104–106 *Nucleic Acids Research, 1997, Vol. 25, No. 1*

The guide RNA database

Augustine E. Souza, Thomas Hermann[1] and H. Ulrich Göringer*

Laboratorium für Molekulare Biologie, Genzentrum der LMU München am MPI für Biochemie, Am Klopferspitz 18 and [1]Max-Planck-Institut für Biochemie, 82152 Martinsried, Germany

Received October 2, 1996; Accepted October 8, 1996

ABSTRACT

The RNA editing process in protozoan parasites is controlled by small RNA molecules known as guide RNAs (gRNAs). The gRNA database is a comprehensive compilation of published guide RNA sequences from eight different kinetoplastid organisms. In addition to the RNA primary sequences, information on the gene localization, the experimental verification of the transcripts, and literature citations are provided. Accessory information includes the secondary structures of four *Trypanosoma brucei* gRNAs as well as a computer modelled three dimensional gRNA structure. The database is made available as a hypertext document accessible via the World Wide Web (WWW) or from the authors in a printed form.

INTRODUCTION

Guide RNAs (gRNAs) are small, metabolically stable mitochondrial transcripts identified only in kinetoplastid organisms such as *Trypanosoma*, *Leishmania* or *Crithidia*. The molecules carry out a central function during the unusual mitochondrial RNA processing reaction known as kinetoplastid (k) RNA editing (for recent reviews see 1,2). During editing uridylate residues get inserted into and deleted from mitochondrial transcripts thus completing the sequence information of these mRNAs. Guide RNAs provide the information for the U insertion/deletion process by base pairing to pre-edited mRNAs. They are encoded on the mitochondrial mini- or maxicircle DNA elements in kinetoplastid organisms and the RNAs are presumably primary transcripts. Guide RNAs have an average length of 50–70 nucleotides (nt) with a strong A/U nucleotide bias. The primary sequence of gRNAs can be divided into three functional domains: first, a region of complementarity located at the 5′-end, termed anchor sequence, which is thought to create the initial contact with the pre-edited mRNA; second, an informational sequence domain which presumably directs the editing reaction; and third, a posttranscriptionally added 3′ oligo(U) extension, sometimes of >20 nt in length. More than 200 different gRNAs have been estimated to be required for the editing of all encrypted genes in *Trypanosoma brucei* (3) and there is an ~3-fold higher coding capacity for gRNA genes in that organism. Thus, in addition to the large number of different gRNAs the potential for gRNA redundancy exists (4). Guide RNAs have been suggested to fold into simple secondary structures, comprising two consecutive stem loop elements with both terminal ends in a single-stranded conformation (5).

DESCRIPTION OF THE DATABASE

Release 1.0 of the database contains 235 gRNA sequence entries including published sequences through September 30, 1996. The sequences stem from eight different kinetoplastid species: *Trypanosoma brucei*, *Trypanosoma cruzi*, *Trypanosoma congolense*, *Trypanosoma equiperdum*, *Leishmania tarentolae*, *Leishmania infantum*, *Leishmania gymnodactyli* and *Crithidia fasciculata*. The compilation is arranged in tabular form, listing for each entry: organism and name of the gRNAs, their primary sequences [not including the 3′ oligo(U) extension] and their localization on the mitochondrial genome (see Fig. 1 for an example). The order in which the gRNAs have been listed is from left-to-right with reference to the linear maxicircle map as given in (6). The nomenclature of gRNAs differs depending on the laboratory involved and the molecules are listed in a 5′ to 3′ order: the gRNA required to edit a 5′ region of a mRNA sequence is listed before that which is involved in editing a 3′ region. The amount of sequence shown for a gRNA may exceed the actual length of the gRNA. This is because in many cases the 5′ and 3′ termini have not been determined experimentally or because heterogeneity has been observed when gRNAs have been analyzed by primer extension or cDNA sequencing. For 159 of the 235 gRNAs, the existence of the molecules within RNA preparations has been experimentally verified by Northern blotting, primer extension, direct cDNA cloning, or by being isolated as part of a gRNA/mRNA chimera. The remaining sequences have to be considered putative gRNAs based on their base complementarity to fully edited mRNA sequence domains. Since all sequences were collected from published information, the corresponding references are provided in an associated hypertext document including MEDLINE identification numbers. In most of the cases these references will provide an alignment of the gRNAs with their cognate mRNAs.

The database contains accessory information such as the experimentally verified secondary structures of four gRNAs from *T.brucei* (gA6-14, gA6-48, gND7-506, gCyb-558) (5) which are presented in Graphic Interchange Format (GIF) (see Fig. 2A for an example) and a three dimensional model for *T.brucei* guide RNA gND7-506 (Hermann and Gsringer, unpublished) (Fig. 2B).

* To whom correspondence should be addressed. Tel: +49 89 8578 2475; Fax: +49 89 8578 3810; Email: goeringe@alf.biochem.mpg.de

The guide RNA database

Maintained by: Augustine E. Souza and H. Ulrich Göringer
Genzentrum München am Max-Planck-Institut für Biochemie
Am Klopferspitz 18
82152 Martinsried
Germany

- Release 1.0 of the database contains most published gRNA sequences along with an explanatory note. The sequences are presented in tabular form and catalogue the following information:
  - kinetoplastid species
  - name of the mRNA to be edited
  - gRNA name (and individual clone names, if present)
  - gRNA sequence
  - mitochondrial genome localization
  - expression
  - references to the literature

- The references cited in the database are listed here. Some of the MEDLINE unique identifiers are useful only if you search the full MEDLINE database and not the free-access, WWW-available, genetics subsection.

- Secondary structures for four gRNAs derived in our laboratory are available as .gif files. Additionally, a working model for the 3D-structure of a gRNA is available. If you derive, or know of, any higher order structural information on gRNAs we would be happy to fit it in here:
  - gA6-14 (4k)
  - gA6-48 (4k)
  - gCyb-558 (4k)
  - gND7-506 (4k)
  - 3D model of a gRNA (58k).

- Links to related WWW sites:
  - Determination of nucleic acid extinction coefficients (USA)
  - The Zuker Home Page (USA)
  - RNA (the journal and the society; UK)
  - RNADraw (Sweden)

Please note that while we have attempted to ensure the accuracy of the given material, sequences extracted from this database should be cross-checked with those in the original references.
Corrections, new information, and other material for inclusion in this page are welcome.

Species	gRNA	Sequence	Source	RNA?	Refs.
	ND8				
T. brucei	gCR1[232]	CUUAAAGGGA AAGGAAAGGU GGAAGGUGGA AGGGAAGAAG AUAGAGUCAG AAUG	mini	ND	5
	gCR1[219]	AUAUAAUAGU AACACAGCAG AUAAGAUACA UAUAGAGAUA AUGACAG	mini	ND	5
	gCR1[205]	AUAUACAAUG GUAACUCGAU GGGUGGGUAA UUAGUGAUAU GAUGAAUUUA AUACUAU	mini	ND	5, 15
	gCR1[204]	AUACAAUGGU UAUUCAGUGG GUAGACAGAU AGUGAUAUGA UAGAC	mini	ND	5
	gCR1[190]	AUAUAAACGC AAUAAAUGGU UAUCAUGAGU CAAUGAAUUA AUGAUAU	mini	y	15
	gCR1[164]	AUAUAAAUCA CAUAUACGGC AAGUUAUAUA GCGUUGUAAA UGAUGUCAAU	mini	y	5, 15

Figure 1. Homepage of the gRNA database (URL: http://www.biochem.mpg.de/~goeringe/) including an example of the data presentation.

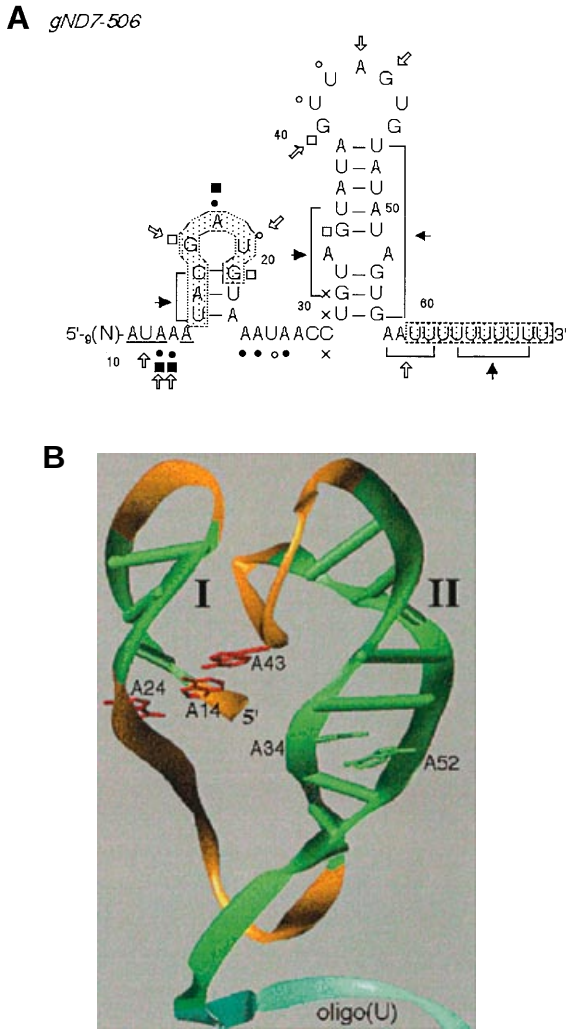

Figure 2. (**A**) Secondary structure model of gRNA gND7-506 from *T.brucei* based on surface probing data (5). Sensitivities of the RNA molecules to the various reagents and enzymes are indicated by the following symbols: kethoxal, open square; DMS, filled circle; DEPC, filled square; CMCT, open circle; CV, filled arrow plus bracket; S1, T1, T2: open arrow. X annotates frequent termination sites in untreated control samples. The oligo(U) tail and the achor region are boxed. (**B**) Three dimensional model of gND7-506 (Göringer and Hermann, unpublished). Helical regions are annotated in green, loop regions in orange, and the oligo(U) tail is in blue.

AVAILABILITY

The gRNA database is accessible via the URL: http://www.biochem.mpg.de/~goeringe/ . A printed version can be obtained upon request from any of the authors who can be contacted by electronic mail (goeringe@alf.biochem.mpg.de/souza@alf.biochem.mpg.de) or by mail at the address given above. Users of the database should cite this publication. Corrections, new entries, errors and omissions or other materials for inclusion in the database are welcome. Submission of new information will be accepted in any form. Unpublished data will be held confidential if required.

ACKNOWLEDGEMENTS

This work was supported by grants from the German ministry for education and research (BMBF) and the German research foundation (DFG) to H.U.G.

REFERENCES

1 Hajduk,S.L. and Sabatini,R.S. (1996) *Molecular Biology of Parasitic Protozoa*. IRL Press, Oxford, pp. 143–158.
2 Arts,G.J. and Benne,R. (1996) *Biochim. Biophys. Acta* **1307**, 39–54.
3 Stuart,K. (1993) *RNA Editing: The Alteration of Protein Coding Sequences of RNA*. Ellis Horwood, New York, pp. 25–52.
4 Riley,G.R., Corell,R.A. and Stuart,K. (1994) *J. Biol. Chem.* **269**, 6101–6108.
5 Schmid,B., Riley,G., Stuart,K. and Göringer,H.U. (1995) *Nucleic Acids Res.*, **23**, 3093–3102.
6 Benne,R. (1994) *Eur. J. Biochem.*, **221**, 9–23.