Nucleic Acids Research, 2006, Vol. 34, Database issue ***D163–D168***
doi:10.1093/nar/gkj142

The tmRDB and SRPDB resources

Ebbe Sloth Andersen, Magnus Alm Rosenblad[1,2], Niels Larsen[3], Jesper Cairo Westergaard[4], Jody Burks[5], Iwona K. Wower[5], Jacek Wower[5], Jan Gorodkin[6], Tore Samuelsson[1] and Christian Zwieb[7,*]

Department of Molecular Biology, University of Aarhus, C.F. Moellers Allé, Building 139, 8000 Aarhus, Denmark, [1]Department of Medical Biochemistry, Goteborg University, Box 440, SE-405 30 Goteborg, Sweden, [2]SWEGENE Bioinformatics, Goteborg University, Box 413, SE-405 30 Goteborg, Sweden, [3]Danish Genome Institute, Gustav Wieds vej 10 C, DK-8000 Aarhus C, Denmark, [4]Department of Natural Sciences, The Royal Veterinary and Agricultural University, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark, [5]Department of Animal Sciences, Auburn University, Auburn, AL 36849-5415, USA, [6]Center for Bioinformatics and Division of Genetics, IBHV, The Royal Veterinary and Agricultural University, Groennegaardsvej 3, 1870 Frederiksberg C, Denmark and [7]Department of Molecular Biology, The University of Texas Health Science Center at Tyler, 11937 US Highway 271, Tyler, TX 75708-3154, USA

Received September 15, 2005; Revised and Accepted October 27, 2005

ABSTRACT

Maintained at the University of Texas Health Science Center at Tyler, Texas, the tmRNA database (tmRDB) is accessible at the URL http://psyche.uthct.edu/dbs/tmRDB/tmRDB.html with mirror sites located at Auburn University, Auburn, Alabama (http://www.ag.auburn.edu/mirror/tmRDB/) and the Royal Veterinary and Agricultural University, Denmark (http://tmrdb.kvl.dk/). The signal recognition particle database (SRPDB) at http://psyche.uthct.edu/dbs/SRPDB/SRPDB.html is mirrored at http://srpdb.kvl.dk/ and the University of Goteborg (http://bio.lundberg.gu.se/dbs/SRPDB/SRPDB.html). The databases assist in investigations of the tmRNP (a ribonucleoprotein complex which liberates stalled bacterial ribosomes) and the SRP (a particle which recognizes signal sequences and directs secretory proteins to cell membranes). The curated tmRNA and SRP RNA alignments consider base pairs supported by comparative sequence analysis. Also shown are alignments of the tmRNA-associated proteins SmpB, ribosomal protein S1, alanyl-tRNA synthetase and Elongation Factor Tu, as well as the SRP proteins SRP9, SRP14, SRP19, SRP21, SRP54 (Ffh), SRP68, SRP72, cpSRP43, Flhf, SRP receptor (alpha) and SRP receptor (beta). All alignments can be easily examined using a new exploratory browser. The databases provide links to high-resolution structures and serve as depositories for structures obtained by molecular modeling.

Ribosomes extend their repertoire of functions by binding to additional ribonucleoprotein particles (RNPs) that can determine the fate of the protein as it emerges from the large ribosomal subunit. Two such complexes are the transfer-messenger RNP (tmRNP) and the signal recognition particle (SRP). The tmRNP, composed of the tmRNA, small protein B (SmpB) and ribosomal protein S1, rescues bacterial ribosomes stalled on faulty mRNAs. The potentially damaging polypeptides are tagged with a short peptide, released from the ribosome and destroyed by intracellular proteases [reviewed in (1)]. Similarly, the SRP binds to emerging signal sequences and directs secretory protein to cellular membranes [recently reviewed in (2)]. The investigations of tmRNP and SRP combined with the knowledge gained from the high-resolution structures of the ribosome (3–5) have contributed significantly to our understanding of protein translation and translocation, but many questions remain to be answered. To assist in the ongoing studies, the updated tmRDB and SRPDB resources offer detailed descriptions of the biological roles of tmRNP and SRP, ordered lists of the components and links to high-resolution structures. Alignment-derived RNA secondary structures are supported by comparative sequence analysis. A new browser allows the user to easily explore the alignments.

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

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

© The Author 2006. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

MATERIALS AND METHODS

Comparative sequence analysis of RNA

New tmRNA sequences provided at the tmRNA website (6) were merged with the previous tmRNA alignment (7). New SRP RNAs were identified using SRPscan (8) or combinations of BLAST (9), RNABOB (Eddy, unpublished data) and Infernal (10) with secondary structure predictions by MFOLD (11). The sequences were placed in phylogenetic order guided by NCBI Taxonomy (12,13). Sequences were aligned automatically with CLUSTAL (14) or manually using BioEdit (15) observing the previously described covariation rules (16). The RNA editor SARSE (A. Lind-Thomsen *et al*., 2005, manuscript in preparation) was used for semi-automated cleanup of the alignments. RNAdbtools (17) was applied to confirm compensatory base changes, check base pairing consistencies and possible RNA helix extensions. Pfold (18) was used to predict the secondary structure of subgroups of the alignment.

Protein alignments

Protein sequences were identified in GenBank (13) using BLAST (9) with a subset of representative sequences from the previous versions of tmRDB (7) and SRPDB (19) as queries. The output was examined manually to generate a set of unique sequences for each protein family. Sequences were aligned using Jalview (20), CLUSTAL (14) and MUSCLE (21).

Alignment browser

The alignments can be viewed, zoomed and scrolled in a www-browser under development for genomes by the Danish Genome Institute (also directly accessible at http://www.genomics.dk:8000/RNA). It currently features basic navigation, with color-dot, grey-dot, character display and zoom to any level. More features will be added.

RESULTS AND DISCUSSION

tmRNA genes

The tmRDB contains a total of 555 tmRNA sequences in the range of 250–434 nt. Because of the continuous rapid emergence of new sequences this dataset is not complete but nevertheless representative. [The tmRNA website (6) can be consulted for the most recent new tmRNA sequences.] All bacterial groups, including the Alphaproteobacteria (55 sequences) previously thought to lack tmRNA, contained tmRNA genes. Consistent with the evolutionary relationship between bacteria and organelles, tmRNAs were found in most of the chloroplast and mitochondrial genomes. However, tmRNA genes were lacking in the chloroplasts of higher plants. Interestingly, tmRNAs could be identified in the genomes of certain bacteriophages.

Most tmRNAs were composed of one continuous molecule. Less frequently, tmRNAs were encoded in the DNA in two sections which, when transcribed, are expected to fold into a tmRNA-like configuration. These two-part tmRNAs were found in the genomes of most Alphaproteobacteria, as well as in some Cyanobacteria and Betaproteobacteria (Table 1). The appearance of this adaptation in these distinct phylogenetic groups suggested that two-part tmRNAs arose in evolution three times independently (22). No tmRNA genes were identified in the archaea or the nuclear genomes of the eukarya.

Features of tmRNA

The tmRNA sequences were aligned using comparative sequence analysis as described previously for SRP RNA (16). An outline of the secondary structure of *Escherichia coli* tmRNA is depicted in Figure 1A. Shown are the tRNA-like domain (TLD), the messenger RNA-like domain (MLD), and the pseudoknot (pk) domain (PKD). Modification to the *E.coli* reference structure includes the reduction or deletion of pseudoknots, the appearance of new helices (e.g. in pk2 of Betaproteobacteria) and structural replacements, e.g. the change of pk4 into two tandem pseudoknots (see diagram **b** in Supplementary Data 1). The phylogenetic distribution of the features is summarized in Table 1.

tmRNA-encoded tag-peptides

A cluster of hydrophobic amino acids at the C-terminus and a variable length of 8–35 amino acids characterized the 539 tmRNA-encoded tag-peptides. Alanine or glycine were the most frequent resume codons. Tag peptide sequences have been experimentally confirmed for *E.coli* and *Bacillus subtilis*.

tmRNA-associated proteins

SmpB. This protein is an essential *trans*-translational co-factor (23) and is present in all bacteria. The protein forms quaternary complexes with aminoacylated tmRNA, EF-Tu and GTP (24). SmpB mutants which lack the C-terminal tail of the protein bind to ribosomes but are unable to tag the truncated proteins (25).

Ribosomal protein S1. This protein contains up to six related domains. The protein binds and cross-links to the MLD and pk2–pk4. The NMR structure of a single protein S1 RNA-binding domain of *E.coli* has been determined (26), but little is known about the arrangement of full-length protein S1 during *trans*-translation. The alignment suggested four groups of sequences which differed in the number of domains. Overall, domains four, five and six were less conserved and absent in some of the S1 homologues. The protein S1 sequences of *Candidatus Tremblaya princeps* and *Clostridium acetobutylicum ATCC 824* were distinct with respect to their low levels of homology to any other aligned sequence.

Alanyl-tRNA synthetase. Aminoacylation of tmRNA constitutes a prerequisite step in *trans*-translation, since uncharged tmRNA mutants do not bind to 70S ribosomes *in vivo* (27). Studies carried out *in vitro* demonstrated that the aminoacyl moiety can be changed without affecting the ability of the tmRNA to participate in protein tagging. The majority of the tmRNAs are expected to be charged with alanine because they posses in their acceptor stem a G-U basepair as the critical determinant for aminoacylation with alanyl-tRNA synthetase.

EF-Tu. Elongation factor Tu, found in all organisms, forms a ternary complex with GTP and Ala-tmRNA *in vitro*. EF-Tu primarily interacts with the acceptor arm of the tRNA-like domain of tmRNA (24). Although Ala-tmRNA has a lower association rate constant for the EF-Tu GTP complex than

Table 1. Phylogenetically ordered properties of the tmRNP

2D	Group *Species*	1	2	3	4	M	5	6	7	8	9	=	10	11	12	SB	S1	RS	Tu
a	Bacteriophages *Bacillus subtilis* phage G	x	x	x	x	x	x	x	x	x	x	–	–	–	x	–	–	–	–
	Bacteriophages CP1639	x	x	x	x	x	x	x	x	x	x	–	x	x	x	x	–	–	–
	Aquificae *Aquifex aeolicus*	x	x	x	x	x	x	x	x	x	x	–	x	x	x	x	x	x	x
	Deinococcus-Thermus *Thermus thermophilus*	x	x	x	x	x	x	x	x	x	x	–	x	x	x	x	x	x	x
	Thermodesulfobacteria *Thermodesulfobacterium commune*	x	x	x	x	x	x	x	x	x	x	–	x	x	x	x	x	x	x
	Thermatogae *Thermatoga maritima*	x	x	x	x	x	x	x	x	x	x	–	x	x	x	x	x	x	x
	Planctomyces Rhodopirellula baltica	x	x	x	x	x	x	!	!	x	x	–	x	x	x	x	x	x	x
	Clamydiae/Verrucomicrobia *Chlamydia trachomatis*	x	x	x	x	x	x	x	x	x	x	–	x	x	x	x	x	x	x
	Chloroflexi *Chloroflexus aurantiacus*	x	x	x	x	x	x	x	x	x	x	–	x	x	x	x	x	x	x
	Bacteroides/Chlorobi *Bacteroides fragilis*	x	x	x	x	x	x	x	x	x	x	–	x	x	x	x	x	x	x
	Bacteroides/Chlorobi *Salinibacter ruber*	x	x	x	x	x	x	x	?	x	x	–	x	x	x	x	x	x	x
b	Cyanobacteria *Synecystis PCC6803*	x	x	x	x	x	x	x	x	x	x	–	pp	pp	x	x	x	x	x
	Cyanobacteria *Cyanobium gracilis*	x	x	x	x	x	x	x	?	–	–	x	–	–	x	x	x	x	x
c	Organelles/Chloroplasts *Guillardia theta*	x	x	x	x	x	x	x	x	–	–	–	–	x	x	x	x	x	x
	Organelles/Chloroplasts *Thalassiosira pseudonana*	x	x	x	x	x	x	–	–	–	–	–	–	x	x	x	x	x	x
	Organelles/Mitochondria *Reclinomonas americana*	x	x	–	–	–	–	–	–	–	–	!	–	–	x	–	x	x	x
	Organelles/Mitochondria *Jakoba libera*	x	x	–	–	–	–	–	–	–	–	–	–	–	x	–	x	x	x
	Fibrobacteres/Acidobacteria *Fibrobacter succinogenes*	x	x	x	x	x	x	x	x	x	x	–	x	x	x	x	x	x	x
	Spirochaetes *Treponema pallidum*	x	x	x	x	x	x	x	x	x	x	–	x	x	x	x	x	x	x
	Nitrospirae *Leptospirillum species*	x	x	x	x	x	?	?	x	x	x	–	x	x	x	x	x	x	x
d	Alphaproteobacteria *Caulobacter crescentus*	x	x	x	x	x	x	x	x	x	x	x	?	x	x	x	x	x	x
	Alphaproteobacteria *Magnetococcus* MC-1	x	x	x	x	x	x	x	–	x	x	–	–	x	x	x	x	x	x
	Betaproteobacteria *Dechloromonas aromatica*	x	x	?	?	x	x	–	–	–	–	x	?	?	x	x	x	x	x
	Betaproteobacteria *Tremblaya princeps*	x	x	x	x	x	x	–	x	x	x	–	x	x	x	x	x	x	x
	Betaproteobacteria *Neisseria gonorrhoeae*	x	x	x	x	x	x	e	–	–	–	–	?	?	!	x	x	x	x
	Gammaproteobacteria *Francisella tularensis*	x	x	x	x	x	x	e	x	x	x	–	x	x	x	x	x	x	x
e	Gammaproteobacteria *Escherichia coli*	x	x	x	x	x	x	x	x	x	x	–	x	x	x	x	x	x	x
	Deltaproteobacteria *Geobacter metallireducens*	x	x	x	x	x	x	x	x	x	x	–	x	x	x	x	x	x	x
	Epsilonproteobacteria *Campylobacter jejuni*	x	x	x	x	x	x	x	x	x	x	–	x	x	x	x	x	x	x
	Fusobacteria *Fusobacterium nucleatum*	x	x	x	x	x	x	x	x	x	x	–	x	x	x	x	x	x	x
	Dictyoglomi *Dictyoglomus thermophilum*	x	x	x	x	x	x	x	x	x	?	–	?	x	x	x	x	x	x
	Actinomycetes *Mycobacterium avium*	x	x	x	x	x	x	x	x	x	x	–	x	x	x	x	x	x	x
	Firmicutes/Bacilli *B.subtilis*	x	x	x	x	x	x	x	x	x	x	–	x	x	x	x	x	x	x
	Firmicutes/Clostridia *Clostridium botulinum*	x	x	x	x	x	x	x	x	x	x	–	x	x	x	x	x	x	x

The names of representative species are given for each phylogenetic group in the tmRDB. The column labeled '2D' marks five tmRNA secondary structure examples *a–e* which are shown in more detail in Supplementary Data 1. The tmRNA features (helices numbered from 1 to 12) are shown in the center part of the table. '=' indicates the interruption in the two-part tmRNAs. SB, Protein SmpB; S1, ribosomal protein S1 and its homologues; RS, alanyl-tRNA synthetase; Tu, Elongation Factor Tu. The table cells are annotated as '–', absent; '?', maybe absent or was not found; '!', expected to be present, and 'x', present. 'e' denotes an extra helix; 'pp' is for a tandem pseudoknot.

Ala-tRNA, chemical and enzymatic footprinting indicate that the architecture of this complex closely resembles canonical ternary complexes.

Phylogeny of tmRNP

A description of the phylogenetic distribution of the secondary structural features of tmRNA based on an alignment of 274 sequences was provided recently (28). From the analysis of 555 sequences the following insights into tmRNA phylogeny were obtained: (i) most tmRNAs consist of a single polynucleotide chain with a TLD, a relatively unstructured MLD, and a variable number of pseudoknots. (ii) The variability of the predicted pseudoknot structures suggests a preservation of RNA folding without the need for sequence conservation. (iii) In the Alphaproteobacteria, some Betaproteobacteria, and some Cyanobacteria, the tmRNAs are composed of two chains. These two-piece tmRNAs contain fewer pseudoknots than the typical one-piece tmRNAs. (iv) Plastids contain one-piece tmRNAs with a reduced number of pseudoknots. (v) Most mitochondria may be devoid of *trans*-translation because they lack SmpB and contain only very short two-piece tmRNAs which appear to have lost the MLD. Examples of tmRNA secondary structure diagrams are shown in Supplementary Data 1.

SRP RNA genes

A total of 393 SRP RNAs were identified using the procedures described in Materials and Methods. SRP RNA genes were found to be present in all major phylogenetic groups as well as the photosynthetic plastids of red algal origin (except the substantially smaller plastid of the haptophyte *Emiliania huxleyi*) and the chloroplasts of some green algae (29). More than one variant were found in 33 organisms. Many novel SRP RNA sequences were found to add to our knowledge of the phylogenetic distribution of the secondary structure features (Table 2).

SRP RNA features

An overview of the SRP RNA secondary structure elements was presented in a recent nomenclature proposal (30) similar to what is shown in Figure 1B. Several new sequences, e.g. from *Eremothecium gossypii*, *Kluyveromyces waltii* and *Kluyveromyces lactis*, provided additional support for the proposed helices. In the Onygenales group within Pezizomycotina

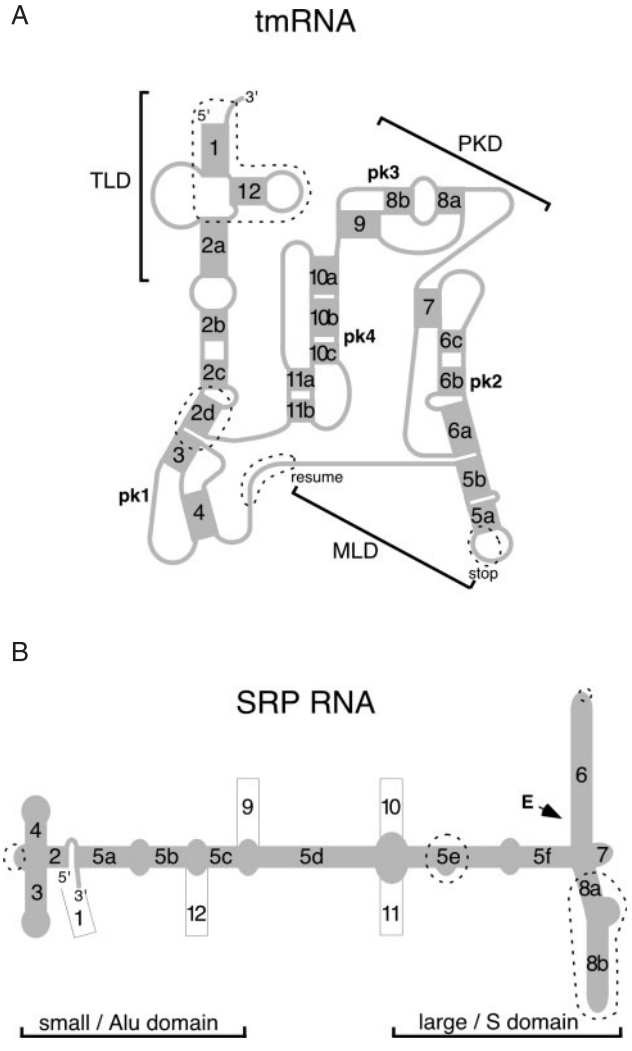

Figure 1. Schematic representation of the secondary structures of (**A**) *E.coli* tmRNA and (**B**) SRP RNA. The tRNA-like (TLD), mRNA-like (MLD) and pseudoknot (PKD) domains are indicated. Helices and their sections are numbered from 1 to 12 and letters a–d. The four pseudoknots are labeled pk1–pk4. The tag peptide-encoding region is located between the resume and stop codon as indicated. Conserved regions are indicated by dashed lines. In the SRP RNA secondary structure diagram B, the features of the mammalian SRP RNA are shown in gray. Helices are numbered from 1 to 12 with helical sections labeled with letters a–f. The approximate boundaries of the small (Alu) and the large (S) domain are shown. The recently discovered extra helix (E) in the SRP RNAs of some Pezizomycotina (see Table 2) is indicated by the arrowhead. Conserved regions are indicated by dashed lines. The conservations are highlighted in the html- and png-formatted alignment files available at the tmRDB and SRPDB, respectively.

(*Histoplasma* and four other species), we found a new helix ('extra' helix E in Figure 1 and Table 2) located toward the 5′ end of helix 6. The phylogenetic distribution of all helices is indicated in Table 2. Representative SRP RNA secondary structure diagrams are shown in Supplementary Data 2.

Most bacteria, including certain chloroplasts, contained a small SRP RNA of 60–115 nt consisting solely of helix 8. The conserved apical tetraloop of this helix typically had the consensus sequence GNRA, with rare G to U mutations in the first position, but occasionally an URRC (8). In some gram-positive bacteria (Bacillales and Clostridia groups) and the deeply-branching gram-negative bacteria *Thermotoga maritima*, the SRP RNA was of the archaeal type but lacked helix 6. Several of these SRP RNAs, as well as some archaeal SRP RNAs, had a non-consensus UGUNR motif (UAUNR, UAUN or CNNNR). In certain Chrenarcheota (*Aeropyrum pernix*) this part seemed to be extended, perhaps forming a helix. The apical loop of the highly conserved helix 8 consisted of 4 nt in most organisms. Plants and certain fungi, however, possessed six nucleotides in this loop. Recently, we found that *Trichomonas*, *Phytophthora*, and *Entamoeba* have a pentaloop with the consensus sequence G[AT][AT]AA.

The eukaryal SRP RNA was highly variable, particularly with respect to the small (Alu) domain (see Table 2 and Supplementary Data 2). Secondary structure models were presented for the *Saccharomyces* SRP RNAs (31,32). These models showed that helices 3 and 4 were missing, whereas helices 9–12 had been acquired. The SRP RNA secondary structures of the non-Ascomycota fungi *Phakopsora* and *Rhizopus* differed from the Ascomycota and were similar to the metazoan SRP RNAs. In Diplomonads and Microsporidia, the small domain seemed to have disappeared to leave an SRP RNA composed only of the large (S) domain.

SRP proteins

SRP9, SRP14 and *SRP21*. A total of 24 SRP9 protein sequences were identified: 16 sequences from the Metazoa, one each from *Dictyostelium discoideum* and *Entamoeba histolytica*, three plant and three from the Alveolata group. SRP14 (a total of 33 sequences) was found in all of the Eukarya examined, including the Fungi. Both SRP9 and SRP14 were absent in Bacteria, Archaea and some eukaryal groups. SRP21 sequences were identified in 12 fungal genomes. Evidence was provided that the metazoan SRP9 is homologous to the fungal SRP21 (31). This finding was consistent with the finding that a gradual evolutionary change from SRP9 to SRP21 had occurred with Pezizomycotina and *Schizosaccharomyces pombe* representing intermediates. However, further studies are required to clarify the functional role of SRP21 in fungi.

SRP19. Protein SRP19 was found in all the examined Eukarya and Archaea. The presence of SRP19 correlated strongly with the appearance of SRP RNA helix 6, thus confirming the important role of SRP19 in the assembly of the large (S) domain (33).

SRP54, also referred to in Bacteria as Ffh (fifty-four homologue), contains the signal sequence binding pocket (34) and thus is likely to be an essential component of every SRP. The SRPDB lists 115 sequences from all phylogenetic groups. We identified homologues to the chloroplast Ffh, cpSRP54, in *Arabidopsis*, *Pisum*, *Chlamydomonas* and *Cyanidioschyzon merolae*.

SRP68 and SRP72. A total of 31 SRP68 and 34 SRP72 sequences from the Fungi, Metazoa, Mycetozoa, Plants, Alveolata and Euglenozoa groups were found. Homologues of these proteins were not identified in the Bacteria and Archaea. Both the proteins are known to form a heterodimer within the large domain of the mammalian SRP, but relatively little is known about their structure. The SRP72 alignment revealed a new lysine-rich domain, originally identified as Pfam B 7529, which will be added to Pfam (35). A corresponding peptide of 63 amino acids located near the C-terminus

Table 2. SRP RNA features and SRP components ordered by phylogeny

2D	Group Species	1	2	3	4	5	6	7	8	9	10	11	12	E	T	9	21	14	19	54	68	72	cp54	cp43
	Plastids *Cyanidioschyzon merolae*	–	–	–	–	–	–	–	x	–	–	–	–	–	–	–	–	–	–	–	–	–	x	x
	Plastids *Arabidopsis thaliana*	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	x	x
a	Bacteria *Escherichia coli*	–	–	–	–	–	–	–	x	–	–	–	–	–	–	–	–	–	–	x	–	–	–	–
b	Bacteria *B.subtilis*	x	x	x	x	x	–	–	x	–	–	–	–	–	–	–	–	–	–	x	–	–	–	–
	Archaea *Aeropyrum pernix*	x	x	X	X	x	x	–	x	–	–	–	–	–	–	–	–	–	x	x	–	–	–	–
c	Archaea *Methanococcus jannaschii*	x	x	x	x	x	x	–	x	–	–	–	–	–	–	–	–	–	x	x	–	–	–	–
d	Ascomycota *Saccharomyces cerevisiae*	–	x	–	–	x	x	X	x	x	X	x	x	–	–	–	x	x	x	x	x	x	–	–
e	Ascomycota *Eremothecium gossypii*	–	x	–	–	x	x	X	x	x	–	x	–	–	–	–	x	x	x	x	x	x	–	–
f	Ascomycota *Coccidioides immitis*	–	x	–	–	x	x	x	x	–	x	–	–	x	–	–	x	x	x	x	x	x	–	–
g	Ascomycota *Schizosaccharomyces pombe*	–	x	–	–	x	x	x	x	–	–	–	–	–	–	–	x	x	x	x	x	x	–	–
	Basidiomycota *Phakospora pachyrhizi*	–	x	x	x	x	x	x	x	–	–	–	–	–	–	x	–	x	x	x	x	x	–	–
	Microsporidia *Encephalitozoon cuniculi*	–	?	?	?	x	x	x	x	–	–	–	–	–	–	?	–	?	x	x	?	?	–	–
h	Metazoa *Homo sapiens*	–	x	x	x	x	x	x	x	–	–	–	–	–	–	x	–	x	x	x	x	x	–	–
	Mycetozoa *Dictyostelium discoideum*	–	x	x	x	x	x	x	x	–	–	–	–	–	–	x	–	x	x	x	x	x	–	–
	Entamoebidae *Entamoeba histolytica*	–	x	x	x	x	x	x	x	–	–	–	–	–	–	x	–	x	x	x	x	x	–	–
	Viridiplantae *Arabidopsis thaliana*	–	x	x	x	x	x	x	x	–	–	–	–	–	–	x	–	x	x	x	x	x	–	–
	Rhodophyta *C.merolae*	–	?	?	?	!	!	!	!	–	–	–	–	–	–	?	–	?	x	x	x	x	–	–
	Heterokonta *Phytophthora sojae*	–	x	x	x	x	x	x	x	–	–	–	–	–	–	?	–	?	x	x	?	?	–	–
	Ciliophora *Tetrahymena thermophila*	–	x	x	s	x	x	x	x	–	–	–	–	–	–	x	–	x	x	x	x	x	–	–
i	Apicomplexa *Plasmodium falciparum*	–	x	X	X	x	x	x	x	–	–	–	–	–	–	x	–	x	x	x	x	x	–	–
j	Apicomplexa *Theileria annulata*	–	x	x	s	x	x	x	x	–	–	–	–	–	–	x	–	x	x	x	x	x	–	–
	Euglenozoa *Trypanosoma brucei*	–	x	x	x	x	x	x	x	–	–	–	–	–	x	–	–	–	x	x	x	x	–	–
	Parabasala *Trichomonas vaginalis*	–	x	x	x	x	x	x	x	–	–	–	–	–	–	!	–	x	x	x	?	?	–	–
	Diplomonadida *Giardia lamblia*	–	?	?	?	x	x	x	x	–	–	–	–	–	–	?	–	?	x	x	x	?	–	–

The names of representative species are given for each group. The column labeled '2D' indicates the secondary structures *a–j* shown in Supplementary Data 2. The RNA features (helices 1–12, the 'extra' helix E and a tRNA-like molecule (40) labeled 'T' are shown in the center section of the table; proteins SRP9–SRP72, as well as the chloroplast proteins cp54 and cp43 are indicated in the right portion. The table cells are annotated as '–', absent; '?', maybe absent or was not found; '!', expected to be present; 'x', present; 'X', this feature was pronounced and may contain several helical sections; 's', this helix was comparatively small or possibly absent.

of human SRP72 with the consensus PDPXRWLPXXER was shown to bind to SRP RNA with high affinity (36).

cpSRP43. It is a unique nuclear encoded protein and part of the post-translational SRP found only in chloroplasts. The protein binds to polypeptides destined for the thylakoid membrane. cpSRP43 contains four ankyrin repeats at the N-terminus and two chromodomains at the C-terminus. It forms a complex with cpSRP54 via its chromodomains (37).

SRP-associated proteins

SRP Receptor (alpha) (FtsY). The SRP receptor is a single polypeptide (FtsY) in the Bacteria and Archaea. In Eukaryotes, the SRP receptor is composed of two subunits, alpha and beta. The alpha subunit is related to FtsY and to SRP54 (Ffh) due to their GTPase domain similarity. Unique to the alpha subunit of the SRP receptor (FtsY) is an N-terminal A-region which is thought to be responsible for interacting with the membrane or the beta subunit [reviewed in (2)].

SRP Receptor (beta) was found in all Eukaryotes including the Fungi. The protein is characterized by a transmembrane anchor and binds to the alpha subunit of the receptor. Like SRP54 (Ffh), the beta subunit also contains a GTPase domain.

FlhF. This protein was characterized first as a flagellar gene from *B.subtilis*. It belongs to the same family of GTP-binding proteins as Ffh and FtsY (38) suggesting a role in SRP function. However, FlhF was shown recently to be dispensable for protein secretion (39).

Phylogeny of SRP

An extensive inventory of SRP RNA and protein components has allowed us to arrive at a comprehensive view of SRP phylogeny (Table 2). Essential elements include (i) the development of an altered Alu domain in the Ascomycota lacking helices 3 and 4, accompanied by the appearance of protein SRP21, (ii) the emergence of the more complex *Saccharomyces* SRP RNAs with multiple insertions, (iii) the retention of a metazoan-type SRP in the Basidiomycota, (iv) the appearance of eukaryotic SRPs that lack the typical mammalian SRP proteins or the small (Alu) domain, (v) the presence of a much reduced SRP in bacteria and chloroplasts composed of only one protein (Ffh) and a small RNA that seems to be absent in the chloroplasts of higher plants and (vi) the conservation of the composition and secondary structure of the archaeal SRP.

Outlook

Exploring RNA and protein alignments has become increasingly difficult with the growing number of sequences. We have implemented and continue to develop a browser which allows to display alignments at various zoom levels like a map. The user can explore and see more clearly the species- and group-specific differences. Further improvements in the quality of the alignments can be expected. Overall, these advances will lead to a better understanding not only of *trans*-translation and co-translational protein translocation, but also of the functional potential of the ribosome.

Access

The data are freely accessible for research purposes at the internet addresses http://psyche.uthct.edu/dbs/tmRDB/tmRDB.html and http://psyche.uthct.edu/dbs/SRPDB/SRPDB.html or at the corresponding mirror sites provided

in the Abstract. This article should be cited in research projects which use the tmRDB and SRPDB resources.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Jorgen Kjems for RNA expertise and support, Allan Lind-Thomsen for assistance with SARSE and Florian Müller for the ERNA-3D modeling program. E.S.A. is supported by the Interdisciplinary Nanoscience Center (iNANO) of the University of Aarhus. M.A.R. is supported by the SWEGENE consortium. J.G. is supported by the Danish Research Council for Technology and Production Sciences and the Danish Center for Scientific Computing. This work was also supported by NIH grants GM-58267 to J.W. and GM-49034 to C.Z. Funding to pay the Open Access publication charges for this article was provided by NIH grant GM-49034 to C.Z.

Conflict of interest statement. None declared.

REFERENCES

1. Karzai,A.W., Roche,E.D. and Sauer,R.T. (2000) The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nature Struct. Biol.*, **7**, 449–455.
2. Halic,M. and Beckmann,R. (2005) The signal recognition particle and its interactions during protein targeting. *Curr. Opin. Struct. Biol.*, **15**, 116–125.
3. Ban,N., Nissen,P., Hansen,J., Moore,P.B. and Steitz,T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
4. Yusupov,M.M., Yusupova,G.Z., Baucom,A., Lieberman,K., Earnest,T.N., Cate,J.H. and Noller,H.F. (2001) Crystal structure of the ribosome at 5.5 Å resolution. *Science*, **292**, 883–896.
5. Ramakrishnan,V. (2002) Ribosome structure and the mechanism of translation. *Cell*, **108**, 557–572.
6. Gueneau de Novoa,P. and Williams,K.P. (2004) The tmRNA website: reductive evolution of tmRNA in plastids and other endosymbionts. *Nucleic Acids Res.*, **32**, D104–D108.
7. Zwieb,C., Gorodkin,J., Knudsen,B., Burks,J. and Wower,J. (2003) tmRDB (tmRNA database). *Nucleic Acids Res.*, **31**, 446–447.
8. Regalia,M., Rosenblad,M.A. and Samuelsson,T. (2002) Prediction of signal recognition particle RNA genes. *Nucleic Acids Res.*, **30**, 3368–3377.
9. McGinnis,S. and Madden,T.L. (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.*, **32**, W20–W25.
10. Eddy,S.R. (2002) A memory-efficient dynamic programming algorithm for optimal alignment of a sequence to an RNA secondary structure. *BMC Bioinformatics*, **3**, 18.
11. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
12. Wheeler,D.L., Chappey,C., Lash,A.E., Leipe,D.D., Madden,T.L., Schuler,G.D., Tatusova,T.A. and Rapp,B.A. (2000) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **28**, 10–14.
13. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J., Rapp,B.A. and Wheeler,D.L. (2000) GenBank. *Nucleic Acids Res.*, **28**, 15–18.
14. Higgins,D.G., Thompson,J.D. and Gibson,T.J. (1996) Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.*, **266**, 383–402.
15. Hall,T. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.*, **41**, 95–98.
16. Larsen,N. and Zwieb,C. (1991) SRP-RNA sequence alignment and secondary structure. *Nucleic Acids Res.*, **19**, 209–215.
17. Gorodkin,J., Zwieb,C. and Knudsen,B. (2001) Semi-automated update and cleanup of structural RNA alignment databases. *Bioinformatics*, **17**, 642–645.
18. Knudsen,B. and Hein,J. (2003) Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Res.*, **31**, 3423–3428.
19. Rosenblad,M.A., Gorodkin,J., Knudsen,B., Zwieb,C. and Samuelsson,T. (2003) SRPDB: Signal Recognition Particle Database. *Nucleic Acids Res.*, **31**, 363–364.
20. Clamp,M., Cuff,J., Searle,S.M. and Barton,G.J. (2004) The Jalview Java alignment editor. *Bioinformatics*, **20**, 426–427.
21. Edgar,R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.
22. Sharkady,S.M. and Williams,K.P. (2004) A third lineage with two-piece tmRNA. *Nucleic Acids Res.*, **32**, 4531–4538.
23. Karzai,A.W., Susskind,M.M. and Sauer,R.T. (1999) SmpB, a unique RNA-binding protein essential for the peptide-tagging activity of SsrA (tmRNA). *EMBO J.*, **18**, 3793–3799.
24. Barends,S., Karzai,A.W., Sauer,R.T., Wower,J. and Kraal,B. (2001) Simultaneous and functional binding of SmpB and EF-Tu-TP to the alanyl acceptor arm of tmRNA. *J. Mol. Biol.*, **314**, 9–21.
25. Jacob,Y., Sharkady,S.M., Bhardwaj,K., Sanda,A. and Williams,K.P. (2005) Function of the SmpB tail in transfer-messenger RNA translation revealed by a nucleus-encoded form. *J. Biol. Chem.*, **280**, 5503–5509.
26. Bycroft,M., Hubbard,T.J., Proctor,M., Freund,S.M. and Murzin,A.G. (1997) The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. *Cell*, **88**, 235–242.
27. Barends,S., Bjork,K., Gultyaev,A.P., de Smit,M.H., Pleij,C.W. and Kraal,B. (2002) Functional evidence for D- and T-loop interactions in tmRNA. *FEBS Lett.*, **514**, 78–83.
28. Burks,J., Zwieb,C., Muller,F., Wower,I. and Wower,J. (2005) Comparative 3-D Modeling of tmRNA. *BMC Mol. Biol.*, **6**, 14.
29. Rosenblad,M.A. and Samuelsson,T. (2004) Identification of chloroplast signal recognition particle RNA genes. *Plant Cell Physiol.*, **45**, 1633–1639.
30. Zwieb,C., Van Nues,R.W., Rosenblad,M.A., Brown,J.D. and Samuelsson,T. (2005) A nomenclature for all signal recognition particle RNAs. *RNA*, **11**, 7–13.
31. Rosenblad,M.A., Zwieb,C. and Samuelsson,T. (2004) Identification and comparative analysis of components from the signal recognition particle in protozoa and fungi. *BMC Genomics*, **5**, 5.
32. Van Nues,R.W. and Brown,J.D. (2004) *Saccharomyces* SRP RNA secondary structures: A conserved S-domain and extended Alu-domain. *RNA*, **10**, 75–89.
33. Walter,P. and Blobel,G. (1983) Disassembly and reconstitution of signal recognition particle. *Cell*, **34**, 525–533.
34. Keenan,R.J., Freymann,D.M., Walter,P. and Stroud,R.M. (1998) Crystal structure of the signal sequence binding subunit of the signal recognition particle. *Cell*, **94**, 181–191.
35. Bateman,A., Coin,L., Durbin,R., Finn,R.D., Hollich,V., Griffiths-Jones,S., Khanna,A., Marshall,M., Moxon,S., Sonnhammer,E.L. *et al.* (2004) The Pfam protein families database. *Nucleic Acids Res.*, **32**, D138–41.
36. Iakhiaeva,E., Yin,J. and Zwieb,C. (2005) Identification of an RNA-binding domain in Human SRP72. *J. Mol. Biol.*, **345**, 659–666.
37. Schuenemann,D., Gupta,S., Persello-Cartieaux,F., Klimyuk,V.V., Jones,J.D.G., Nussaume,L. and Hoffman,N.E. (1998) A novel signal recognition particle targets light-harvesting proteins to the thylakoid membranes. *Proc. Natl Acad. Sci. USA*, **95**, 10312–10316.
38. Carpenter,P.B., Hanlon,D.W. and Ordal,G.W. (1992) flhF, a *Bacillus subtilis* flagellar gene that encodes a putative GTP-binding protein. *Mol. Microbiol.*, **6**, 2705–2713.
39. Zanen,G., Antelmann,H., Westers,H., Hecker,M., van Dijl,J.M. and Quax,W.J. (2004) FlhF, the third signal recognition particle-GTPase of *Bacillus subtilis*, is dispensable for protein secretion. *J. Bacteriol.*, **186**, 5956–5960.
40. Liu,L., Ben-Shlomo,H., Xu,Y.X., Stern,M.Z., Goncharov,I., Zhang,Y. and Michaeli,S. (2003) The trypanosomatid signal recognition particle consists of two RNA molecules, a 7SL RNA homologue and a novel tRNA-like molecule. *J. Biol. Chem.*, **278**, 18271–18280.