Nucleic Acids Research, 1998, Vol. 26, No. 1 **163–165**

The tmRNA Website

Kelly P. Williams* and David P. Bartel

Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, MA 02142, USA

Received September 10, 1997; Accepted September 25, 1997

ABSTRACT

tmRNA (also known as 10Sa RNA) is so-named for its dual tRNA-like and mRNA-like nature. It is employed in a remarkable *trans*-translation process to add a C-terminal peptide tag to the incomplete protein product of a broken mRNA; the tag targets the abnormal protein for proteolysis. tmRNA sequences have been identified in genomes of diverse bacterial phyla, including the most deeply branching. They have also been identified in plastids of the 'red' lineage. The tmRNA Website (http://www.wi.mit.edu/bartel/tmRNA/home) contains a database currently including sequences from 37 species, with provisional alignments, as well as the tentatively predicted proteolysis tag sequences. A brief review and guide to the literature is also provided.

tmRNA STRUCTURE AND FUNCTION

The bacterial RNA with the provisional biochemical designation 10Sa (1,2) was renamed tmRNA (3,4) when its combined tRNA-like and mRNA-like properties were recognized. A half-tRNA structure, with a coaxially stacked T stem–loop and acceptor stem–tail, was identified in tmRNA upon determination of the ends of the *Escherichia coli* and *Bacillus subtilis* molecules and comparison with other available gene sequences (5,6); the CCA tail is not always fully encoded in the genome, but has nonetheless been found at the 3′ end of the *B.subtilis* tmRNA (6). The acceptor stem has the simple identity elements of tRNA(Ala) (7,8), and tmRNA is alanylated *in vitro* and *in vivo* (5). The secondary structure of *E.coli* tmRNA was revealed by phylogenetic analysis (9) and by chemical probing (10). A long disrupted stem (P2) exits the tRNA-like domain, capped by a large loop composed of a pseudoknot, the tag reading frame with the stop codon in the loop of a hairpin, and a string of three pseudoknots.

Coding by tmRNA was revealed in a careful study of the smaller products that accumulated at low abundance during overexpression of a foreign gene in *E.coli* (11). The series of proteins were all truncated to a different extent at the C-terminus, and all had the same C-terminal peptide tag (A)ANDENYALAA, unrelated to the overexpressed gene. The reading frame for all but the parenthetical alanyl residue of the tag was found in the *E.coli* tmRNA sequence. Knowledge of the determinants of susceptibility to an *E.coli* protease (12) led to the finding that the tmRNA-directed tag sequence causes degradation of tagged proteins in the periplasm and in the cytoplasm (13). Tagging can be directed to proteins translated from 'broken' mRNAs, i.e. mRNAs whose reading frame has no stop codon, demonstrated *in vivo* by placing a terminator of transcription inside the reading frame of a target gene (13).

In the *trans*-translation model for tmRNA action (13): (i) the ribosome, having translated to the end of a broken mRNA, signals entry of the tmRNA into its A site, (ii) the nascent polypeptide is transferred to the alanyl residue that charges the tmRNA, which becomes the parenthetical residue of the (A)ANDENYALAA tag, (iii) the broken mRNA is replaced at the decoding site by the tmRNA reading frame, and (iv) translation resumes at a specific codon of the tmRNA and stops normally, yielding a substrate for proteolysis. Thus tmRNA promotes the destruction of the abnormally short products of broken mRNAs in bacteria.

The tmRNA gene *ssrA* has been grossly disrupted in three different *E.coli* K-12 strains (5,14,15) and another mutant alters a single but important base of the mature tmRNA that determines tRNA(Ala) identity (15,16). All these mutants are viable, but diverse phenotypes have been noted whose relationships are unclear (5,13–18).

CONSTRUCTION OF A tmRNA DATABASE

To better understand the structure of tmRNA and the unusual aspects of *trans*-translation, we have attempted to identify and collect all available tmRNA sequences in a single resource. This was especially important since public files containing tmRNA sequences have not all been identified, and several are present only in smaller genome-dedicated databases.

The computer programs Blast (19) and PatScan (R.Overbeek) have allowed the identification in sequence databases of many tmRNA genes from diverse bacterial phyla, including the deepest phylogenetic branches. Comparative sequence analysis revealed that he *E.coli* secondary structure model generally fits distant relatives. The new sequences indicate that the long P2 paired region (9) is more general than an alternative proposal (10). The paired regions in tmRNAs from thermophilic species exhibit trends found in RNase P RNAs from thermophiles (20): G-C richness, reduced non-Watson–Crick pairing and fewer disruptions. Although only two cyanobacterial sequences are currently available, covariations suggest that the downstream pseudoknot is replaced by two directly apposed pseudoknots. tmRNA genes have been found in three plastids of the 'red' lineage (from a red

* To whom correspondence should be addressed. Tel: +1 617 258 8346; Fax: +1 617 258 9872; Email: kelly@wi.mit.edu

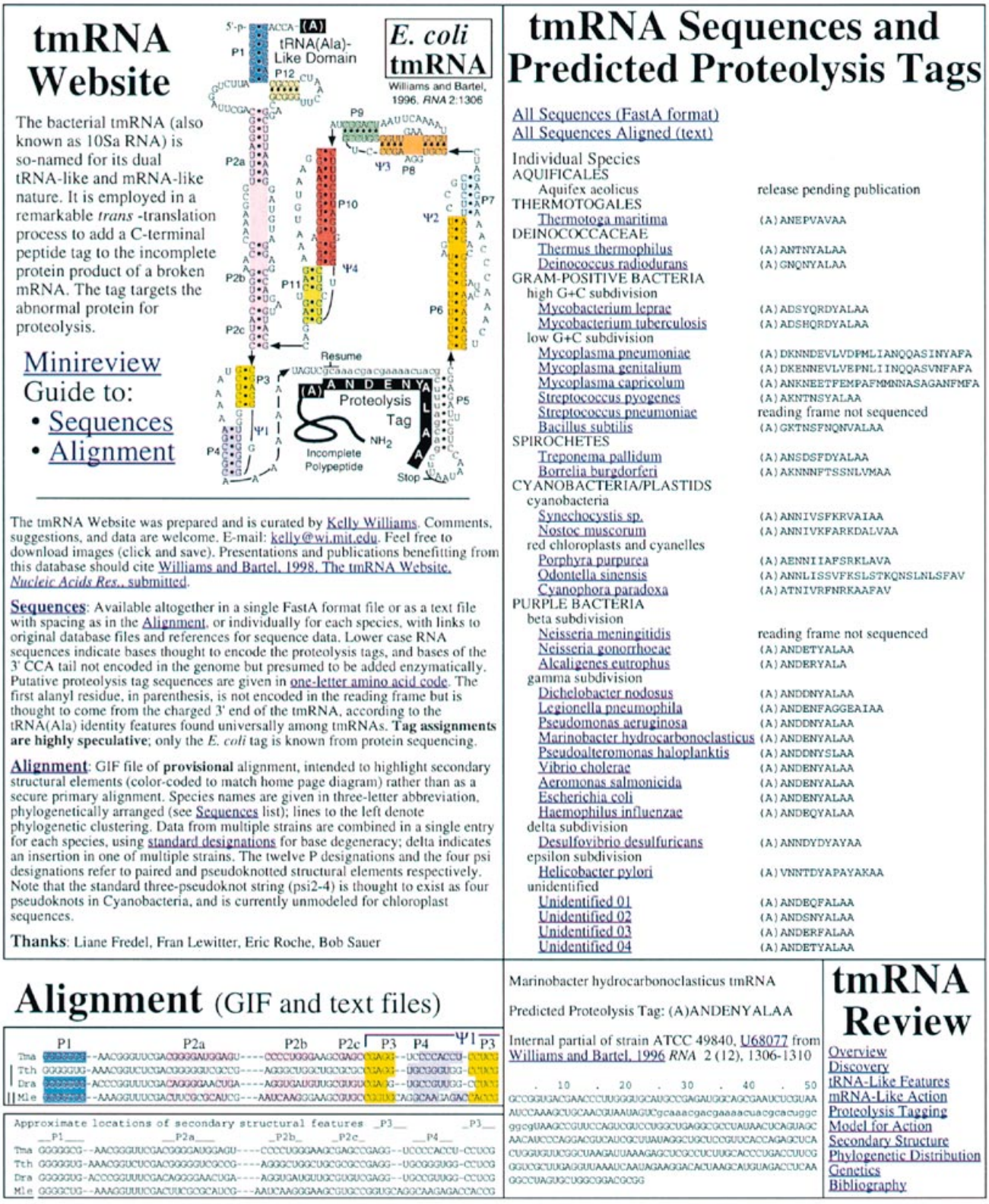

Figure 1. The tmRNA Website. The home (upper left) and other pages from the website are shown, with links in blue and underlined.

alga, a diatom and a colorless alga), but not yet in any of the green lineage (green algae and higher plants). Primary alignment of the plastid sequences in the three-pseudoknot string region is currently considered so unclear as to preclude secondary structure prediction, however it can be noted that these regions in plastids are by far the shortest known, which is puzzling in light of the

apparent expansion to four pseudoknots in their closest bacterial relatives. tmRNA genes have not yet been found in Archaea or in eukaryotic nuclei or mitochondria.

tmRNA primary sequence is most conserved at the termini. This allows the amplification and determination of tmRNA sequences from type specimens (9), or, as pioneered for rRNA and RNase P (21–23), from the ensemble of microorganisms in a field sample. For example, in a preliminary search for a tmRNA gene in the nucleus of a green plant, using PCR primers from red plastids and an *Arabidopsis* leaf DNA sample, four new tmRNA sequences were found (K.P.W. and D.P.B., unpublished). The sequences were more similar to those of purple bacteria than to those of cyanobacteria or chloroplasts, and were therefore ascribed to bacteria present in the leaf sample.

The tmRNA Website (Fig. 1) is accessible via WWW at http://www.wi.mit.edu/bartel/tmRNA/home . It currently contains tmRNA sequences from 37 species, in a single FastA format file, and individually with links to original database files and references to sources of sequence data. Predicted proteolysis tag sequences are provided; most of these are highly speculative. The tag reading frame has been ascertained only for *E.coli* (11,13) and although the stop codon can be readily identified in other species by the hydrophobicity of the encoded C-terminal residues and other considerations, determinants of the resume codon are not yet known.

A provisional alignment of the tmRNA sequences, intended to highlight secondary structural elements rather than as a secure primary alignment, is provided as a GIF document, with structural features color-coded to match the diagram in the home page. The alignment is also available as an (uncolored) text file. A brief review of tmRNA and bibliography are also provided. Presentations and publications benefitting from the tmRNA Website should cite this article. Users are encouraged to submit new data, which can be withheld until release is desired by the author.

ACKNOWLEDGEMENTS

We thank Fran Lewitter and Liane Fredel for help in preparing the website, Christian Luschnig for an *Arabidopsis* DNA sample, and Bob Sauer and Eric Roche for useful discussions.

REFERENCES

1 Ray,B.K. and Apirion,D. (1979) *Mol. Gen. Genet.*, **174**, 25–32.
2 Jain,S.K., Gurevitz,M. and Apirion,D. (1982) *J. Mol. Biol.*, **162**, 515–533.
3 Jentsch,S. (1996) *Science*, **271**, 955–956.
4 Atkins,J.F. and Gesteland,R.F. (1996) *Nature*, **379**, 769–771.
5 Komine,Y., Kitabatake,M., Yokogawa,T., Nishikawa,K. and Inokuchi,H. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 9223–9227.
6 Ushida,C., Himeno,H., Watanabe,T. and Muto,A. (1994) *Nucleic Acids Res.*, **22**, 3392–3396.
7 Francklyn,C. and Schimmel,P. (1989) *Nature*, **337**, 478–481.
8 Shi,J.P. and Schimmel,P. (1991) *J. Biol. Chem.*, **266**, 2705–2708.
9 Williams,K.P. and Bartel,D.P. (1996) *RNA*, **2**, 1306–1310.
10 Felden,B., Himeno,H., Muto,A., McCutcheon,J.P., Atkins,J.F. and Gesteland,R.F. (1997) *RNA*, **3**, 89–103.
11 Tu,G.F., Reid,G.E., Zhang,J.G., Moritz,R.L. and Simpson,R.J. (1995) *J. Biol. Chem.*, **270**, 9322–9326.
12 Keiler,K.C. and Sauer,R.T. (1996) *J. Biol. Chem.*, **271**, 2589–2593.
13 Keiler,K.C., Waller,P.R. and Sauer,R.T. (1996) *Science*, **271**, 990–993.
14 Oh,B.K. and Apirion,D. (1991) *Mol. Gen. Genet.*, **229**, 52–56.
15 Kirby,J.E., Trempy,J.E. and Gottesman,S. (1994) *J. Bacteriol.*, **176**, 2068–2081.
16 Retallack,D.M., Johnson,L.L. and Friedman,D.I. (1994) *J. Bacteriol.*, **176**, 2082–2089.
17 Retallack,D.M. and Friedman,D.I. (1995) *Cell*, **83**, 227–235.
18 Ando,H., Kitabatake,M. and Inokuchi,H. (1996) *Genes Genet. Syst.*, **71**, 47–50.
19 Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) *Nucleic Acids Res*, **25**, 3389–3402.
20 Brown,J.W., Haas,E.S. and Pace,N.R. (1993) *Nucleic Acids Res.*, **21**, 671–679.
21 Reysenbach,A.L., Giver,L.J., Wickham,G.S. and Pace,N.R. (1992) *Appl. Environ. Microbiol.*, **58**, 3417–3418.
22 Amann,R.I., Ludwig,W. and Schleifer,K.H. (1995) *Microbiol. Rev.*, **59**, 143–169.
23 Brown,J.W., Nolan,J.M., Haas,E.S., Rubio,M.A., Major,F. and Pace,N.R. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 3001–3006.