

# Virtual Ribosome—a comprehensive DNA translation tool with support for integration of sequence feature annotation

Rasmus Wernersson\*

Center for Biological Sequence Analysis, BioCentrum-DTU, Technical University of Denmark, Building 208, DK-2800 Lyngby, Denmark

Received February 14, 2006; Revised March 1, 2006; Accepted March 20, 2006

## ABSTRACT

**Virtual Ribosome is a DNA translation tool with two areas of focus. (i) Providing a strong translation tool in its own right, with an integrated ORF finder, full support for the IUPAC degenerate DNA alphabet and all translation tables defined by the NCBI taxonomy group, including the use of alternative start codons. (ii) Integration of sequences feature annotation—in particular, native support for working with files containing intron/exon structure annotation. The software is available for both download and online use at <http://www.cbs.dtu.dk/services/VirtualRibosome/>.**

## INTRODUCTION

A large number of software packages for translating DNA sequences already exist, as services on the World Wide Web [e.g. the Expasy Translate Tool (<http://www.expasy.ch/tools/dna.html>)], as command-line tools [e.g. the GCG package (1)] and as user-friendly graphical applications [e.g. DNA strider (a personal favorite) (2) and ApE (<http://www.biology.utah.edu/jorgensen/wayned/ape/>)]. However, many of these fine tools do not support translating sequences containing degenerate nucleotides, have no or limited support for alternative translation tables (including alternative initiation codons) and in general have problems handling special case situations. The software described here aims at addressing these issues and providing a comprehensive solution for translation. The software is build on the experience gained from writing and maintaining the Rev-Trans server (3).

Another part of the rationale for creating Virtual Ribosome is to create an easy and consistent way to map the underlying intron/exon structure of a gene onto its protein product.

This makes it easy to build datasets that can be used for analyzing how the underlying exon structure is reflected in the protein [e.g. how exon modules maps onto the 3D structure of the protein, see the FeatureMap3D server (4) elsewhere in this issue].

## SOFTWARE FEATURES

### Support for the degenerate nucleotide alphabet

The software has full support for the IUPAC alphabet (Table 1) for degenerate nucleotides. For example, the codon TCN correctly translates to S (serine) and not X (unknown) as often seen in other translators.

### Support for a wide range of translation tables

Full support for all translation tables defined by the NCBI taxonomy group (5) (see the list below). The command-line version of the software also has support for

Table 1. IUPAC alphabet of degenerate nucleotides

Letter	Description	Bases represented
A	Adenine	A
T	Thymine	T
G	Guanine	G
C	Cytosine	C
Y	pYrimidine	C T
R	puRine	A G
S	Strong	G C
W	Weak	A T
K	Keto	T G
M	aMino	A C
B	Not A	C G T
D	Not C	A G T
H	Not G	A C T
V	Not T/U	A C G
N	aNy	A C G T

\*Tel: +45 45252485; Fax: + 45 45931585; Email: raz@cbs.dtu.dk



If a GenBank or TAB file is supplied as input, only the exonic parts of DNA sequences is used for the translation. Furthermore, the underlying exon structure will be rejected in the translated sequence (also in the TAB format). By default, each amino acid will be annotated with a number indicating the exon that encoded this particular amino acid (see example below).

```
Name: 'J00044, Goat adult alpha-ii-globin gene'
MVLSAADKSNVKAAGWKGSGNAGAYGAEALERMFLSPFTTKTYPPHFDLSHGSAQVKGHG 60
1111111111111111111111111111111222222222222222222222222222222 60
EKVAAALTKAVGHLLDPLGTLSDLSDLHAHKLRVDPVNFKLLSHSLVTLACHHPSDFTP 120
2222222222222222222222222222222222222222222222233333333333333333 120
AVHASLKDPLANVSTVLTSKYR* 143
3333333333333333333333333333333333333333333333333333333333333333 143
```

Alternatively, the positions and the phase of the introns can be indicated.

- Phase 0: an intron exists right before the codon encoding the amino acid.

- Phase 1: an intron exists in between positions 1 and 2 of the codon.
- Phase 2: an intron exists in between positions 2 and 3 of the codon.

The following example illustrates the principle.

```
Name: 'J00044, Goat adult alpha-ii-globin gene'
MVLSAADKSNVKAAGWKGSGNAGAYGAEALERMFLSPFTTKTYPPHFDLSHGSAQVKGHG 60
.....2..... 60
EKVAAALTKAVGHLLDPLGTLSDLSDLHAHKLRVDPVNFKLLSHSLVTLACHHPSDFTP 120
.....0..... 120
AVHASLKDPLANVSTVLTSKYR* 143
..... 143
```

### Easy to use interface

The interface to the Virtual Ribosome server has been designed to be intuitive and easy to use. Figure 2 shows the basic part of the interface. Notice that it is possible to submit a

**Virtual Ribosome - version 1.1**

The Virtual Ribosome is a comprehensive tool for translating DNA sequences to the corresponding peptide sequences.

Besides being a strong translation tool in its own right (with an integrated ORF finder, support for all translation tables defined by the NCBI taxonomy group, and a number of options regarding START and STOP codons), the Virtual Ribosome can work directly on files containing annotation of gene structure. This makes it easy to map various aspects of Intron/Exon structure onto the translated sequence.

Instructions    Output format    Software download    Article abstract

Paste in DNA sequences in FASTA, GenBank or TAB format

```
>gi|74268119:41-469 Bos taurus hemoglobin alpha chain
ATGGTGCTGCTGCCGCGCAAGGGCAATGCAAGGCCGCTGGGGCAAGGTGGCGGCCACGCTGCGAG
AGTATGGCGCAGAGGCCCTGGAGAGGATGTCCTGAGCTTCCCACCACCAAGACCTACTTCCCACCTT
CGACCTGAGCCAGCGCTCCGCGAGGTCAAGGGCCACGGCGGAAGTGGCCGCCGCTGACCAAAGCG
GTGGAACACCTGGACGACCTGCCCGTCCCTGCTGAAGTGAAGTGAAGTGCACGCTCACAAAGCTCCGCTG
TGGACCCTGCAACTTCAAGCTTCTGAGCCACTCCCTGCTGGTGACCTTCCACCTCCCAAGTGA
```

Choose File    no file selected

View [example DNA files](#)

Submit query    Clear fields

Instructions: Basic usage - Paste in or upload one or more DNA sequences in FASTA (sequence only), GenBank (CDS sections are processed) or TAB (sequence and intron/exon annotation) format and hit submit. The Virtual Ribosome will then translate the DNA sequences using the standard genetic code (by default). Options can be customized in the section below.

Options

Figure 2. Screenshot of the basic part of the Virtual Ribosome interface.

sequence for translation using the default parameters, without having to scroll through a page of obscure options. The options are grouped into logical sections further down the web page. For each option a short explanation is provided together with a link to a detailed description.

## ACKNOWLEDGEMENTS

Special thanks to Ulrik de Lichtenberg for comments on the content and layout of the manuscript, and to Henrik Nielsen for inspiration for implementing the 'Intron phase' functionality. This work is supported by a grant from The Danish National Research Foundation and The Danish Research Agency. Funding to pay the Open Access publication charges for this article was provided by The Danish Research Agency.

*Conflict of interest statement.* None declared.

## REFERENCES

1. Dölz,R. (1994) GCG: translation of DNA sequence. *Methods Mol. Biol.*, **24**, 129–142.
2. Douglas,S.E. (1994) DNA Strider. A Macintosh program for handling protein and nucleic acid sequences. *Methods Mol. Biol.*, **24**, 181–194.
3. Wernersson,R. and Pedersen,A.G. (2003) RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res.*, **31**, 3537–3539.
4. Wernersson,R., Rapacki,K., Stærfeldt,H.-H., Sackett,P.W. and Mølgaard,A. (2006) FeatureMap3D: a tool to map protein features and sequence conservation onto homologous structures in the PDB. *Nucleic Acids Res.*, **34**, W536–W540.
5. Wheeler,D.L., Chappey,C., Lash,A.E., Leipe,D.D., Madden,T.L., Schuler,G.D., Tatusova,T.A. and Rapp,B.A. (2000) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **28**, 10–14.
6. Abramczyk,D., Tchorzewski,M. and Grankowski,N. (2003) Non-AUG translation initiation of mRNA encoding acidic ribosomal P2A protein in *Candida albicans*. *Yeast*, **20**, 1045–1052.
7. Wernersson,R. (2005) FeatureExtract—extraction of sequence annotation made easy. *Nucleic Acids Res.*, **33**, W567–W569.