

The database of the smallest known auto-replicable RNA species: viroids and viroid-like RNAs

Martin Pelchat, Patrick Deschênes and Jean-Pierre Perreault*

Département de Biochimie, Faculté de Médecine, Université de Sherbrooke, Sherbrooke, Québec J1H 5N4, Canada

Received September 14, 1999; Accepted September 16, 1999

ABSTRACT

This is an online database in order to facilitate research on viroid, viroid-like RNAs and human hepatitis delta virus by presenting a large number of sequences and related data in a comprehensive and user-friendly format (e.g., position of their self-catalytic domains, open reading frame, prediction of the most stable secondary structures, etc.). This online database is available on the WWW at <http://www.callisto.si.usherb.ca/~jpperra>

INTRODUCTION

Viroids, plant satellite viroid-like RNAs, and human hepatitis delta virus (vHDV) share a common proposed replication pattern known as a DNA-independent rolling circle mechanism. These pathogens are small, circular, single-stranded RNA species. Together, they form the 'brotherhood' of the smallest known auto-replicable RNAs. Since 1996, we have built an online database in order to facilitate research on viroid and related RNAs by presenting a large number of sequences and related data in a comprehensive and user-friendly format (e.g., position of their self-catalytic domains, open reading frame of the vHDV, prediction of the most stable secondary structures, etc.) (1–4). Today, more than 550 sequences are now part of this database comprising data from 29 viroid species, nine species of plant satellite viroid-like RNAs, four related species of RNA and vHDV (23 complete and 165 partial). This database is updated as soon as sequences become available and contains more entries than in the GenBank and EMBL nucleotide sequence libraries. Users of the viroid and viroid-like RNA database are encouraged to provide corrections, new information or other information for inclusion in the database.

DESCRIPTION

The database is organized in four levels. The welcome page (i.e. the first level) shows in a frame, which is always available and allows easy access to any subdivision, the four sections:

viroids satellite RNAs HDV others

The choice of a section will lead to the second level which is a summary table for each sequence subdivision. The viroids, satellite RNAs and other RNAs are divided according to the nature of the compiled RNA species. The choice of either a viroid or viroid-like RNA species in a summary level will lead

to the third level. Briefly, each species of RNA appearing in the database is listed by its complete name and number of sequence variants. This is followed, for each species, by a complete list of the sequence variants and their assigned nomenclature. The identification of a variant is based on its usual acronym followed by a number. The procedure for sequence identification and information compiled was presented previously (1). Additional data for each entry include accession numbers, bank loci (when available), number of nucleotides (total and by type), complete publication information, and the sequence in blocks of 10 nucleotides. For several sequences a table shows structural features, for example the position of the conserved sequences of the RNA self-catalytic motifs (e.g., hammerhead and delta). In addition, a secondary structure prediction of the most likely ancestral variant of most entries was derived using the RNAfold structure prediction package. These predicted structures are appended to the database in connect file format to allow easy manipulation (fourth level).

Unlike the viroid section, the vHDV section comprises all available sequences, irrespective of their completeness, since the majority of the partial sequences are informative (i.e., they correspond to the sequence of either the open reading frame or the ribozymes). Both vHDV complete and partial sequences were arbitrarily subdivided into several parts in order to accelerate the display. For vHDV sequence, a 'p' preceding the specific number is attributed to partial sequence (2).

There is also other useful information appended to the database. For example, a compilation of all natural sequence variants of the delta self-catalytic RNA motifs is available. Furthermore, sequence alignments of all complete vHDV nucleotide sequences and all amino acid sequences of the vHDV mRNA antigen are available through the summary tables (second level). This database provides an excellent reference point for further phylogenetic and structure–function studies of these RNA species.

ACKNOWLEDGEMENTS

This work was supported in part by grants from the Natural Sciences and Engineering Research Council (NSERC), the Medical Research Council (MRC) of Canada, and in part by a scientific team grant to J.-P.P. from 'Fonds pour la Formation des Chercheurs et l'Avancement de la Recherche du Québec' (FCAR). P.D. is recipient of the 'Fonds de la Recherche en Santé du Québec' (FRSQ). J.-P.P. is an MRC scholar.

*To whom correspondence should be addressed. Tel: +1 819 564 5310; Fax: +1 819 564 5340; Email: jperre01@courrier.usherb.ca

REFERENCES

1. Bussière, F., Lafontaine, D. and Perreault, J.-P. (1996) *Nucleic Acids Res.*, **24**, 1793–1798.
2. Lafontaine, D., Mercure, S. and Perreault, J.-P. (1997) *Nucleic Acids Res.*, **25**, 123–125.
3. Lafontaine, D.A., Mercure, S., Poisson, V. and Perreault, J.-P. (1998) *Nucleic Acids Res.*, **26**, 190–191.
4. Lafontaine, D.A., Deschênes, P., Bussière, F., Poisson, V. and Perreault, J.-P. (1999) *Nucleic Acids Res.*, **27**, 186–187.