissflogii*. We were unable to find the gene for tmRNA in the recently completed genomic sequence of

*To whom correspondence should be addressed. Tel: +1 903 877 7689; Fax: +1 903 877 5731; Email: zwieb@uthct.edu

Table 1. Catalog of the tmRNA sequences used in this survey

Thermophilic Oxygen Reducers			Purple Bacteria (cont.)		
<i>Aquifex aeolicus</i>	AQU.AEO.	(G) (38)	<i>Pseudomonas aeruginosa</i>	PSE.AER.	(G) (55)
Thermotogales			<i>Marinobacter hydrocarbonoclasticus</i>	MAR.HYD.	(A) (13)
<i>Thermotoga maritima</i>	THE.MAR.	(G) (39)	<i>Pseudoalteromonas haloplanktis</i>	PSE.HAL.	(A) (13)
Green Non-sulfur Bacteria & Relatives			<i>Shewanella putrefaciens</i>	SHE.PUT.	(G) (39)
<i>Thermus thermophilus</i>	THE.THE.	(D) (40)	<i>Vibrio cholerae</i>	VIB.CHO.	(G) (39)
<i>Deinococcus radiodurans</i>	DEI.RAD.	(G) (39)	<i>Aeromonas salmonicida</i>	AER.SAL.	(A) (13)
Flexibacter Cytophaga Bacteroides			<i>Escherichia coli</i>	ESC.COL.	(R) (29)
<i>Porphyromonas gingivalis</i>	POR.GIN.	(G) (39)	<i>Salmonella typhimurium</i>	SAL.TYP.	(G) (56)
Green Sulfur Bacteria			<i>Yersinia pestis</i>	YER.PES.	(G) (52)
<i>Chlorobium tepidum</i>	CHL.TEP.	(G) (39)	<i>Actinobacillus actinomycetemcomitans</i>	ACT.ACT.	(G) (57)
Planctomyces & Relatives			<i>Haemophilus influenzae</i>	HAE.INF.	(G) (58)
<i>Chlamydia trachomatis</i>	CHL.TRA.	(G) (41)	<i>Desulfovibrio desulfuricans</i>	DES.DES.	(A) (13)
Cyanobacteria & Organelles			<i>Helicobacter pylori</i>	HEL.PYL.	(G) (59)
<i>Anabaena species</i>	ANA.SPE.	(cD) (42)	<i>Campylobacter jejuni</i>	CAM.JEJ.	(G) (52)
<i>Synechococcus species</i>	SYN.SP-A	(G) (43)	Gram Positive		
<i>Synechocystis species</i>	SYN.SP-B	(cD) (44)	<i>Mycobacterium leprae</i>	MYC.LEP.	(G) (52)
<i>Odontella sinensis</i> chloroplast	ODO.SIN.CL	(G) (45)	<i>Mycobacterium tuberculosis</i>	MYC.TUB.	(D) (60)
<i>Porphyra purpurea</i> chloroplast	POR.PUR.CL	(G) (46)	<i>Bacillus megaterium</i>	BAC.MEG.	(A) (61)
<i>Guillardia theta</i> chloroplast	GUI.THE.CL	(G) (47)	<i>Bacillus subtilis</i>	BAC.SUB.	(D) (17)
<i>Thalassiosira weissflogii</i> chloroplast	THA.WEI.CL	(cD) (48)	<i>Staphylococcus aureus</i>	STA.AUR.	(X) (39)
<i>Cyanophora paradoxa</i> cyanelle	CYN.PAR.CY	(G) (49)	<i>Enterococcus faecalis</i>	ENT.FAE.	(G) (39)
Spirochetes & Relatives			<i>Streptococcus pyogenes</i>	STR.PYO.	(G) (57)
<i>Borrelia burgdorferi</i>	BOR.BUR.	(G) (50)	<i>Streptococcus gordonii</i>	STR.GOR.	(G) (39)
<i>Treponema pallidum</i>	TRE.PAL.	(G) (39)	<i>Streptococcus mutans</i>	STR.MUT.	(G) (57)
Purple Bacteria			<i>Streptococcus pneumoniae</i>	STR.PNE.	(G) (39)
<i>Neisseria gonorrhoeae</i>	NEI.GON.	(G) (51)	<i>Mycoplasma capricolum</i>	MYC.CAP.	(D) (17)
<i>Neisseria meningitidis</i>	NEI.MEN.	(G) (52)	<i>Mycoplasma genitalium</i>	MYC.GEN.	(G) (62)
<i>Alcaligenes eutrophus</i>	ALC.EUT.	(D) (53)	<i>Mycoplasma pneumoniae</i>	MYC.PNE.	(G) (63)
<i>Dichelobacter nodosus</i>	DIC.NOD.	(G) (54)	<i>Clostridium acetobutylicum</i> (partial)	CLO.ACE.	(G) (64)
<i>Legionella pneumophila</i>	LEG.PNE.	(A) (13)			

Species are grouped as described by the Ribosomal Database Project (32) with the full species name given on the left column of each panel followed by the abbreviated code of the tmRDB (12). Sources of the isolates appear in parentheses in the third column of each panel and indicate if the sequence was obtained from DNA (D), from genomic sequencing (G), by PCR amplification (A), cDNA cloning (cD) or from RNA (R). References containing the sequence information appear in the last column of each panel.

the mitochondria-related α -proteobacterium *Rickettsia prowazekii*, which challenges the previous conjecture that tmRNAs are present in all bacteria.

SEQUENCE ALIGNMENT

Details of the alignment procedure have been described previously (11). Briefly, sequences from close relatives were aligned first by using primary structure similarities. Groups of aligned sequences were then aligned with each other to identify the conserved residues. Conserved positions were used as signposts to align the sequences of the more distant relatives. Unique to tmRNA, the tmRNA-encoded open reading frame assisted in aligning this portion of the molecule, including the conserved 'resume' and stop codons. Finally, common secondary structural elements were used as additional markers.

To determine the secondary structure and prove or disprove the existence of a particular base pair, we observed covariances between the nucleotide residues at two corresponding alignment positions. As a guideline, we considered compensating base changes (CBCs) first. (A CBC is a covariance of the Watson-Crick type, including G-U pairings that are commonly found in RNA.) For the conserved alignment positions, which by their nature lack

any covariation, no argument for the existence of a base pair could be made. Thus, a conserved base pair was included only when there was substantial support for its existence from biochemical experiments or rigid structural analysis or when the conserved pair was adjacent to a well-supported interaction.

The recent alignment of the tmRNA sequences is available from psyche.uthct.edu/dbs/tmRDB/ in a variety of formats. The sequences are grouped phylogenetically with the abbreviated species names corresponding to those listed in Table 1. In the tmRNA alignment (available from the tmRDB web site) helices supported by CBCs are shown highlighted by upper case letters and are numbered from 1 to 12 (starting from the 5'-end) in lines labeled 'hel-10', 'hel-1' and 'hel-x'. Helical sections are named with lower case letter extensions as described below. Near the top of the aligned sequences, the locations of pseudoknots 1-4 and the 'resume' and stop codons, respectively, are indicated. Furthermore, the degree of covariational support for Watson-Crick and G-U base pairs is shown on a scale from 1 to 9, with 1 indicating low support and 9 indicating the highest level of support as determined with the program Covariation (19). Finally, highly conserved positions with low covariation scores, but which nevertheless are included as being paired, are labeled with stars and positions which

correspond to invariant nucleotide residues are indicated. The alignment was created using the sequence editor provided by the program Sequp (20).

SECONDARY STRUCTURES

Secondary structure diagrams derived from comparative sequence analyses are presented in Figure 1 for *E.coli* (Fig. 1A), *Thermotoga maritima* (Fig. 1B), *Neisseria gonorrhoeae* (Fig. 1C), *Synechocystis* spp. (Fig. 1D), *Mycoplasma pneumoniae* (Fig. 1E) and the cyanelle of *C.paradoxa* (Fig. 1F). Supported Watson–Crick base pairs are juxtaposed as indicated by a line, while G–U pairs are indicated by circles. Below, we review the 12 tmRNA helices, numbered from 1 to 12 starting at the 5′-end. We consider the characteristics of each helix, including the connecting bulges and loops, as well as properties peculiar to the bacterial subgroups.

Helix 1

Helix 1 corresponds to the amino acid acceptor arm of tRNA. The first four and the last of the 7 bp are conserved or invariant (Fig. 2). Although the overall phylogenetic support for helix 1 is very weak, there is substantial biochemical evidence of its existence, including the ability to charge the *B.subtilis*, *E.coli* and *M.capricolum* tmRNAs with alanine (17). The uridine at position 1 of *Alcaligenes eutrophus* should be a guanine in order to form a Watson–Crick pair with the cytosine near the 3′-terminus. Similarly, the gap in the 3′-half of helix 1 in *Anabaena* spp. is likely to be filled with a guanine residue to form a G–C pair. Both inconsistencies may be due to sequencing errors.

There are 10–13 nt in the loop between helices 1 and 2. Two invariant G residues are present near the center of this loop and are also conserved in the dihydrouridine (DHU) loop of tRNA (21). The 3′-part of the loop ends with an invariant GA dinucleotide which is preceded by a pyrimidine, usually a cytidine. Overall, the tmRNA loop differs substantially from the DHU stem–loop region of tRNA, as its size is reduced and the formation of an equivalent stem is unsupported by CBCs.

Helix 2

Helix 2 is composed of four helical sections, 2a–2d. Typically, section 2a consists of variable residues that form eight base pairs; however, the first pair is an invariant C–G. Because this pair is located adjacent to a well-supported interaction with a high covariation score, it is included as base paired. Sections 2b and 2c form well-supported helices that, in most tmRNAs, are likely to be continuously stacked onto each other (Fig. 1B and E). However, non-conventional A–G and A–C ‘pairings’ might occur between the helical sections (Fig. 1A, D and F). The number of residues (two to nine) and the sequences of the internal loops between sections 2a and 2b are variable. There is the potential for base pairing, but no conclusive proof or disproof by CBCs is provided.

In contrast to 2a and 2b, section 2d contains numerous conserved residues (Fig. 2). The first base pair is always a C–G pair, with one exception in the *P.purpurea* chloroplast, where it is a U–G. Up to four residues may be present in the 5′-portion of the loop between helical sections 2c and 2d, but this loop is absent in many purple bacteria (Fig. 1A). The 3′-portion of the loop between sections 2c and 2d consists of three to seven residues of which U308, located adjacent to the invariant paired G307 of

helix 2d (*E.coli* tmRNA numbering; Fig. 1A), is the most conserved. No single-stranded residues connect section 2d to helix 3, indicating that section 2d and helix 3 might be stacked continuously.

Helix 3

Helix 3 typically consists of five base pairs, lacks invariant residues, but displays a similar degree of conservation to that of section 2d. In most secondary structures, helix 3 is an integral part of pseudoknot 1. However, this pseudoknot is not formed in two *Mycoplasma* species in which helix 4 is absent (Fig. 1E). *Mycoplasma pneumoniae* exhibits only 2 nt in the terminal loop and there is strong support for the terminal base pair of helix 3. Whether this is possible sterically will have to be verified experimentally; however, a similar two base loop was encountered in the 23S-like mitochondrial rRNAs of higher eukaryotes (22).

Helix 4

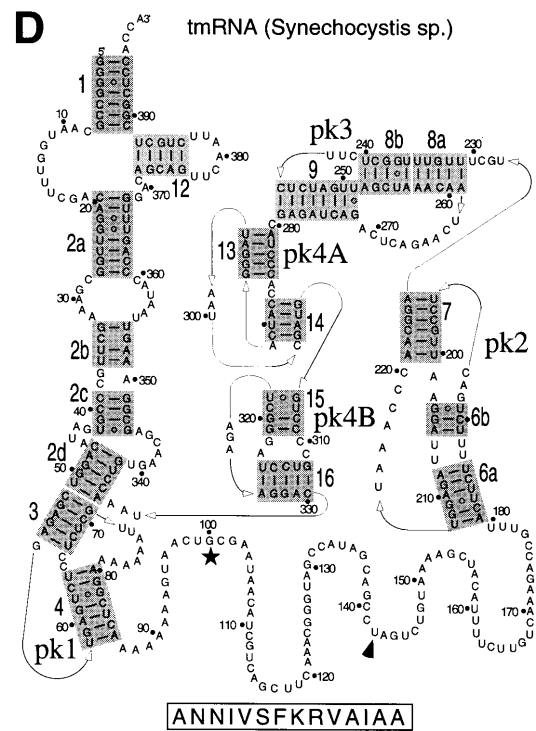
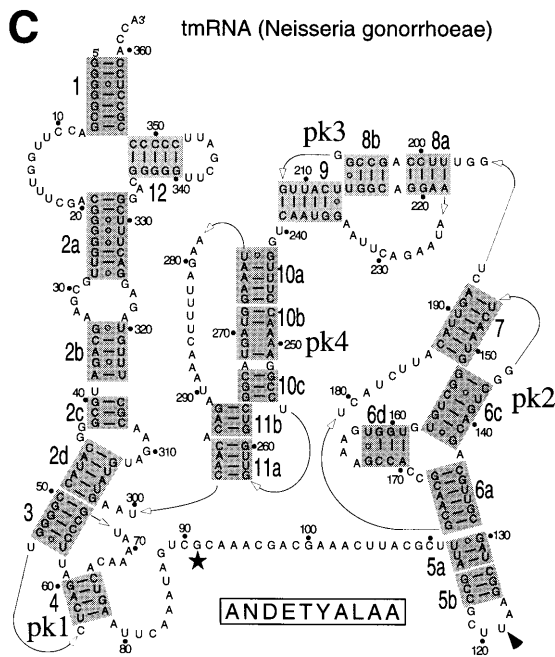
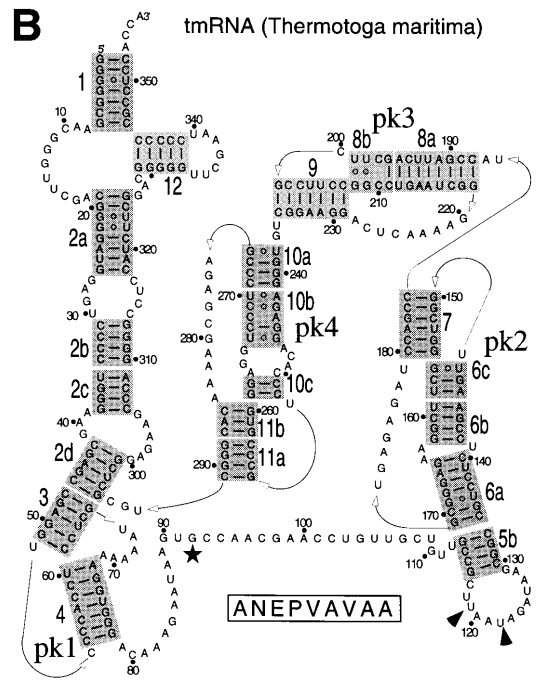
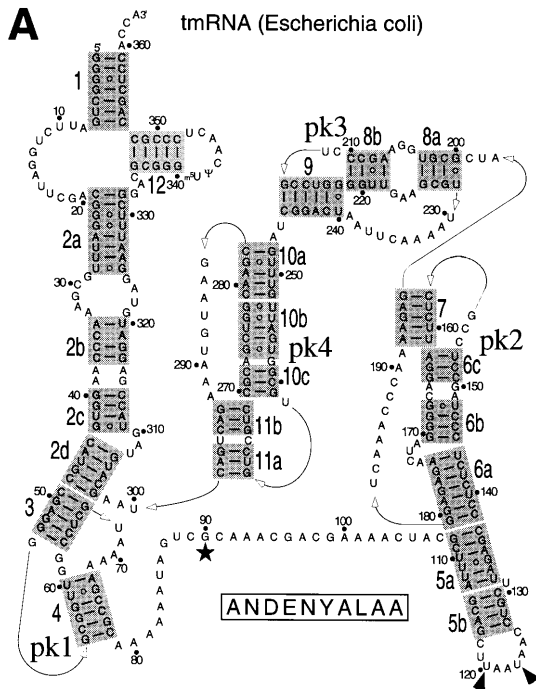
Helix 4 is variable in sequence and size (3–7 bp). Furthermore, it is absent in two *Mycoplasma* species, as discussed in the previous paragraph. Usually, the 5′-halves of helices 3 and 4 are joined by an immediate transition of no more than three residues, although one example of a six residue connection is found in the green sulfur bacterium *Chlorobium tepidum*. A small number of loop residues connect the 5′-half of helix 4 and the 3′-half of helix 3.

The loop between the 3′-halves of helices 3 and 4 contains four to 13 residues, some of which are highly conserved (Fig. 2). This loop is rich in adenosine residues, causing the alignment in this region to be ambiguous. The adenine at the third position of the loop appears to be invariant (A70 in *E.coli*; Fig. 1A); however, one exception is found in the *P.purpurea* chloroplast. The first residue of the loop (U68; Fig. 1A) is almost always a uridine, but there are a few exceptions, as in *Porphyromonas gingivalis*, *Anabaena* spp., *Synechocystis* spp. and the cyanelle of *C.paradoxa*.

Tag peptide coding region

Criteria for assigning a tag peptide coding region within the tmRNA sequences were: first, the requirement for an open reading frame between the ‘resume’ and the stop codon (no frameshifting was allowed); second, a guanine as the first residue of the ‘resume’ codon less than 25 nt from the end of helix 4; third, a tendency for the stop codon to be located within the terminal loop of helix 5. We emphasize that the assignment of these criteria is arbitrary and, with the exception of the *E.coli* sequence, the predicted tag peptide sequences are hypothetical and await experimental verification. Furthermore, although there is proof that the tag peptide serves as a relatively promiscuous signal for proteolysis in *E.coli* (23) and *B.subtilis* (24), knowledge about similar proteolytic systems in other bacteria is limited.

The base triplet preceding the ‘resume’ codon is relatively conserved (84–AUA–86 in *E.coli*; Fig. 1A). Higher degrees of conservation are also found in the first and second codon of the tag sequence. We do not know to what degree these constraints are related to requirements of the tag peptide. Although the RNA of the tag peptide coding region can be folded using thermodynamic rules (25), because of its variable length (30–87 residues) we were unable to align this region and identify base pairs supported sufficiently by CBCs.



Helix 5

Helix 5 is composed of two sections with relatively weak support. Some tmRNAs lack this helix (Fig. 1D and F). If helix 5 is present, the stop codon is located within the terminal loop and

preceded by the sequence GCNU (where G is invariant, the C is conserved but may be a uridine, N is any nucleotide and U is the first residue of the stop codon). In many sequences the invariant G (G117 in *E. coli*; Fig. 1A) is involved in forming the closing pair of helix 5.

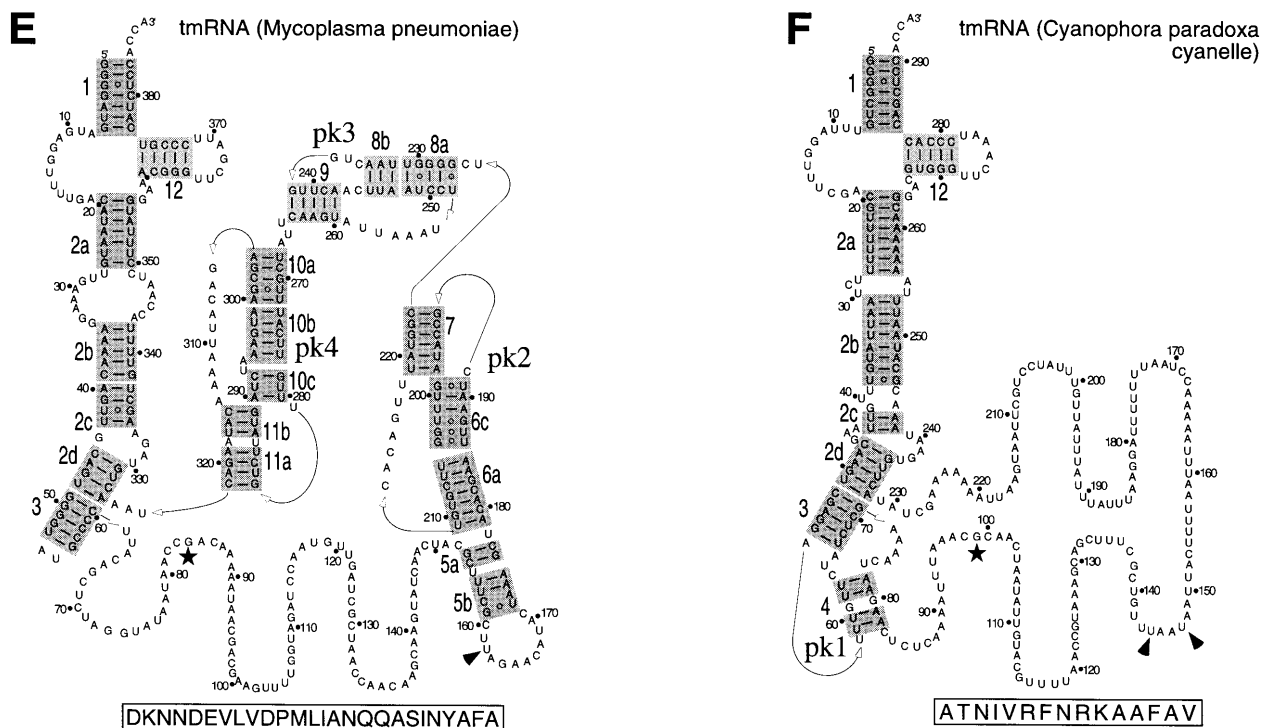


Figure 1. (Opposite and above) tmRNA secondary structure diagrams of (A) *E. coli* (purple bacterium), (B) *Thermotoga maritima* (Thermotogales), (C) *N. gonorrhoeae* (purple bacterium), (D) *Synechococcus* spp. (Cyanobacterium), (E) *M. pneumoniae* (Gram-positive) and (F) the cyanelle of *C. paradoxa*. Perfect base pairs are connected with a line and G-U pairs with open circles. Helices are highlighted in gray and numbered from 1 to 12 from the 5'-end. Helical sections are given extensions with lower case letters. The sequences are labeled with a dot and numbered (if allowed by the available space) in increments of 10. The 5'→3' direction of the RNA chain is indicated by lines with open arrowheads. The location of two modified nucleotides in helix 12 of the *E. coli* structure (A) has been determined by Felden *et al.* (28). The four pseudoknots are marked pk1–pk4 [pk4A and pk4B in (D)]. The solid star marks the beginning of the tmRNA coding region. In-frame stop codons are marked by a solid arrowhead. The deduced tmRNA-encoded sequence of the tag peptide is shown at the bottom of each panel in the one letter amino acid code.

Helix 6

Helix 6 consists of four helical sections, 6a–6d, and, together with helix 7, forms the second pseudoknot. With the exception of the plastid sequences (which lack pseudoknots 2–4 altogether), 6a is present in all tmRNAs and is well supported by CBCs. There are indications (Fig. 1A and B) that a near-continuous stack forms with sections 6b and 6c. In most Gram-positive bacteria, section 6b is absent (Fig. 1E). Interestingly, in three purple bacteria (*Neisseria gonorrhoeae*, *Neisseria meningitidis* and *A. eutrophus*) 6b is replaced by a short hairpin (named section 6d) that is closed by a tetranucleotide loop (tetraloop) of the GNRA-type (26; Figs 1C and 3B).

Helix 7

Absent in plastids, helix 7 is well supported by CBCs in all sequences and preceded by a loop of at least six residues. This number of residues seems to meet the steric requirement for maintaining the pseudoknot (Fig. 3A and B). Usually, helix 7 is connected to helix 8 by no more than five residues, except in *Desulfovibrio desulfuricans* (15 residues) and *Legionella pneumophila* (28 residues).

Helix 8

Helix 8 is divided into two sections, 8a and 8b, part of pseudoknot 3, and is present in all sequences except the plastids. The helical sections are likely to be continuously stacked, in some cases containing non-Watson-Crick interactions (Fig. 1B), as demonstrated by several instances where a single helix 8 is formed in other organisms (Fig. 1D).

Helix 9

Helix 9 is part of pseudoknot 3 (absent in the plastids) and well supported by CBCs. Most secondary structures display an immediate transition between the 5'-half of helix 9 and the 3'-half of helical section 8b. No more than three residues are inserted between the 5'-half of helical section 8 and helix 9. Modeling using ERNA-3D software (27) showed that a supported pairing in the helical section 8b of *Aquifex aeolicus* is sterically unfavorable (not shown). We prefer an equally well-supported structure which can exist in three dimensions and where this potential pair is replaced with a G-C located in helix 9. Six to 11 residues connect helix 9 with the 3'-half of helix 8a. These extended connections are found in all four pseudoknots and are usually rich in adenosine residues. Preliminary 3-dimensional model building indicates that the elongated single strands are required to maintain the pseudoknot fold (Fig. 3).

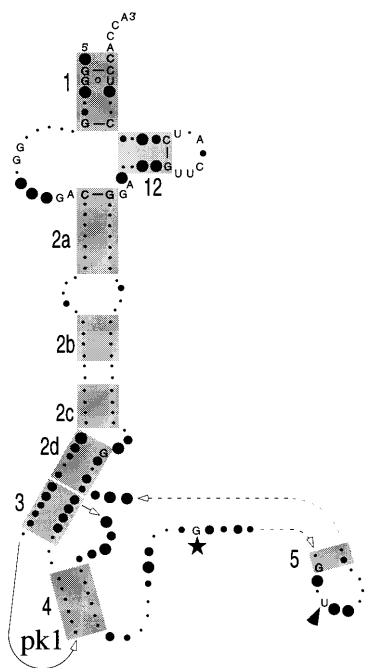


Figure 2. Secondary structure of helices 1–5 and helix 12 of *E.coli* with the connecting region indicated by the dashed arrows. The region between helices 2 and 5 corresponds to pseudoknots 2–4 and has been omitted because it lacks conserved residues. The overall sequence conservation between the corresponding bases in the other tmRNA sequences is shown. Invariant bases are shown by letters; dot diameters indicate the extent of conservation: large dots mark positions that are conserved in >90% of sequences; medium size dots indicate conservation in 75% of sequences; small dots indicate conservation in >50% of sequences. The secondary structure is annotated as described in the legend to Figure 1.

Helix 10

Helix 10 forms pseudoknot 4 with helix 11 (pseudoknot 4 is absent in plastids and Cyanobacteria) and is divided into three sections. Sections 10a and 10b appear as a continuous stack with up to 10 bp (e.g. *M.pneumoniae*; Fig. 1E). Some sequences contain small insertions with up to four single-stranded residues between sections 10b and 10c. An adenosine-rich strand of at least eight residues connects the 3'-half of helix 10a with the 3'-half of helical section 11b (Fig. 3C).

Helix 11

Helix 11 is a part of pseudoknot 4 with two short helical sections, 11a (four pairs) and 11b (three pairs), which might form one uninterrupted helix (e.g. *N.gonorrhoeae*; Fig. 1C). Three and sometimes four residues connect helix section 11a with section 2d. With only two exceptions (*C.tepidum* and the chloroplast of *P.purpurea*), the first residue of the triplet is a uridine. The second residue is a conserved adenosine, which is replaced by a guanine in *T.maritima* and *C.acetobutylicum*. The third residue is variable, but never a guanine.

Helix 12

Helix 12 is located in the tRNA-like part of the tmRNA molecule and is preceded by a triplet of conserved nucleotides. The first

residue is a guanine, with one exception in the *C.paradoxa* cyanelle, where it is an adenine. The second residue is conserved as an adenosine and the third residue is almost always a cytidine, except in *Mycoplasma genitalium* and *M.pneumoniae*.

The five base pairs and the seven residue loop formed by helix 12 are equivalent to the TΨC stem-loop of tRNA. Similarities include an invariant G-C closing pair and the same distribution of invariant residues in the loop (21). Furthermore, *E.coli* tmRNA was shown to contain a 5-methylated uridine and a pseudouridine at positions 341 and 342, respectively (28). As in tRNA, the m⁵UΨ in the loop might pair to the conserved guanosines in the DHU loop equivalent. However, this possible pairing cannot be supported nor disproven by CBCs and will have to be investigated experimentally. As in tRNA, helix 12 connects directly to helix 1, the amino acid acceptor stem, discussed above. Finally, an invariant ACCA forms the 3'-end of all tmRNAs.

Helix 13

Together with helices 14–16, helix 13 is present only in the Cyanobacteria, where pseudoknot 4 is replaced by two smaller pseudoknots, 4A and 4B (Fig. 1D). Because only three sequences are available, the comparative sequence analysis support for these pairings is weak and no covariation score was assigned to the proposed helices. However, they are included here as they provide the best preliminary explanation for the folding of the tmRNA in this region. Of the four helices, the 5 bp forming helix 13 are supported by one C-G→U-G change. The 5'-half of helix 13 connects to the 5'-half of helix 14 with one or two residues.

Helix 14

Helices 13 and 14 form pseudoknot 4A in the Cyanobacteria. Considering the small number of sequences, there is good phylogenetic support for three of the five proposed base pairs. The 5'-half of helix 14 and the 3'-half of helix 13 are in immediate transition, whereas three to four residues connect the 3'-halves of helices 13 and 14. No nucleotides are inserted between helices 14 and 15.

Helix 15

Helices 15 and 16 form pseudoknot 4B in the Cyanobacteria. Helix 15 is supported only by one G-U→A-U change and even contains a U-C mismatch in *Synechocystis* (Fig. 1D). Two cytosine residues connect the 5'-half of helix 15 to helix 16.

Helix 16

Helix 16 contains 4 or 5 bp of which three pairs are supported by CBCs. The 5'-half of this helix uses two or three residues to connect to the 3'-half of helix 15. Three to four residues are present between the 3'-halves of helices 15 and 16.

Plastid tmRNA secondary structure

Thermodynamic calculation (25) of the plastid sequences in the region between helices 2 and 4 suggest the potential to form two to four helices. However, because this region is unusually AU-rich and of variable length, it is difficult to align unambiguously for the identification of covariations. The comparative sequence analysis approach may become feasible when more plastid sequences become available. Experimental evidence will be required to

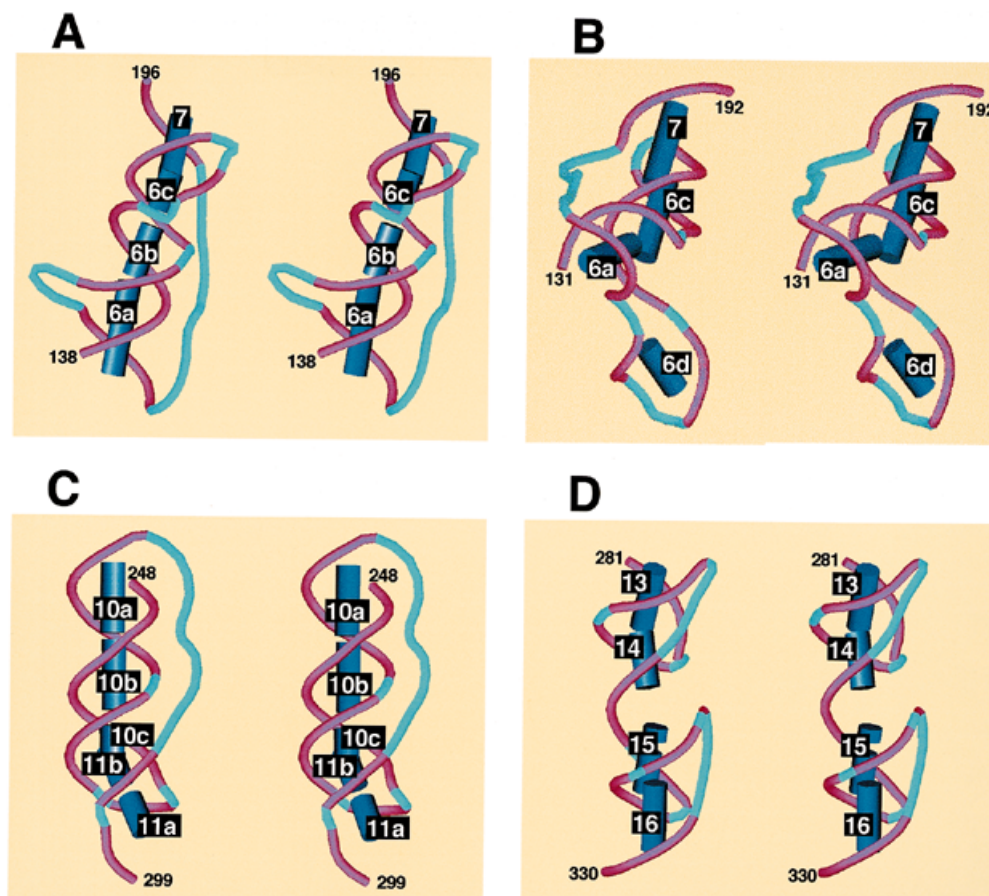


Figure 3. Stereodiagrams of representative tmRNA pseudoknots. (A) Pseudoknot 2 of *E.coli*; (B) pseudoknot 2 of *N.gonorrhoea*; (C) pseudoknot 4 of *E.coli*; (D) pseudoknot 4 of *Synechococcus* spp. Helices are indicated by blue cylinders of a diameter of 5 Å and are labeled according to the nomenclature shown in the respective secondary structures (see Sequence alignment and Fig. 1). The sugar–phosphate backbone is displayed as a tube (diameter 2 Å) where the helical regions are colored purple and the single-stranded regions are colored blue–green. The termini of the RNA chains are numbered to correspond to the nucleotide positions of the respective sequences. The models were generated with the program ERNA-3D (27) on a Silicon Graphics Indigo 2 Extreme workstation; the PDB coordinates of the models are available from the tmRDB at the internet address psyche.uthct.edu/dbs/tmRDB/

show which base pairings occur in this region or if, in fact, it lacks secondary interactions.

PSEUDOKNOTS

Four pseudoknots (labeled pk1–pk4 in Fig. 1) are present in most tmRNAs. Pseudoknot 1 precedes the tag peptide-encoding region and three pseudoknots flank the 3′-part of the same region. Sequence comparison shows that the pseudoknots lack conserved residues, thus emphasizing the requirement for RNA folding. In some bacteria, certain pseudoknots appear to be a dispensable feature of the tmRNAs, as shown in the case of *M.genitalium* and *M.pneumonia*, where pseudoknot 1 is absent. The plastid tmRNA may lack pseudoknots 2–4 altogether (Fig. 1F). However, further analysis, both comparative and experimental, will be required to demonstrate base pairing in this region. Particularly interesting is the case of the Cyanobacteria, where pseudoknot 4 is replaced by two smaller pseudoknots, again indicating that the RNA must be folded somehow, but there is no need for maintaining a particular pseudoknot fold. Preliminary model building indicates 3-dimensional similarities between a typical single pseudoknot 4

and the tandem pseudoknots 4A and 4B (Fig. 3). An argument for preservation of RNA folding without the need for sequence conservation can also be made for three relatives of the purple bacteria (*N.gonorrhoeae*, *N.meningitidis* and *A.eutrophus*) which replace helical section 6b with 6d (see above).

HIGHER ORDER INTERACTIONS

The similarity of tmRNA to the amino acceptor stem–loop and TΨC stem–loop of tRNA is supported not only by comparative sequence analysis but also by the biochemical reactions for which tmRNA is a substrate (17,29,30). Moreover, the 3′ maturation of tmRNA closely resembles that of tRNA in that multiple exoribonucleases can participate in 3′ trimming reactions and nucleotidyl transferase can rebuild its CCA terminus (17,31). However, our data indicate that the tmRNA region which is equivalent to the DHU stem–loop of tRNA, is significantly different, as is the region corresponding to the anticodon stem and loop. More detailed biophysical studies of tmRNA structure and of the interactions of tmRNA with the ribosome will be required to resolve the degree of structural and functional similarity between

tRNA and tmRNA. Besides the base pairings shown in Figure 1, we were unable to identify additional, so-called higher order, tertiary interactions within the tmRNA molecule. This may be due to the limitations of the comparative sequence analysis approach when only a relatively small number of sequences is available.

As tmRNA is involved in translation, we also searched for covariations between tmRNA and the RNAs of the small and large ribosomal subunits with rRNA alignments provided by the Ribosomal Database Project (32). Most likely due to the limited number of organisms for which both the rRNA and tmRNA sequences are known, we found no firm evidence for RNA–RNA interactions between tmRNA and rRNA. The earlier proposed pairing of tmRNA with a region in the 3'-domain of *E.coli* 16S rRNA (33) was unsupported by phylogenetic criteria (data not shown). Because of common structural features, interactions between tmRNA and the ribosome may involve rRNA regions that have been shown to facilitate binding of tRNA and mRNA by zero length cross-linking (34,35) and footprinting (36). However, these potential interactions may not be detectable within the limitations of the comparative sequence analysis method because of the involvement of conserved residues.

OUTLOOK

The phylogenetic approach provides a rational basis from which to infer minimal secondary structure models of tmRNA for all organisms in which this molecule is present. The database is sufficiently large to identify interesting structural features, some of which are specific to certain bacterial subgroups. Currently, the number of available sequences is too small to prove or disprove possible higher order interactions conclusively within tmRNA or between tmRNA and rRNA. This limitation is expected to disappear in the future when sequence data are more complete.

In the current early stage of tmRNA research the present sequence alignment and secondary structure analyses will assist in the design and interpretation of biochemical experiments such as enzymatic and chemical modification, site-directed mutagenesis and cross-linking. Furthermore, present results will be useful in comparative molecular modeling analyses (37) aimed to resolve the 3-dimensional structures of a wide variety of tmRNA molecules. Ultimately, these studies are expected to provide insight into which components of the cell might associate with the tmRNA and, in particular, how tmRNA interacts with other elements of the translation machinery.

DATA DISTRIBUTION

The sequence alignment is available at <http://psyche.uthct.edu/dbs/tmRDB/tmRDB.html> in several formats including printable PostScript and PDF versions or by writing to the first author (Email: zwieb@uthct.edu).

ACKNOWLEDGEMENTS

We thank Drs Florian Müller and Richard Brimacombe (Max-Planck Institut für Molekulare Genetic, Berlin, Germany) for providing the ERNA-3D RNA modeling software. This work was supported by a National Institutes of Health grant, GM 58267, to J.W.

REFERENCES

- Lee, S.Y., Bailey, S.C. and Apirion, D. (1978) *J. Bacteriol.*, **133**, 1015–1023.
- Ray, B. and Apirion, D. (1979) *Mol. Gen. Genet.*, **174**, 25–32.
- Keiler, K. and Sauer, R. (1996) *J. Biol. Chem.*, **271**, 2589–2593.
- Keiler, K., Waller, P. and Sauer, R. (1996) *Science*, **271**, 990–993.
- Muto, A., Ushida, C. and Himeno, H. (1998) *Trends Biochem. Sci.*, **23**, 25–29.
- Atkins, J.F. and Gesteland, R.F. (1996) *Nature*, **379**, 769–771.
- Jentsch, S. (1996) *Science*, **271**, 955–6.
- Retallack, D.M. and Friedman, D.I. (1995) *Cell*, **83**, 227–235.
- Fox, G. and Woese, C. (1975) *Nature*, **256**, 505–507.
- Kaine, B. and Merkel, V. (1989) *J. Bacteriol.*, **171**, 4261–4266.
- Larsen, N. and Zwieb, C. (1991) *Nucleic Acids Res.*, **19**, 209–215.
- Wower, J. and Zwieb, C. (1999) *Nucleic Acids Res.*, **27**, 167.
- Williams, K. and Bartel, D. (1996) *RNA*, **2**, 1306–1310.
- Felden, B., Himeno, H., Muto, A., McCutcheon, J.P., Atkins, J.F. and Gesteland, R.F. (1997) *RNA*, **3**, 89–103.
- Hickerson, R.P., Watkins-Sims, C.D., Burrows, C.J., Atkins, J.F., Gesteland, R.F. and Felden, B. (1998) *J. Mol. Biol.*, **279**, 577–587.
- Williams, K. (1999) *Nucleic Acids Res.*, **27**, 165–166.
- Ushida, C., Himeno, H., Watanabe, T. and Muto, A. (1994) *Nucleic Acids Res.*, **22**, 3392–3396.
- Felden, B., Himeno, H., Muto, A., McCutcheon, J., Atkins, J. and Gesteland, R. (1997) *RNA*, **3**, 89–103.
- Brown, J. (1991) *Comput. Appl. Biosci.*, **7**, 391–393.
- Gilbert, D. (1996) Version 0.6 of the program Seqpup is available from <ftp://iubio.bio.indiana.edu/molbio/seqpup/>
- Soll, D., Abelson, J. and Schimmel, P. (1980) *Transfer RNA*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Egebjerg, J., Larsen, N. and Garrett, R. (1990) In Hill, W., Dahlberg, A., Garrett, R., Moore, P., Schlessinger, J. and Warner, J. (eds), *The Ribosome: Structure, Function and Evolution*. American Society of Microbiology, Washington, DC, pp. 168–179.
- Gottesman, S., Roche, E., Zhou, Y. and Sauer, R.T. (1998) *Genes Dev.*, **12**, 1338–1347.
- Gerth, U., Kruger, E., Derre, I., Msadek, T. and Hecker, M. (1998) *Mol. Microbiol.*, **28**, 787–802.
- Zuker, M. and Stiegler, P. (1981) *Nucleic Acids Res.*, **9**, 133–148.
- Woese, C.R., Winker, S. and Gutell, R.R. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 8467–8471.
- Müller, F., Döring, T., Erdemir, T., Greuer, B., Jünke, N., Osswald, M., Rinke-Appel, L., Stade, K., Thamm, S. and Brimacombe, R. (1995) *Biochem. Cell Biol.*, **73**, 767–773.
- Felden, B., Hanawa, K., Atkins, J.F., Himeno, H., Muto, A., Gesteland, R.F., McCloskey, J.A. and Crain, P.F. (1998) *EMBO J.*, **17**, 3188–3196.
- Komine, Y., Kitabatake, M., Yokogawa, T. and Nishikawa, K. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 9223–9227.
- Himeno, H., Sato, M., Tadaki, T., Fukushima, M., Ushida, C. and Muto, A. (1997) *J. Mol. Biol.*, **268**, 803–808.
- Li, Z., Pandit, S. and Deutscher, M.P. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 2856–2861.
- Maidak, B., Cole, J., Parker, C.J., Garrity, G., Larsen, N., Li, B., Lilburn, T., McCaughey, M., Olsen, G., Overbeek, R., Pramanik, S., Schmidt, T., Tiedje, J. and Woese, C. (1999) *Nucleic Acids Res.*, **27**, 171–173.
- Muto, A., Sato, M., Tadaki, T., Fukushima, M., Ushida, C. and Himeno, H. (1996) *Biochimie*, **78**, 985–991.
- Wower, J., Hixson, S.S. and Zimmermann, R.A. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 5232–5236.
- Wower, J., Wower, I.K., Kirillov, S.V., Rosen, K.V., Hixson, S.S. and Zimmermann, R.A. (1995) *Biochem. Cell Biol.*, **73**, 1041–1047.
- Moazed, D. and Noller, H.F. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 3725–3728.
- Zwieb, C. and Müller, F. (1997) *Nucleic Acids Symp. Ser.*, **36**, 69–71.
- Deckert, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., Overbeek, R., Snead, M.A., Keller, M., Aujay, M., Huber, R., Feldman, R.A., Short, J.M., Olsen, G.J. and Swanson, R.V. (1998) *Nature*, **392**, 353–358.
- TIGR, http://www.tigr.org/tigr_home/index.html
- Bekke, A., Kiefmann, M., Kremerskothen, J., Vormlocher, H.P., Sprinzl, M. and Brosius, J. (1998) *DNA Seq.*, **9**, 31–35.
- Stephens, R.S., Kalman, S., Lammel, C., Fan, J., Marathe, R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, R.L., Zhao, Q., Koonin, E.V. and Davis, R.W. (1998) *Science*, **282**, 754–759.

- 42 McCam,D.F., Whitaker,R.A., Alam,J., Vrba,J.M. and Curtis,S.E. (1988) *J. Bacteriol.*, **170**, 3448–3458.
- 43 Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y., Miyajima,N., Hirosawa,M., Sugiura,M., Sasamoto,S., Kimura,T., Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Naruo,K., Okumura,S., Shimpo,S., Takeuchi,C., Wada,T., Watanabe,A., Yamada,M., Yasuda,M. and Tabata,S. (1996) *DNA Res.*, **3**, 109–136.
- 44 Watanabe,T., Sugita,M. and Sugiura,M. (1997) GenBank accession no. AB003477.
- 45 Kowalik,K., Stoebe,B., Schaffran,I., Kroth-Pancic,P. and Freier,U. (1995) GenBank accession no. Z67753.
- 46 Reith,M. and Munholland,J. (1995) GenBank accession no. U38804.
- 47 Douglas,S., Durnford,D. and Morden,C. (1989) GenBank accession no. AF063017.
- 48 Gueneau,P., Morel,F., Laroche,J. and Erdner,D. (1998) GenBank accession no. AF049491.
- 49 Stirewalt,V., Michalowski,C., Luffelhardt,W., Bohnert,H. and Bryant,D. (1995) GenBank accession no. U30821.
- 50 Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R., Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K. *et al.* (1997) *Nature*, **390**, 580–586.
- 51 Williams,K. and Bartel,D. (1996) GenBank accession no. U68080.
- 52 The Sanger Center, <http://www.sanger.ac.uk/>
- 53 Brown,J.W., Hunt,D.A. and Pace,N.R. (1990) *Nucleic Acids Res.*, **18**, 2820.
- 54 Billington,S., Huggins,A., Johanesen,P., Crellin,P., Cheung,J., Katz,M., Wright,C., Haring,V. and Rood,J. (1995) GenBank accession no. U20246.
- 55 Pseudomonas Genome Project, <http://www.pseudomonas.com/>
- 56 Genome Sequencing Center Washington University, <http://genome.wustl.edu/gsc/bacterial/salmonella.shtml>
- 57 ACGT University of Oklahoma, <http://www.genome.ou.edu/>
- 58 Fleischmann,R., Adams,M., White,O., Clayton,R., Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J., Dougherty,B.A., Merrick,J.M. *et al.* (1995) *Science*, **269**, 496–512.
- 59 Tomb,J.F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G., Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S., Dougherty,B.A. *et al.* (1997) *Nature*, **388**, 539–547.
- 60 Tyagi,J.S. and Kinger,A.K. (1992) *Nucleic Acids Res.*, **20**, 138.
- 61 Zwieb,C. and Black,S. (1997) GenBank accession no. AF019911.
- 62 Fraser,C.M., Gocayne,J.D., White,O., Adams,M.D., Clayton,R.A., Fleischmann,R.D., Bult,C.J., Kerlavage,A.R., Sutton,G., Kelley,J.M. *et al.* (1995) *Science*, **270**, 397–403.
- 63 Himmelreich,R., Hilbert,H., Plagens,H., Pirkl,E., Li,B.C. and Herrmann,R. (1996) *Nucleic Acids Res.*, **24**, 4420–4449.
- 64 Genome Therapeutics Corporation, http://www.genomecorp.com/sequence_center/bacterial_genomes/