

DNAssist: the integrated editing and analysis of molecular biology sequences in Windows

Hugh-G. Patterton¹ and Steven Graves^{2,*}

¹Department of Biochemistry, University of Cape Town, University Private Bag, Rondebosch 7700, South Africa and ²Los Alamos National Laboratories, Biosciences Division, M888, Los Alamos, NM 87545, USA

Received on December 14, 1999; revised on February 24, 2000; accepted on February 25, 2000

Abstract

Motivation: *The programs currently available for the analysis of nucleic acid and protein sequences suffer from a variety of problems: Web-based programs often require inconvenient reformatting of sequences when proceeding from one analysis to the next, and commercial-console-based programs are cost prohibitive. Here, we report the development of DNAssist, an inexpensive, multiple-document, interface program for the fully integrated editing and analysis of nucleic acid and protein sequences in the familiar environment of Microsoft Windows.*

Availability: *DNAssist is available as shareware and may be downloaded from <http://www.dnassist.com>.*

Contact: *support@dnassist.com*

DNAssist is a program running under Windows 95, 98 or NT4.0, where the investigator is presented with an integrated environment in which sequences are edited and analyses performed (Figure 1), similar to the DNA Strider (Marck, 1988) application for the Apple Macintosh. All file and window management, sequence editing and analysis operations are selected from the main menu, a context-sensitive pop-up menu, or from the toolbar. A selected analysis method is directly applied to the sequence or sequence selection in the editor window with input focus. There are four different editor window types that will accept DNA, degenerate DNA, RNA or protein sequences represented by IUPAC-IUB specified sequence symbols. A sequence can be entered into an editor window from the keyboard or loaded from disk, and can be of a length that is limited only by the physical memory of the microcomputer. DNAssist recognizes sequence files in text, Fasta, GenBank, GCG, DNA Strider or DNAssist format. Sequences can also be dragged-and-dropped within or between compatible editor window types or pasted from the clipboard.

The analysis methods available in DNAssist include sequence conversion, DNA translation, open reading frame identification, pattern searching, restriction enzyme analysis, transcription-factor site analysis, nucleosome position identification, multiple-sequence alignment and the calculation of physicochemical properties. In addition, all data such as codon tables and restriction enzyme databases used by DNAssist for analysis are available on the Internet.

DNAssist can convert a nucleic acid sequence to a protein sequence according to the entries in the selected codon table. The codon files used by DNAssist can be downloaded from the Internet (http://www.gcg.com/techsupport/data/codon_freq_tables.html). DNAssist can also reversibly inter-convert DNA and RNA sequences, translate a DNA sequence in multiple reading frames, and reverse translate a protein sequence to a degenerate DNA sequence.

Investigators can use DNAssist to calculate general physicochemical properties of nucleic acid and protein sequences such as the melting temperature, iso-electric point and molar absorption coefficient, as well as the hydrophobicity, hydrophilicity and antigenicity profiles of protein sequences.

Restriction enzyme analyses make use of data from Rebase (Roberts and Macelis, 1999). Updated versions of the enzyme data file can be downloaded from the Rebase Web site (<http://rebase.neb.com/rebase/rebase.files.html>). The result of a restriction enzyme analysis can be displayed as a graphic map, a sequence listing, or in tabulated format (see Figure 1).

Multiple alignments of DNA, degenerate DNA, RNA or protein sequences can be performed, and the aligned sequences can be displayed with identical nucleotides or amino acid residues, or conserved amino acid residues, shown in font foreground and background colors assigned by the user. Aligned sequences can be copied or saved to disk as RTF text and edited in third-party word processors.

*To whom correspondence should be addressed: Steven Graves, P.O. Box 1024, Los Alamos, NM 87544, USA, E-mail: steveg@dnassist.com.

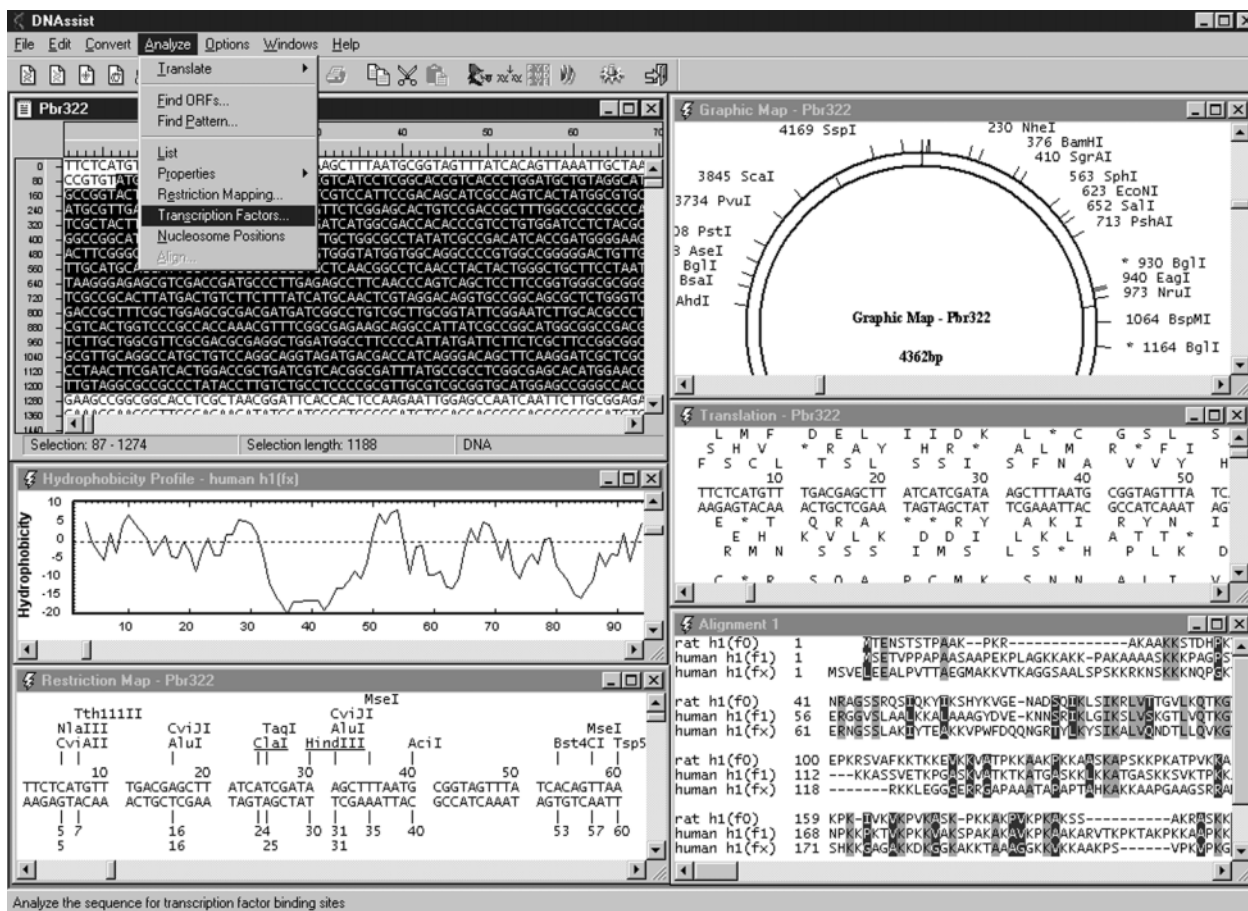


Fig. 1. DNAssist allows the user to enter and to edit nucleic acid and protein sequences in an environment that behaves like a Windows-based word processor. Sequences are analysed by selecting the required method from the main menu, toolbar or from a context-sensitive pop-up menu. The selected analysis method is directly applied to the sequence or sequence selection in the editor window with focus, and the result displayed in an MDI child window. Analysis results can be printed, saved or copied to the clipboard in rich text format or enhanced windows metafile format, and imported into compatible third-party applications.

The investigator can use DNAssist to analyse a DNA sequence for the presence of transcription-factor recognition sequences listed in the Transfac database (Heinemeyer *et al.*, 1999). Updated versions of the Transfac database may be downloaded from the Transfac Web site (<http://transfac.gbf.de/TRANSFAC/index.html>). The result of a transcription factor analysis can be displayed as a graphic map, sequence listing or in tabulated form, where the name and binding position of the selected transcription factors are given.

DNAssist is a convenient, integrated sequence editing and analysis program for the Microsoft Windows platform. The program presents an excellent alternative to programs that are currently available, but that are difficult to use or, most importantly, financially prohibitive to many molecular biology laboratories with modest budgets. We aim to maintain and support DNAssist as a

general sequence editing and analysis tool, and also plan to incorporate additional features in future as the need becomes evident. In addition, the use of DNAssist as a client-side interface to other popular databases on the Internet will be explored.

References

- Heinemeyer, T., Chen, X., Karas, H., Kel, A.E., Kel, O.V., Liebich, I., Meinhardt, T., Reuter, I., Schacherer, F. and Wingender, E. (1999) Expanding the TRANSFAC database towards an expert system of regulatory molecular mechanisms. *Nucleic Acids Res.*, **27**, 318–322.
- Marck, C. (1988) DNA Strider: a C program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Res.*, **16**, 1829–1836.
- Roberts, R.J. and Macelis, D. (1999) REBASE—restriction enzymes and methylases. *Nucleic Acids Res.*, **27**, 312–313.