156–159 *Nucleic Acids Research, 1998, Vol. 26, No. 1*

5S rRNA Data Bank

Maciej Szymanski, Thomas Specht[1], Miroslawa Z. Barciszewska, Jan Barciszewski and Volker A. Erdmann[2,*]

Institute of Bioorganic Chemistry of the Polish Academy of Sciences, Noskowskiego 12, 61704 Poznan, Poland, [1]MetaGen GmbH. Ihnestrasse 63, Berlin, Germany and [2]Institut für Biochemie, Freie Universität Berlin, Thielallee 63, 14195 Berlin, Germany

Received October 2, 1997; Accepted October 3, 1997

ABSTRACT

In this paper we present the updated version of the compilation of 5S rRNA and 5S rDNA nucleotide sequences. It contains 1622 primary structures of 5S rRNAs and 5S rRNA genes from 888 species. These include 58 archaeal, 427 eubacterial, 34 plastid, nine mitochondrial and 1094 eukaryotic DNA or RNA nucleotide sequences. The sequence entries are divided according to the taxonomic position of the organisms. All individual sequences deposited in the 5S rRNA Database can be retrieved using the WWW-based, taxonomic browser at http://rose.man.poznan.pl/5SData/5SRNA.html or http://www.chemie.fu-berlin.de/fb_chemie/agerdmann/5S_rRNA.html . The files with complete sets of data as well as sequence alignments are available via anonymous ftp.

INTRODUCTION

This data bank was prepared in August and September 1997. It is just 30 years since the first nucleotide sequence of 5S rRNA was determined (1) and 35 years since this RNA species was been identified for the first time as a component of the large subunit of *Escherichia coli* ribosomes (2). Since that time a large amount of sequence data has been collected, however we are still far from the detailed knowledge of the tertiary structure and function of 5S rRNA. 5S rRNA appears to be a ubiquitous component of all prokaryotic and eukaryotic ribosomes. However, it has not been found in mitochondrial ribosomes of some fungi and animals. In prokaryotes and organelles, 5S ribosomal RNAs are synthesized as parts of single long transcripts, together with 16S and 23S rRNAs. The three individual components are then separated in a maturation process. In eukaryotes, 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. These genes are transcibed by polymerase III. The synthesis of 5S rRNA in eukaryotic cells is dependent on the binding of a 40 kDa protein, transcription factor IIIA (TFIIIA), to the internal control region of 5S rRNA genes. One of the remarkable features of TFIIIA 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.

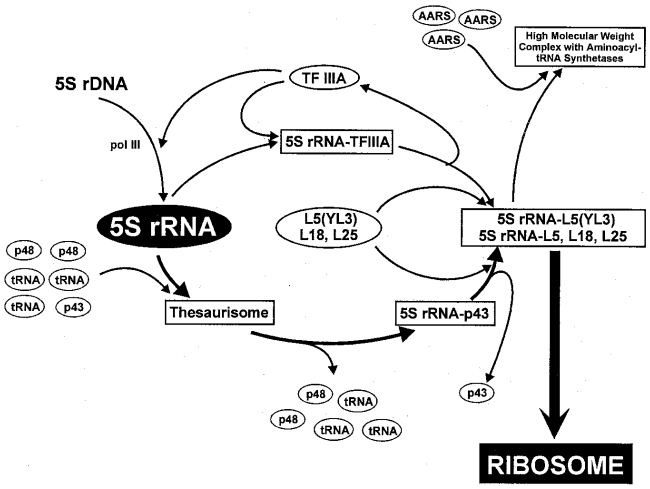

Figure 1. The activities and functions of 5S rRNA.

5S rRNA is a relatively small RNA molecule (120 nt long) and has been subject to various studies concerning its structure as well as biological function. It turns out that the nucleotide sequences of 5S rRNAs are strongly conserved and provide data that can be used in evolutionary analyses for the reconstruction of phylogenetic relationships between distant taxa.

5S ribosomal RNA is the only known rRNA species that binds ribosomal proteins before it is incorporated into the ribosomes both in prokaryotes and eukaryotes (Fig. 1). In eukaryotes the 5S rRNA molecule binds only ribosomal protein L5, whereas in bacteria it interacts with three ribosomal proteins L5, L18 and L25 (3,4). The 5S rRNA–protein interactions have been studied extensively by the analysis of the ribonucleoprotein complexes, mutant forms of 5S rRNA as well as proteins with the use of a variety of chemical and enzymatic probes. Recently several approaches have been made for crystallization of the whole 5S rRNA molecule and also its domains. However, no crystals suitable for structural analysis have been grown. The only X-ray and NMR data available are for the isolated domains of 5S rRNA (7–10).

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

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

Table 1. A summary of nucleotide sequence entries in the 5S rRNA Data Bank for particular taxonomic groups

Taxonomic group	No. of species	No. of sequences
ARCHAEA	**49**	**58**
Crenarchaeota	6	8
Pyrodictiales	1	1
Sulfolobales	5	7
Euryarchaeota	43	50
Halobacteriales	15	23
Methanobacteriales	5	8
Methanococcales	4	7
Methanomicrobiales	8	9
Thermococcales	2	2
Thermoplasmatales	1	1
EUBACTERIA	**332**	**427**
Chlorobium group	5	6
Chloroflexaceae	2	2
Cyanobacteria	4	5
Cytophagales	12	12
Deinococcaceae	5	7
Firmicutes	129	171
Actinobacteria	81	98
Clostridiobacteria	48	73
Fusobacteria	3	4
Planctomycetales	6	7
Proteobacteria	161	208
alpha subdivision	54	87
beta subdivision	17	21
gamma subdivision	81	89
delta subdivision	5	6
epsilon subdivision	4	5
Spirochaetales	4	4
Verrucomicrobiales	1	1
ORGANELLES	**43**	**43**
Plastids	34	34
Mitochondria	9	9
EUKARYOTA	**464**	**1094 (364)**
PROTISTA	34	49
FUNGI	129	194 (18)
Oomycota	6	6
Eumycota	123	188 (18)
Chytridiomycotina	2	2
Deuteromycotina	23	27
Zygomycotina	11	11
Ascomycotina	28	87 (17)
Basidiomycotina	59	61 (1)
PLANTAE	178	527 (221)
Thallobionta	35	37
Embryobionta	143	490 (221)
Bryophyta	4	4
Pteridophyta	5	5
Telomophyta	134	481 (221)
ANIMALIA	123	322 (135)
Parazoa	4	4
Porifera	4	4
Metazoa	119	318
Placozoa	1	1
Mesozoa	1	1
Cnidaria	6	7
Plathelminthes	2	3
Nemertini	2	3
Nematoda	10	20 (1)
Rotatoria	1	1
Brachiopoda	1	1
Bryozoa	1	1
Echiurida	1	1
Sipunculida	1	1
Pogonophora	1	1
Mollusca	9	9
Annelida	3	3
Arthropoda	43	134 (90)
Echinodermata	6	8
Hemichordata	1	2
Chordata	29	121 (44)
TOTAL:	**888**	**1622 (374)**

The number of partial sequences (included in the total number) is shown in brackets.

The aim of this work was to update and extend the 1997 compilation (6) with all currently known nucleotide sequences of 5S rRNAs and their genes.

DESCRIPTION OF THE DATA BANK

The 1998 edition of the data base contains 1622 nucleotide sequences of 5S rRNAs and 5S rDNAs published through September 1997. The sequence entries use the format of the EMBL Nucleotide Sequence Data Bank. There are 178 sequences that are not included in EMBL or GenBank. These can be easily identified, since they have an AC field blank. Table 1 shows the number of species and sequence entries in the data base for particular taxonomic groups.

The 5S rDNA sequence entries, in addition to the 5S rRNA coding sequence, contain information on the length of the original clone and the location of the structural gene.

An additional feature in the new edition of the 5S rRNA Data Bank is the collection of the eukaryotic intergenic spacer sequences available via WWW at the Eukaryotic 5S rRNA home page at: http://www.man.poznan.pl/5SData/Eukar_5S.html

THE SECONDARY STRUCTURE

The secondary structure of all 5S rRNAs consists of five helices (I–V), two hairpin loops (C and E), two internal loops (B and D) and a hinge region (A), formed as a three helix junction. The general secondary structures of eukaryotic and eubacterial 5S rRNAs are shown in Figure 2. Some sequences show the deviations from the general structure, that can be recognized as insertions or deletions in the multiple sequence alignment (5). The insertions and deletions found in eukaryotic sequences are listed in Table 2.

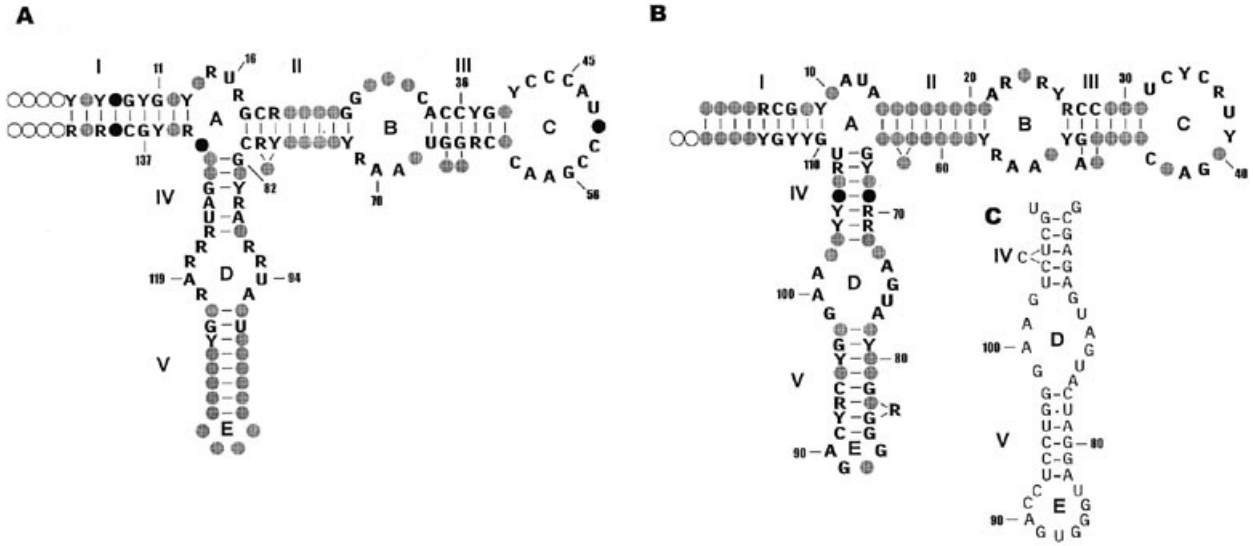

Figure 2. The general secondary structures of eubacterial (**A**) and eukaryotic (**B**) 5S rRNAs. An alternative base-pairing scheme for the region of helix IV–loop D–helix V–loop E, proposed for the plant 5S rRNAs is shown in C. The contstant positions are marked with letters (R, purine; Y, pyrimidine), variable positions are marked with grey and open circles. The black circles show differences between eukaryotic and eubacterial structures.

Table 2. A summary of insertions and deletions in 5S rRNA structures from eukaryotes

Nucleotide position	Organisms
INSERTIONS	
–2	*Chilomonas paramecium*
–1	*Chilomonas paramecium*, *Trypanosoma cruzi*, *Mastigamoeba invertans*, *Entosphenus japonicus*
1a	all Pteridophyta, Chlorophyta (except for *Spirogyra grevilleana*), some Zygomycetes
17a	*Porphyra yezoenzis*, *Porphyra tenera*, *Porphyra umbilicalis*
18a	*Gracilaria compressa*, *Gelidium amansi*, *Batrachospermum ectocarpum*
19a	*Carpopeltis crispata*, *Cyanidium caldarium*
21a	*Gloiopeltis complanata*
38a	all Rhodophyta, *Coemansia mojavensis*
49a	some Ascomycetes, some Deuteromycetes
50a	*Dugesia japonica*
73a	some Ascomycetes, some Deuteromycetes
90a	*Cyanidium caldarium*
105a	*Crypthecodinium cohnii*
105b	*Crypthecodinium cohnii*
105c	*Crypthecodinium cohnii*
106a	some Ascomycetes, some Deuteromycetes, some Basidiomycetes
106b	*Microstroma juglandis*
106c	*Microstroma juglandis*
117a	Chlorophyta (except for *Spirogyra grevilleana* and *Coleochaete scutata*)
DELETIONS	
1	*Trihoplax adhaerens*, *Anthocero punctatus*, *Gossypium hirsutum*
2	*Dicyema misakiense*, *Trichoplax adhaerens*
3	*Dicyema misakiense*
4	*Chlamydomonas* sp., *Crypthecodinium cohnii*
82	*Euglena gracilis*
115	*Chlamydomonas* sp
116	*Dicyema misakiense*
117	*Dicyema misakiense*, *Trichoplax adhaerens*
118	*Trichoplax adhaerens*

Table 3. A summary of the occurence of modified nucleosides found in 5S rRNAs

Nucleotide position	Modified nucleotide	Organisms
ARCHAEA		
32	2′-*O*-metylcytidine (Cm)	*Sulfolobus solfataricus*
35	*N*-4-acetyl-2′-*O*-methylcytidine (ac^4Cm)	*Pyrodictium occultum*
41 or 77	*N*-4-acetylcytidine (ac^4C)	*Pyrodictium occultum*
EUKARYOTA		
38a	pseudouridine (Ψ, F)	*Euglena gracilis*
49a	pseudouridine (Ψ, F)	some Ascomycetes

MODIFIED NUCLEOTIDES

With the exception of some fungi and archaebacteria the 5S rRNA molecules do not undergo posttranscriptional modification. The presence of only four modified nucleotides has been reported so far. These are pseudouridine, 2′-*O*-methylcytidine, *N*-4-acetylcytidine and *N*-4-acetyl-2′-*O*-methylcytidine. The summary of occurrence of modifications in 5S rRNAs is shown in Table 3.

DATABASE ACCESS

The files with sequence data and alignments are available via anonymous ftp at ftp.fu-berlin.de in the directory /science/db/5SrRNA. Individual sequences can be retrieved via WWW using the taxonomy browser or alphabetical listing of organisms at: http://rose.man.poznan.pl/5SData/5SRNA.html or http://www.chemie.fu-berlin.de/fb_chemie/agerdmann/5S_rRNA.html

Comments, suggestions and corrections can be addressed to: Maciej Szymanski (mszyman@ibch.poznan.pl) or Thomas Specht (thomas.specht@metagen.de).

ACKNOWLEDGEMENTS

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

REFERENCES

1 Brownlee,G.G., Sanger,F. and Barell,G. (1967) *Nature*, **215**, 735–736.
2 Rosset,R. and Monier,R. (1963) *Biochim. Biophys. Acta*, **68**, 653–656.
3 Horne,J. and Erdmann,V.A. (1972) *Mol. Gen. Genet*., **119**, 337–344.
4 Barciszewska,M., Erdmann,V.A. and Barciszewski,J. (1996) *Biol. Rev*., **71**, 1–25.
5 Szymanski,M., Barciszewska,M.Z., Barciszewski,J., Specht,T. and Erdmann,V.A. (1997) *Biochim. Biophys. Acta*, **1350**, 75–79.
6 Specht,T., Szymanski,M., Barciszewska,M.Z., Barciszewski,J. and Erdmann,V.A. (1997) *Nucleic Acids Res*., **25**, 96–97.
7 Betzel,C., Lorenz,S., Fürste,J.P., Bald,R., Zhang,M., Schneider,T., Wilson,K.S. and Erdmann,V.A. (1994) *FEBS Lett*., **351**, 159–164.
8 Nolte,A., Klussman,S., Bald,R., Betzel,C., Dauter,Z., Wilson,K.S., Furste,J.-P. and Erdmann,V.A. (1995) *FEBS Lett*., **374**, 292–294.
9 Correll,C.C., Freeborn,B., Moore,P.B. and Steitz,T.A. (1997) *J. Biomol. Struct. Dyn*., **14**, 792–793.
10 Varani,G. (1997) *Accounts Chem. Res*., **30**, 189–195.