236–237 *Nucleic Acids Research, 1996, Vol. 24, No. 1*

The Ribonuclease P Database

James W. Brown

Department of Microbiology, North Carolina State University, Raleigh, NC 27695, USA

Received August 10, 1995; Revised and Accepted September 19, 1995

ABSTRACT

Ribonuclease P is the endoribonuclease responsible for the removal of leader sequences from tRNA precursors. Ribonuclease P is a ribonucleoprotein, and in bacteria the RNA alone is capable of pre-tRNA processing *in vitro*, i.e. it is a catalytic RNA. The Ribonuclease P Database is a compilation of ribonuclease P sequences, sequence alignments, secondary structures, three-dimensional models and accessory information, in the form of a hypertext document available via the Worldwide Web.

INTRODUCTION

Ribonuclease P is the ribonucleoprotein endonuclease that cleaves transfer RNA precursors, removing 5′ precursor sequences and generating the mature 5′ terminus of the tRNA [see reference (1) for review]. In bacteria, in which RNase P is best-understood, the holoenzyme consists of two subunits, an RNA of ~400 nucleotides (140 kDa) and a protein of ~120 amino acids (14 kDa) (2). The RNA is the catalytic subunit of the holoenzyme; bacterial RNase P RNA is active, at high ionic strength *in vitro,* in the absence of the protein subunit (3). Our knowledge of the higher-order structure of RNase P RNA has come primarily from comparative analysis. In the last year, the number of bacterial RNase P RNA sequences available for analysis has more than doubled (to nearly 100), thanks to improved cloning methods and the determination of sequences cloned from mixed microbial populations (4).

On a yearly basis since 1991, RNase P RNA sequence alignments and secondary structures have been compiled and distributed in an informal publication, ‘The Book of P’, providing workers in the field a single source for comparative information. In 1994 the Ribonuclease P Sequence Database was established as an electronic means of distributing the Book of P and its second release (described here) has been developed into a rich Worldwide Web hypertext database.

DESCRIPTION

The Ribonuclease P Sequence Database is a compilation of RNase P sequences, sequence alignments, secondary structures, three-dimensional models and accessory information available electronically via Internet (see below for details on accessing the database). The hypertext format of the database makes access to its various components straightforward. The database primarily contains information on bacterial RNase P, focusing on the RNA subunit of the enzyme. Some information is also included on the archaeal, eucaryal and organellar RNase P RNAs. The database is updated as additional sequences and structural information becomes available.

The bacterial and archaeal RNase P RNA sequences in the database are available from lists arranged phylogenetically as individual sequences or an alignment in GenBank format. Secondary structures of each RNA are presented in Graphic Interchange Format (GIFs) and can be downloaded as high-quality Postscript files. Models of the three-dimensional structure of the RNA are provided in Protein Database (PDB) format and as Quicktime movies. The handful of currently available bacterial RNase P protein subunit sequences are also available as individual sequences or as an alignment. Accessory information, such as reference structures, phylogenetic trees, and GenBank accession numbers and reference listings are also made available as part of the database.

The hard-copy version of the RNase P Database, the ‘Book of P’, contains all of this information compiled into a single document. The Book of P is available in the database in Apple Macintosh formats (MacDraw and Miscrosoft Word) that can be readily imported into most drawing or word-processing programs on this machine. The Book of P is also provided in Postscript format, which can be readily downloaded to any Postscript-capable printer from any computer platform.

AVAILABILITY

The Ribonuclease P Database is accessible using any Web reader at the URL ftp://iubio.bio.indiana.edu/molbio/rnase-p/home.html (see Fig. 1). Those without Web access can obtain the files by anonymous FTP. Hard copies of the Book of P are also available for those without electronic access to the database upon request. The administrator of the Ribonuclease P database (J. W. Brown) can be contacted by electronic mail (jwbrown@mbio.ncsu.edu) or by mail at the address given above. Users of the Ribonuclease P Database should cite this publication, and are encouraged to provide corrections, new information, or other materials for inclusion in the database; unpublished information will be held confidential at the authors request.

ACKNOWLEDGEMENT

The author would like to thank Dr Don Gilbert, the administrator of the Indiana University Biology Archives, not only for technical assistance with establishing and providing a home for the Ribonuclease P Database, but for his years of service to the biocomputing community worldwide.

REFERENCES

1 Pace, N.R. and Brown, J.W. (1995) *J. Bacteriol.* **177**, 1919–1928.
2 Brown, J.W. and Pace, N.R. (1992) *Nucleic Acids Res.* **20**, 1451–1456.
3 Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) *Cell* **35**, 849–857.
4 Brown, J.W., Nolan, J.N., Haas, E.S., Rubio, M.A., Major, F., and Pace, N.R. (1995) *Proc. Natl. Acad. Sci. USA* (in press).
5 Brown, J.W., Haas, E.S., Gilbert, D.G. and Pace, N. R. (1994) *Nucleic Acids Res.* **22**, 3660–3662.

The Ribonuclease P Database

Copyright © 1995

The Ribonuclease P Database is a compilation of sequences, sequence alignments, secondary structures, and other information relating to ribonuclease P, in the form of a Worldwide Web system administrated by *Dr. James W. Brown*. Please submit any information you wish to be added to the Database to the curator. Although the Database currently contains only bacterial and archaeal sequences, it is hoped that eucaryal and organellar sequences will be included in later releases as further insight becomes available.

Every effort has been made to ensure the accuracy of this material; any sequence taken from this document, however, should be checked for accuracy before use either from the original publication (where possible and where the original sequence has not been found to contain errors) or from the GenBank/EMBL data file (if available).

Database release 2, 6/16/95

Read me!

Check here for the '*Release notes*', containing last-minute information, corrections, etc.

The Book of P, 6th Edition

The Book of P is the 'hard copy' version of the RNase P Database. It contains all of the sequence alignments, secondary structures, and some of the other graphics in the database. Individual sequences are not included (although they are in the alignment and structures). The current edition of the Book of P is #6.

The Book of P primarily contains information available to the public, either from published papers or sequence databases. Some otherwise unavailable information is presented with permission. For more information on unpublished information not presented in the Ribonuclease P Database and the Book of P, contact the database curator.

The Book of P is available in the following formats:

Macintosh MacDraw and Microsoft Word formats
Most Macintosh word-processing programs can read MS Word version 5.1 files, and most drawing programs can read MacDraw II files, so most Macintosh users should download this version of the book. The files are in the form of a single, 1.5Mbyte binhexed self-extracting archive.

Postscript
For non-Macintosh users, or those that just want to print the book out without messing around, the Book of P is provided in Postscript format (plain text, level 2). Most computers can send this directly for printing on Postscript laser printers. The book is broken-up into several parts (usually one page each), each as a separate Postscript file. Be sure to download ALL of the 39 files to get the entire book.

RNase P RNAs

At this time, only bacterial and archaeal RNase P RNA sequences are included in the RNase P Database.

The sequences in the RNase P Database are available as:

Individual sequences
Individual sequences of RNase P RNAs are in GenBank format (text). This link will take you to a listing of the sequences from which you can choose any sequence you want.

Aligned sequences
The entire collection of released bacterial RNase P RNAs is available as a single alignment, in GenBank multiple-sequence format (text).

Secondary structures
Individual sequences are also available as secondary structures in a variety of formats, including GIFs which can be viewed directly if you are using Netscape or Mosaic with this Web document.

RNase P proteins

At this time, only bacterial RNase P protein sequences are included in the RNase P Database.

These protein sequences are available as:

Individual sequences
Individual sequences of RNase P proteins are in GenBank format (text). This link will take you to a listing of the sequences from which you can choose the sequence you want.

Aligned sequences
The small collection of released bacterial RNase P protein is available as a single alignment, in GenBank multiple-sequence format (text).

Three-dimensional models

There are currently two published models for the three-dimensional structure of bacterial RNase P RNA. Both are preliminary, or 'working', models and are clearly correct in some aspects and incorrect in others. The two models share a few features, but are overall quite different. Have fun!

- *Harris/Pace model*
- *Westhof/Altman model*

Figures for general reference

A number of general-use 'reference' figures are available:

- *E.coli RNA secondary structure.*
- *E.coli RNA structure with tertiary interactions.*
- *B.subtilis RNA secondary structure.*
- *Consensus bacterial secondary structure.*
- *Tree of bacterial RNA structural variation.*
- *Tree of bacterial RNA sequences.*
- *PCR to obtain sequences from natural populations.*

Software

Covariation This is version 4.0 of the Macintosh Covariation program described in Brown, J. W., 1991, CABIOS 7:391-393. This version can perform standard covariation, compensatory changes, and mutual information analyses on RNA alignments of up to 32,000 nucleotides to identify secondary and tertiary structure. The program comes as a 900K binhexed self-extracting archive.

Citations & Accession numbers

Sequence citations This list provides reference citations and GenBank/EMBL accession numbers for each of the RNase P sequences.

RNase P bibliography This is a fairly inclusive bibliography of papers on RNase P. In fact, this is a downloaded copy of the curators 'Endnote' reference list!

Database curator

If you have comments, suggestions, even criticisms, or desire further information on the Ribonuclease P Database and The Book of P, contact:

Dr. James W. Brown
Department of Microbiology, Box 7615
North Carolina State University
Raleigh, NC 27695 USA
(919) 515-1231 (voice)
(919) 515-7867 (fax)
jwbrown@mbio.ncsu.edu

Figure 1. The Ribonuclease P Database 'Home page'. The appearance of the Ribonuclease P Database home page is shown as it appears in graphics-capable Worldwide Web browsers, e.g. Mosaic or Netscape. Links to the various sections of the Ribonuclease P Database are italicized.