Nucleic Acids Research, 1994, Vol. 22, No. 17 3483–3484

The signal recognition particle database (SRPDB)

Christian Zwieb* and Niels Larsen[1]+
Department of Molecular Biology, The University of Texas Health Science Center, PO Box 2003, Tyler, TX 75710 and [1]Department of Microbiology, 131 Burrill Hall, 407 South Goodwin Avenue, University of Illinois, Urbana, IL 61801, USA

ABSTRACT

The SRPDB (signal recognition particle database) provides aligned SRP RNA and protein sequences, annotated and phylogenetically ordered. This release includes 82 SRP RNAs (including 22 bacterial and 9 archaeal homologs) and a total of 20 protein sequences representing SRP9, SRP14, SRP19, SRP54, SRP68, and SRP72. The offerings also include representative RNA secondary structure diagrams.

DESCRIPTION

The signal recognition particle data base (SRPDB; currently at The University of Texas Health Science Center at Tyler), offers phylogenetically arranged alignments of SRP RNAs and proteins, the corresponding annotated sequences and representative secondary structure diagrams.

This release includes 82 SRP RNA sequences (including archaeal and bacterial homologs), which is an increment of 18 sequences since the previous announcement (1). The added sequences comprise 14 variants from *Lycopersicon esculentum* (tomato), one from *Oryza sativa* (rice), as well as three tentative SRP RNAs, one from *Homo sapiens* (man) and two from *Caenorhabditis elegans* (a roundworm). The new SRP RNA sequences were aligned using the previously described rules (2).

The known SRP proteins, a total of 20, are listed in Table 1. Included are alignments of the sequences of proteins SRP19 (four representatives), SRP54 (nine representatives), and SRP9 (two representatives). Our protein alignments highlight conserved amino acids and regions found experimentally to be required for binding to SRP RNA.

In addition to alignments and sequences, we offer representative SRP RNA secondary structure models, one for each of the major phylogenetic domains (we hope to add more soon). These models are minimal models, i.e. there is comparative evidence for every base pair included in the model (2); this is in contrast to pairing schemes adopted from free energy predictions or from mere sequence complementarity, both of which give no evidence for the pairings' real existence. For ease of reference, each helix shown in the models is given an arabic numeral (according to conventions used for e.g. ribosomal RNAs, see ref. 3). The new sequences provide no evidence for additional base pairings, but all sequences conform to and in some cases support, the previously established models (2).

New SRP sequences are identified primarily by their appearance in the primary sequence databases Genbank (4) and EMBL (5). Using a modified version of the program BLAST (6) we first searched a local copy of Genbank (release 83) for sequence similarity with each of the sequences in the previous SRPDB release. To capture less similar SRP RNA sequences, we then searched with a motif describing the conserved primary and secondary structure features of helix 8, using the program RNABOB (7). This search revealed two previously unidentified SRP RNAs in chromosome IV of *Caenorhabditis elegans* (entry CELZC155, approximate positions 12846–13149; entry CEB0284, approximate positions 13959–14249 on the complementary strand).

SRP RNA sequences can be obtained in formats used by the GCG sequence analysis package (8), GenBank (4), EMBL (5), and in a format recognized by the secondary structure drawing program LoopDLoop (9). SRP protein sequences are provided in Swissprot format (10).

Alignments are available as concatenated GenBank and EMBL entries (with inserted gaps), in the ALMA alignment editor format (11), and as files used by the AE2 alignment editor (12). We offer the alignment in a printable PostScript version, which (for the RNAs) highlights base pairs and helices numbered using the convention described above. The secondary structure figures are available in PostScript format.

Electronic copies of the SRPDB are freely available: Connect with ftp to diana.uthct.edu (currently 192.88.11.4), login with name anonymous and give your full electronic mail address as password (thereby we can mail you about updates). A help file named 00README explains where to find the data. Hardcopies are also available through written contact or E-mail, but electronic transfer is preferred.

Submission of SRP related data will be accepted in any form. The submitter may request that the data may not be released after a given date or upon notification. Submitted sequences will be aligned and the alignment will be returned in the requested format.

ACKNOWLEDGEMENTS

We thank the Free Software Foundation for excellent software. This work was assisted by a Grant-in-Aid award to C.Z. from the American Heart Association, Texas Affiliate, identification number 91G-556. We thank Michael McCaughey for assistance with the searches of Genbank.

*To whom correspondence should be addressed (E-mail: zwieb@jason.uthct.edu)

+E-mail: niels@darwin.life.uiuc.edu

REFERENCES

1. Larsen, N. and Zwieb, C. (1993) Nucleic Acids Res., **21**, 3019–3020
2. Larsen, N. and Zwieb, C. (1991) Nucleic Acids Res., **19**, 209–215
3. Leffers, H., Kjems, J., Østergaard, L., Larsen, N. and Garrett, R.A. (1987) J. Mol. Biol., **195**, 43–61
4. Burks,C., Cinkosy, M.J., Fischer, W.M., Gilna, P., Hayden, J.E.-D., Keen, G.M., Kelley, T.A., Kelley, M., Kristofferson, D. and Lawrence, J. (1992) Nucleic Acids Res., **20**, 2071–2074
5. Higgins, D.G., Fuchs, R. and Stoehr, P.J. (1992) Nucleic Acids Res., **20**, 2071–2074
6. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol., **215**, 403–410.
7. Eddy, S.R. unpublished. The RNABOB program is available by anonymous ftp from cele.mrc-lmb.cam.ac.uk.
8. Devereux, J., Haeberli P. and Smithies, O. (1984). Nucleic Acids Res., **12**, 387–395
9. Gilbert, D.G., The LoopDLoop program is obtained by anonymous ftp from ftp.bio.indiana.edu
10. Bairoch, B. and Boeckmann B. (1993), Nucleic Acids Res., **21**, 3093–3096
11. Thirup, S. and Larsen, N. (1990) Proteins: Structure, Function and Genetics, **7**, 291–295
12. Larsen, N., Olsen, G.J., Maidak, B.L., McCaughey, M.J., Overbeck, R., Macke, T.J., Marsh, T.L. and Woese, C.R. (1993), Nucleic Acids Res., **21**, 3021–3023