The Signal Recognition Particle Database (SRPDB)

Tore Samuelsson and Christian Zwieb¹,*

Department of Medical Biochemistry, University of Göteborg, Medicinareg. 9A, S-413 90, Göteborg, Sweden and
¹Department of Molecular Biology, The University of Texas Health Science Center at Tyler, PO Box 2003, Tyler, TX 75710, USA

Received October 1, 1998; Revised October 5, 1998; Accepted October 13, 1998

ABSTRACT

The signal recognition particle database (SRPDB) is located at the University of Texas Health Science Center at Tyler and includes tabulations of SRP RNA, SRP protein and SRP receptor sequences. The sequences are annotated with links to the primary databases. They are ordered alphabetically or phylogenetically and are available in aligned form. As of September, 1998, there were 108 SRP RNA sequences, 83 SRP protein sequences and 28 sequences of the SRP receptor α subunit and its homologues. In addition, the SRPDB provides search motifs consisting of conserved amino acid and nucleotide residues, and a limited number of SRP RNA secondary structure diagrams and 3-D models. The data are available freely at the URL http://psyche.uthct.edu/dbs/SRPDB/SRPDB.html

DESCRIPTION

The signal recognition particle (SRP) is a small cytosolic ribonucleoprotein particle that directs secretory proteins to the proper cellular compartment (reviewed in ref. 1). As of September, 1998, the SRPDB provides 106 annotated SRP RNA sequences, 83 SRP protein sequences of SRP9, SRP14, SRP19, SRP21, SRP68, SRP72 and 28 SRP receptor protein sequences. To determine the conserved features of the SRP components and thus provide clues for their functions, the collected sequences are ordered phylogenetically and aligned using the previously described rules (2,3).

SRP RNAs

We searched a non-redundant collection of the primary databases with the 2.0 version of BLAST (4) using 45 representative SRP RNA sequences from the 1998 release of the SRPDB (5). Also, we used several conserved RNA secondary structure search motifs as input for the pattern matcher PatScan (http://www-c.mcs.anl.gov/home/overbeek/PatScan/HTML/patscan.html). The SRP RNA candidates were either rejected or identified as positives by comparison with the established aligned sequences in the SRPDB. The eight newly identified SRP RNA sequences are from Aquifex aeolicus (not annotated in the primary databases), Borrelia burgdorferi, Caulobacter crescentus, Clostridium perfringens, Haemophilus ducreyi, Leptomonas collosoma, Pyrococcus horikoshii and Treponemana pallidum (not annotated). Also available at the SRPDB web site are several tentative 3-D SRP RNA models in PDB format, generated with ERNA-3D (6), for the human SRP RNA (7) and the SRP RNAs of Methanococcus jannaschii, Bacillus subtilis, Escherichia coli and Mycoplasma mycoides (8).

The SRP RNA alignment was generated manually with the help of the alignment editors ALMA (9) and ALE (2), and is available as concatenated GenBank (10), EMBL (11), or GCG-formatted entries with gaps inserted in the sequences. The alignment can be viewed at the web site in textual form or as a printable PostScript version (if a PostScript viewer is installed) where helices are numbered and highlighted according to the nomenclature by Larsen and Zwieb (3).

SRP proteins

With the known SRPDB protein sequences as queries, we used BLAST (4) to identify SRP proteins in the primary databases. A candidate for the Schizosaccharomyces pombe SRP9 was identified (accession number 3183357, annotated as a hypothetical protein) to bring the number of known SRP9 sequences to six. We found only one partial sequence of Macaca radiata SRP14, but identified an open reading frame in the Saccharomyces cerevisiae genome (accession YDL092w) as being the SRP14 homologue. There were five new SRP19 sequences from Archaeoglobus fulgidus, Candida albicans, Methanobacterium thermautotrophicum, Kluveromyces lactis and Pyrococcus horikoshii, for a total of 11 SRP19 entries. No homologues to yeast SRP21 (accession P32342) were found. To identify candidates for protein SRP54, the characteristic methionine-rich domain (SRP54M) was selected for a query. The number of SRP54 sequences increased by 12 to a total of 46 sequences derived from the following organisms: A. aeolicus (mistakenly annotated as an SRP receptor protein), Arabidopsis thaliana (two sequences), A. fulgidus, B. burgdorferi, C. albicans, C. trachomatis, M. thermautotrophicum, P. horikoshii, Streptomyces coelicolor, Streptomyces lividans and T. pallidum. Two new SRP68 sequences were identified (a total of five); one from S. pombe (annotated as hypothetical protein, accession 3560159), the other from the human SRP68, determined in our laboratory (Gowda,K. and Zwieb.C., unpublished data). The SRP72 equivalents from the same two organisms were identified for a total of six entries. The
protein alignments can be viewed directly or downloaded as concatenated entries in the most popular formats.

**SRP receptor proteins**

The bacterial FtsY proteins are homologous to the α subunit of the SRP receptor (SRα). We identified seven new sequences using BLAST (4) or TFASTA from the following species: *A. aeolicus*, *A. fulgidus*, *B. burgdorferi*, *C. trachomatis*, *M. thermoautotrophicum*, *P. horikoshii*, and *T. pallidum*. Multiple sequence alignments of the SRα/FtsY family were created with CLUSTALW (12) and BOXSHADE (available by ‘anonymous ftp’ to ftp.isrec.isb-sib.ch) revealing that most FtsY proteins contain a charged N-terminal domain which is highly variable in length and sequence, but absent in *B. burgdorferi*, *C. trachomatis*, *Helicobacter pylori*, *Rickettsia prowazekii*, and *T. pallidum*. As SRα/FtsY and the N and G-domains of SRP54 structures are closely related (13,14), the SRPDB also presents alignments which highlight this relationship.

**ACCESS**

The data in the SRPDB are freely available at the URL http://psyche.uthct.edu/dbs/SRPDB/SRPDB.html or at its European mirror site at http://www.medkem.gu.se/dbs/SRPDB/SRPDB.html. Please note that the previously available ‘anonymous ftp’ access has been discontinued. Direct data submission to the SRPDB will be accepted. The submitters may request that the data not be released until after a given date or upon notification. We will make an effort to align the submitted sequences and return the alignment in the requested format. Hard copies of the alignments are available by Email request to the last author or through written contact. The first author can be contacted by Email at tore.samuelsson@medkem.gu.se and C.Z. at zwieb@uthct.edu. Please cite this article if the SRPDB assists your research.

**ACKNOWLEDGEMENTS**

We thank Dr Niels Larsen for help with ALE, and Dr Cathy Wu, University of Texas Health Science Center at Tyler and Jing Zhao for help with the PatScan searches. This work is supported by NIH grant GM-49034 to CZ.

**REFERENCES**