158–160 Nucleic Acids Research, 1999, Vol. 27, No. 1 © 1999 Oxford University Press

5S Ribosomal RNA data bank

Maciej Szymanski, Miroslawa Z. Barciszewska, Jan Barciszewski and Volker A. Erdmann[1,*]

Institute of Bioorganic Chemistry of the Polish Academy of Sciences, Noskowskiego 12, 61704 Poznan, Poland and [1]Institut für Biochemie, Freie Universitat Berlin, Thielallee 63, 14195 Berlin, Germany

Received October 13, 1998; Accepted October 15, 1998

ABSTRACT

This paper presents the updated version of the data base of ribosomal 5S ribonucleic acids (5S rRNA) and their genes (5S rDNA). This edition of the data bank contains 1889 primary structures of 5S rRNA and 5S rDNA. These include 60 archaebacterial, 439 eubacterial, 63 plastid, 9 mitochondrial and 1318 eukaryotic sequences. The nucleotide sequences of 5S rRNAs or 5S rDNAs are divided according to the taxonomic position of organisms. The sequences stored in the database can be viewed and retrieved using the taxonomic browser at the URL: http://rose.man.poznan.pl/5SData/5SRNA.html

INTRODUCTION

In all organisms, messenger-directed protein synthesis is catalyzed by ribosomes. Although they are universal cellular organelles and catalyze the formation of peptide bonds, the entire protein biosynthesis process is highly complicated (1). Bacterial ribosomes sediment at about 70S and consist of two subunits, a small one—30S, and a large one—50S. The 50S subunit containing 33 different proteins, 23S rRNA of 2900 nucleotides and 5S rRNA has been recently crystallized and its structure has been solved with 9 Å resolution (2). Ribosomal 5S RNA is a 120 nucleotide long RNA molecule found in virtually all ribosomes with the exception of mitochondria of some fungi, higher animals and most protists (3,4). It is located in the central protuberance of the large ribosomal subunit near the peptidyl transferase and factor-binding sites (1,2). Since its discovery in 1963 as a component of *Escherichia coli* ribosome (5), 5S rRNA has been an object of very intensive study by many different methods (6,7). During the last 35 years, a large amount of sequence data has also been collected for this RNA species (8), but we are still far away from a detailed knowledge of the tertiary structure and detailed function of 5S rRNA, although in 1997 new, very important, structural data was published (9,10). The crystal structures of a 62 nt domain of *E.coli* 5S ribosomal RNA and the duplex dodecamer encompassing an internal loop E have been determined at 3.0 and 1.5 Å resolution, respectively (9). In addition, the solution structure of a 42 nt derivative of *E.coli* 5S rRNA which includes loops D and E has been determined by nuclear magnetic resonance spectroscopy (10). It was demonstrated that this portion of the 5S rRNA is an overall double helical region with several irregularities which might be important for specific RNA–protein interactions. These structures fit well with the NMR structures. Recently, great progress has been observed in crystallization of short fragments of 5S rRNA. The X-ray structure of single domain E and of helix E octamer and heptamer of *Thermus flavus* 5S rRNA have been solved at atomic resolution (11; M.Vallazza, C.Forster, Eickmann, C.Lipmann, M.Perbandt, C.Betzel and V.A.Erdmann, manuscript submitted).

In prokaryotes and organelles, 5S rRNAs are synthesized as a part of a single long transcript, together with 16S and 23S rRNAs. Eukaryotic 5S rRNAs of cytoplasmic ribosomes are usually encoded by separate genes arranged in tandem arrays of repeating units. Their number varies significantly up to several thousands in vertebrates and plants. The 5S rRNA genes are transcribed by polymerase III which is strongly inhibited by p53 (12) and depends strongly in eukaryotic cells on the binding of a 40 kDa protein—transcription factor IIIA (TF IIIA)—to the internal control region of 5S rRNA genes (13). There is also evidence for direct interactions of upstream regulatory elements and a new independent upstream promoter element centered about –17 to –20 (14). One of the remarkable features of TF IIIA is that it is capable of specific binding to the 5S rRNA gene and the gene product with high affinity and specificity, although three-dimensional structures of RNA and DNA are clearly different. The X-ray structure of the TFIIIA–DNA complex shows how zinc fingers have been deployed to bind to separated promoter elements (15).

5S rRNA is the only known RNA species that binds ribosomal proteins before it is incorporated into the ribosomes both in prokaryotes and eukaryotes (Fig. 1). In eukaryotes, 5S rRNA molecule binds only ribosomal protein L5, whereas in bacteria it interacts with three ribosomal proteins L5, L18 and L25. Protein L5 being a central component of the 5S rRNA export system interacts also with eukaryotic initiation factor 5A which on the other hand bind HIV-1 Rev. (16). 5S rRNA also binds the last protein (17) and enhances the methionyl- and isoleucyl-tRNA synthetase activities by direct interactions with MetRS and $tRNA^{Met}$ in the macromolecular aminoacyl-tRNA synthetases complex (18). Recently 5S rRNA has been identified in the degradosome complex (19). It has been also demonstrated that it is complementary with 12 and 15 nt strings to the intron 1 sequences of cobrotoxin b and cobrotoxin genes (20).

It seems that in addition to RNA–protein interactions an important role is played by the contacts of 5S rRNA with 23S rRNA. It has been shown that uridine 89 of 5S rRNA cross-links

*To whom correspondence should be addressed: Tel. +49 30 838 6402; Fax. +49 30 838 6403; Email: erdmann@chemie.fu-berlin.de

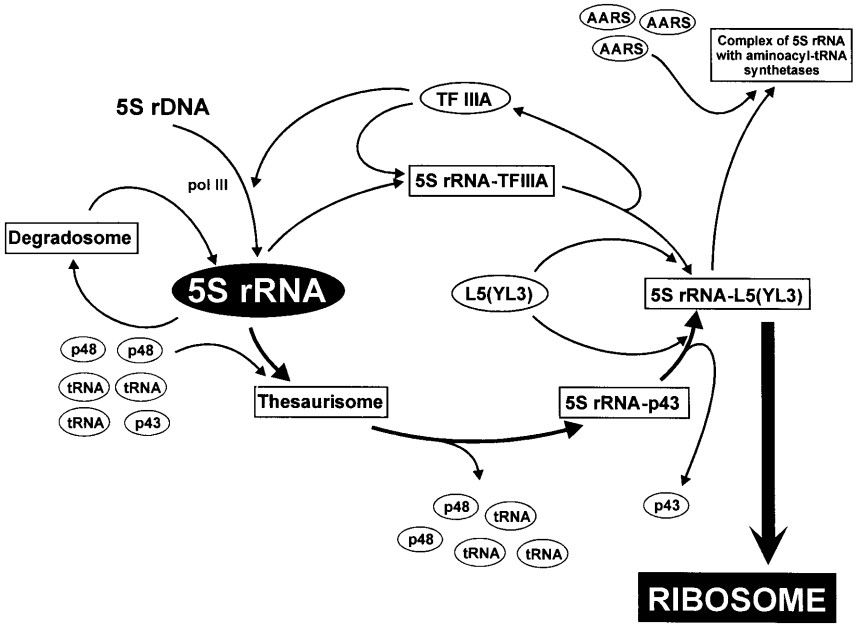

Figure 1. The activities and functions of 5S ribosomal RNA.

to U958, G1022 and G1138 of 23S rRNA, and thus 5S rRNA is topographically implicated in the peptidyl transferase and GTPase-asociated regions (21). Base paired interaction between 5S rRNA (residues 91–110) in the large subunit and 18S rRNA in the small subunit could contribute to the reversible association of the ribosomal subunits (22).

The 5S rRNA is an attractive model system for exploring fundamental issues of RNA conformation and RNA–protein interaction due to its relatively small size, easy to prepare and rich array of non-canonical base pairs (10).

The sequence of 5S rRNA is highly conserved throughout nature and phylogenetic analysis alone provided an initial model for its secondary structure. This model was later refined to include five helical regions, three internal loops, and two hairpin loops forming unknown three dimensional structure by chemical modification, site directed mutagenesis, physical characterization and computer modeling (6,7).

To get a consistent picture of structure–function relationships of 5S rRNA, detailed knowledge concerning the primary structure of this RNA species from different sources is required. Such is the content of the revised data bank.

DESCRIPTION OF THE DATA BANK

This edition of the database contains 1889 nucleotide sequences of 5S rRNAs and 5S rDNAs published up to September 1998. In comparison with the 1997 edition of the database, 267 entries are new (8). Most of them are partial sequences of plant genes. In Table 1 we show the distribution of the sequence entries for the main taxonomic groups. The database entries use the format of the EMBL Nucleotide Sequence Data Bank. The 5S rDNA nucleotide entries contains 5S rRNA coding sequence as well as information on the length of the original clone and location of the structural gene.

Table 1. A summary of the nucleotide sequence entries in the 5S rRNA data bank for major taxonomic groups

Taxonomic group	Number of entries
EUBACTERIA	439
ARCHAEA	60
ORGANELLES	72
Mitochondria	9
Plastids	63
EUKARYOTA	1318
Protista	53
Fungi	203
Animals	344
Plants	718
TOTAL:	1889

THE SECONDARY STRUCTURE OF 5S rRNA

The secondary structure of all analyzed 5S rRNAs consists of five helices (I–V), two hairpin loops (C and D), two internal loops (B and E) and a hinge region (A), organized into the three helix junction. The general secondary structure models of eukaryotic and prokaryotic 5S rRNAs are shown in Figure 2. Most of the 5S rRNA sequences can be folded acording to one of these models, although Eubacteria and Archaea show much higher variability than the Eukaryotes. The ability of the sequence to adopt correct, consensus secondary structure can be used to discriminate between the genes and pseudogenes that are often found in eukaryotic genomes.

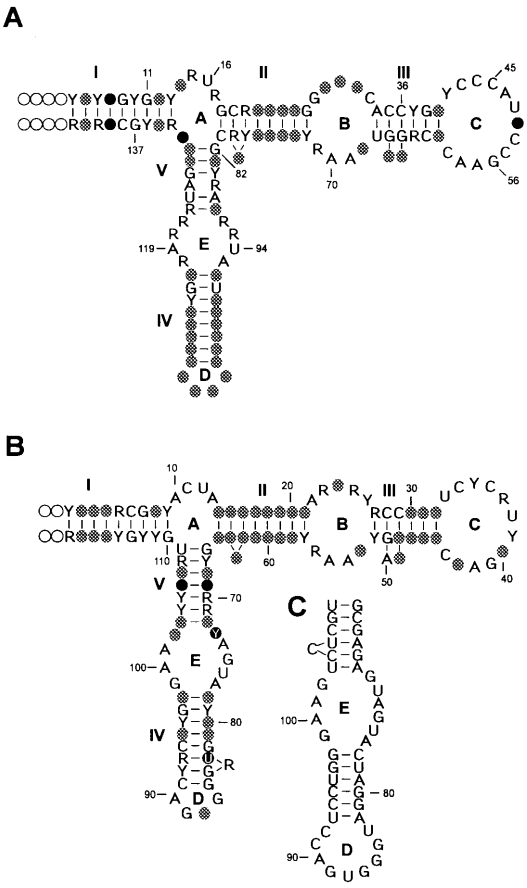

Figure 2. The general secondary structure models of eubacterial (**A**) and eukaryotic (**B**) 5S rRNAs and an alternative base-pairing scheme for the region of helix IV–loop E–helix V–loop D for plant 5S rRNA (**C**). The constant positions are marked with letters (R-purine; Y-pyrimidine), variable positions are marked with grey and open circles. The differences in nucleotide sequences of eukaryotic and eubacterial 5S rRNA are with black circles.

THE DATABASE ACCESS

Files with nucleotide sequence data and sequence alignments are available via the WWW at http://rose.man.poznan.pl/5SData/5SRNA.html . Given nucleotide sequences can be retrieved using the taxonomy browser or alphabetical list of organisms.

Send any comments, suggestions and corrections to Maciej Szymanski (mszyman@ibch.poznan.pl).

ACKNOWLEDGEMENTS

This work has been supported by the Deutsche Forschungsgemeinschaft (Gottfried Wilhelm Leibnitz Prize to VAE, the Sonderforschungsbereich 344-C8, the Deutsche Agentur für Raumfahrtangelegenheiten GmbH, the Fonds der Chemischen Industrie e.V and the Polish State Committee for Scientific Research.

REFERENCES

1 Yonath,A. and Francheschi,F. (1998) *Structure*, **6**, 679–684.
2 Ban,N., Freeborn,B., Nissen,P., Penczek,P., Grassucci,R.A., Sweet,R., Frank,J., Moore,P.B. and Steitz,T.A. (1998) *Cell*, **93**, 1105–1115.
3 Gray,M.W., Lang,B.F., Cedergren,R., Golding,G.B., Lemieux,C., Sankoff,D., Turmel,M., Brossard,N., Delage,E., Littlejohn,T.G., Plante,I., Rioux,P., Saint-Louis,D., Zhu,Y. and Burger,G. (1998) *Nucleic Acids Res.*, **26**, 865–878.
4 Spruyt,N., Delarbe,C., Gachelin,G. and Laudet,V. (1998) *Nucleic Acids Res.*, **26**, 3279–3285.
5 Rosset,R. and Monier,R. (1963) *Biochim. Biophys. Acta*, **68**, 653–656.
6 Barciszewska,M., Erdmann,V.A. and Barciszewski,J. (1996) *Biol. Rev.*, **71**, 1–25.
7 Zheng,P., Burrows,C.J. and Rokita,S.E. (1998) *Biochemistry*, **37**, 2207–2214.
8 Szymanski,M., Specht,T., Barciszewska,M., Barciszewski,J. and Erdmann,V.A. (1998) *Nucleic Acids Res.*, **26**, 156–159.
9 Correll,C.C., Freeborn,B., Moore,P.B. and Steitz,T.A. (1997) *Cell*, **91**, 705–712.
10 Dallas,A. and Moore,P.B. (1997) *Structure*, **5**, 1639–1653.
11 Perbandt,M., Nolte,A., Lorentz,S., Bald,R., Betzel,C. and Erdmann,V.A. (1998) *FEBS Lett.*, **429**, 211–215.
12 Cairns,C.A. and White,R.J. (1998) *EMBO J.*, **17**, 3112–3223.
13 Shastry,B.S. (1996) *J. Cell Sci.*, **109**, 535–539.
14 Lee,Y., Wong,W.M., Guyer,D., Erkine,A.M. and Nazar,R.N. (1997) *J. Mol. Biol.*, **269**, 676–683.
15 Nolte,R.T., Conlin,R.M., Harrison,S.C. and Brown,R.S. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 2938–2943.
16 Schatz,O., Oft,M., Dascher,C., Schebesta,M., Rosorius,O., Jaksche,H., Dobrovnik,D., Bevec,D. and Hauber,J. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 1607–1612.
17 Lam,W.-H., Seifert,J.M., Amberger,F., Graf,C., Auer,M. and Millar,D. (1998) *Biochemistry*, **37**, 1800–1809.
18 Ogata,K., Ohno,R., Morioka,S. and Terao,K. (1996) *J. Biochem.*, **120**, 869–880.
19 Bessarab,D.A., Kaberdin,V.R., Wei,C.-L, Liou,G.-G. and Lin-Chao,S. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 3157–3161.
20 Chang,L.-S., Cho,Y.-C., Lin,S.-R., Wu,B.-N., Lin,J., Hong,E., Sun,Y.-J. and Hsiao,C.-D. (1997) *J. Biochem.*, **122**, 1252–1259.
21 Sergiev,P., Dokudovskaya,S., Romanova,E., Tropin,A., Bogdanov,A., Brimacombe,R. and Dontsova,O. (1998) *Nucleic Acids Res.*, **26**, 2519–2525.
22 Azad,A.A., Failla,P. and Hanna,P.J. (1998) *Biochem. Biophys. Res. Commun.*, **248**, 51–56.