Nucleic Acids Research, 1998, Vol. 26, No. 1 177–178

The Signal Recognition Particle Database (SRPDB)

Niels Larsen, Tore Samuelsson[1] and Christian Zwieb[2,*]

Department of Microbiology, Giltner Hall, Michigan State University, East Lansing, MI 48824, USA, [1]Department of Medical Biochemistry, University of Göteborg, Medicinareg. 9A, S-413 90, Göteborg, Sweden and [2]Department of Molecular Biology, The University of Texas Health Science Center, PO Box 2003, Tyler, TX 75710, USA

Received October 2, 1997; Accepted October 3, 1997

ABSTRACT

This release of the SRPDB (signal recognition particle database, http://pegasus.uthct.edu/SRPDB/SRPDB.html) adds four SRP RNA sequences (a total of 99 SRP RNA sequences), 23 SRP protein sequences (a total of 63 protein sequences from SRP9, SRP14, SRP19, SRP21, SRP54, SRP68 or SRP72), and, for the first time, sequences of the alpha subunit of the eukaryotic SRP receptor and its homologous bacterial proteins (a total of 21 sequences). Sequences are offered phylogenetically ordered, annotated with links to the primary databases, and in aligned form. Also downloadable are sample SRP RNA secondary structure diagrams, three-dimensional models of representative SRP RNAs, and search motifs.

DESCRIPTION

The signal recognition particle database (SRPDB) at the University of Texas Health Center at Tyler, Texas, provides 99 annotated sequences of SRP RNAs, 63 SRP protein sequences, and 21 SRP receptor protein sequences. The sequences were grouped and ordered according to the phylogeny derived by the Ribosomal Database Project (1) and alignments were created using the rules described previously (2).

SRP RNAs

Using PatScan (3), we searched release 102 of GenBank (4) with a series of motifs that each describe the most conserved primary and secondary structure features. These search motifs are available at the SRPDB and allow anyone to change the pattern definitions and to repeat the search.

Four new SRP RNAs were identified, one from *Brevibacillus brevis*, one from *Synechocystis* sp. (previously not annotated in the *Synechocystis* genome sequence), one from *Fugu rubripes* (the first SRP RNA from a fish), and a second SRP RNA from the archaea *Sulfolobus solfataricus*. The SRP RNA alignment is available as concatenated GenBank (4) and EMBL (5) entries with gaps inserted in the sequences, as a human readable text format, and as a printable PostScript version where helices are numbered and highlighted. The alignment can also be viewed directly at the SRPDB web site. The corresponding sequences are available as separate GenBank and EMBL entries. Representative SRP RNA secondary structure models (in PostScript format) are from *Bacillus subtilis* (Bacteria), *Halobacterium halobium* (Archaea), and *Canis sp.* (Eukaryota). Tentative three-dimensional models in PDB format are offered for human SRP RNA (6), and the SRP RNAs from *Methanococcus jannaschii*, *B.subtilis*, *Escherichia coli* and *Mycoplasma mycoides.*

SRP proteins

SRP protein sequences were identified in keyword searches of the primary databases at GenBank (4) or EBI (5), by using BLAST at NCBI (7; http:///www3.ncbi.nlm.nih.gov/BLAST) or the FASTA program (8) with the known SRP protein sequences as inputs. New sequences include *Zea mays* SRP9 (a total of five SRP9 sequences), and SRP14 sequences of *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Oryza sativa* and human (now a full-length sequence). Sixteen new protein SRP54 sequences for a total of 35 were added from these species: *Acidianus ambivalens*, *Aspergillus niger*, *Entamoeba histolytica*, *Haemophilus influenzae*, *Helicobacter pylori*, *Lycopersicon esculentum*, *Mycobacterium leprae*, *Mycobacterium tuberculosis*, *Mycoplasma pneumoniae*, *Rattus norvegicus* (partial sequence), *Streptococcus mutans*, *Sulfolobus acidocaldarius*, *Synechococcus* sp., *Synechocystis* sp., *Thermus aquaticus* and *Yarrowia lipolytica*. Sequences for the SRP68 and SRP72 proteins were found in the DNA of *C.elegans*. A region identical to amino acid residue positions 315–548 of human SRP72 was identified in an truncated isoform of human Ca^{2+}/calmodulin-dependent protein kinase II, designated γ_{SRP} (9). Since the human SRP72 and the CaM kinase II γ genes are located on different chromosomes (chromosome 18 and 10q22, respectively) this cDNA may be the result of *trans*-splicing (9).

SRP receptor proteins

The mammalian SRP receptor is composed of two subunits, SRα and SRβ. SRα and SRP54 form a separate class of GTP binding proteins as their 'N+G' domains are structurally related (for a review see ref. 10). Proteins with homology to SRα have been identified in numerous bacteria, and in *E.coli* this protein is referred to as FtsY (11,12). As the SRP54 and SRα/FtsY proteins are not only highly homologous but also functionally related, they are now included in the SRPDB. SRα/FtsY proteins, a total of 21, are presented from the following species: *B.subtilis*, *Canis*

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

familiaris, *Desulfurolobus ambivalens*, *E.coli*, *H.influenzae*, *H.pylori*, *Homo sapiens*, *M.jannaschii* (not annotated in the genome sequence), *Mycobacterium leprae*, *M.tuberculosis*, *Mycoplasma genitalium*, *Mycoplasma hominis*, *Mycoplasma mycoides*, *M.pneumoniae*, *Neisseria gonorrhoeae*, *Rickettsia prowazekii*, *Saccharomyces cerevisiae*, *S.acidocaldarius*, *S.solfataricus*, *Synechocystis sp.* PCC6803 and *Thermococcus sp.* AN1. The sequences were identified with BLAST as described above and by using the sequence analysis tools of the Genetics Computer Group (13). Multiple sequence alignments, performed with PILEUP (13) and CLUSTALW (14), highlight the homologous N+G domains between all SRα/FtsY proteins and SRP54. Characteristic of the SRα/FtsY proteins is a variable N-terminal, highly charged extension. Interestingly however, in *H.pylori* and *R.prowazekii*, this N-terminal extension is completely absent**.**

ACCESS

All data and software are freely accessible at the SRPDB web site and are downloadable by connecting to the URL http://pegasus.uthct.edu/SRPDB/SRPDB.html . The SRPDB can also be accessed directly by anonymous ftp to 'diana.uthct.edu'. (Currently 192.88.11.4; login with the name 'anonymous', without the quotes, and give your full electronic mail address as the password.) The complete SRPDB can be downloaded from the same site as a single tar-file. Hardcopies of the alignments are available through written contact or through Email to the authors.

Submission of SRP related data will be accepted in any form. We will align the sequences and return the alignment to the submitter in the requested format. The submitter may request that the data not be released until after a given date or upon notification. Please cite this article if your research is assisted by the SRPDB. N.L. can be contacted by Email at niels@truth.mph.msu.edu, T.S. at tore.samuelsson@medkem.gu.se and C.Z. at zwieb@uthct.edu.

ACKNOWLEDGEMENTS

N.L. was supported by the Center for Microbial Ecology at Michigan State University (DEB #9120006), T.S. by grant B-BU 9676-311 from the Swedish Natural Science Research Council, and C.Z. by NIH grant GM-49034. We also thank the Free Software Foundation and Larry Wall for their excellent software.

REFERENCES

1 Maidak, B.L., Olsen, G.J., Larsen, N., Overbeek, R., McCaughey, M.J. and Woese, C.R. (1997) *Nucleic Acids Res.*, **25**, 109–110.
2 Larsen, N. and Zwieb, C. (1991) *Nucleic Acids Res.*, **19**, 209–215.
3 Overbeek, R. and Price, M. (1993) *Accessing Integrated Genomic Data Using GenoBase: A Tutorial (Part 1)*, Report ANL/MCS-TM-173, Argonne National Laboratory.
4 Benson, D.A., Boguski, M., Lipman, D.J. and Ostell, J. (1997) *Nucleic Acids Res.*, **25**, 1–6. [See also this issue *Nucleic Acids Res.* (1998) **26**, 1–7.]
5 Stoesser, G., Sterk, P., Tuli, M.A., Stoehr, P.J. and Cameron, G.N. (1997) *Nucleic Acids Res.*, **25**, 7–13. [See also this issue *Nucleic Acids Res.* (1998) **26**, 8–15.]
6 Zwieb, C., Müller, F. and Larsen, N. (1996) *Folding Design*, **1**, 315–324.
7 Altschul, S.F., Gish, W., Miller, W., Myers, E.W.and Lipman, D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.
8 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.
9 Breen, M. and Ashcroft, S.J.H. (1997) *FEBS Lett.*, **409**, 375–379.
10 Lütcke, H. (1995) *Eur. J. Biochem.*, **228**, 531–550.
11 Bernstein, H.D., Poritz, M.A., Strub, K., Hoben, P.J., Brenner, S. and Walter, P. (1989) *Nature*, **340**, 482–486.
12 Römisch, K., Webb, J., Herz, J., Prehn, S., Frank, R., Vingron, M., and Dobberstein, B. (1989) *Nature*, **340**, 478–482.
13 *Program Manual for the Wisconsin Package, Version 8,* September 1994, Genetics Computer Group, 575 Science Drive, Madison, WI 53711, USA.
14 Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) *Nucleic Acids Res.*, **22**, 4673–4680.