CABIOS APPLICATIONS NOTE

Vol. 12 no. 3 1996
Pages 247–249

RNAdraw: an integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows

Ole Matzura* and Anders Wennborg[1,2]

Abstract

The use of algorithms for calculation and analysis of RNA secondary structures has largely been limited to mainframe computers, mainly due to the 16-bit memory restrictions imposed by MS-DOS. The program presented here, RNAdraw, moves calculations to the 32-bit Microsoft Windows environments with an intuitive user interface with extensive viewing, editing and printing possibilities. The algorithms for secondary structure/basepair probability matrix/heat curve calculation have been ported directly from a 32-bit Unix environment. RNAdraw also offers novel features such as the options to edit energy parameters, extract structures of different probability levels, create de novo secondary structures interactively, and combine viewing of structures and basepair probabilities.

Introduction

The biological significance of RNA secondary structure is increasingly apparent from research in many different biological systems. The importance of structure for the processive maturation and function of ribosomal RNA and transfer RNA is well recognized. Furthermore, structural features of messenger RNA (mRNA) have been shown to be involved in several processes, such as their effect on antisense RNA pairing and in the regulation of gene expression through specific binding of regulatory proteins (Theil, 1993). There is therefore a growing interest in molecular biology laboratories for easy access to RNA secondary structure calculation algorithms.

The program described here, RNAdraw, offers calculations of RNA optimal structures, basepair probability matrices and heat curves on Intel x86 compatible computers running Windows 3.1–3.11/95/NT, eliminating the need for mainframe computer systems. The program provides a consistent user interface with many possibilities to view, print, import/export and edit calculation results.

Department of Medical Biophysics and [1]Microbiology and Tumorbiology Center, Karolinska Institute, Box 280, S-171 77 Stockholm, Sweden

[2]Present address: Department of Bioscience, Karolinska Institute, S-14157 Huddinge, Sweden

**To whom correspondence should be addressed. E-mail: info@base8.se*

RNAdraw implements the following methods of RNA secondary structure analysis:

1. Calculation of the thermodynamically optimal secondary structure using experimentally derived energies for substructural motifs with a dynamic programming algorithm (Zuker and Stiegler, 1981). It is possible to enter basepairing constraints prior to calculations.
2. Calculation of a basepair probability matrix from the partition function resulting in an array of basepair probabilities (McCaskill, 1990). It is furthermore possible to extract structures from this matrix, resulting in structures of different probability levels.
3. Calculation of the specific heat capacity for a sequence in a specified temperature range from the partition function by numerical differentiation (McCaskill, 1990).

The calculation algorithms were all ported directly from the Vienna RNA package (Hofacker *et al.*, 1994). The secondary structure graphic coordinates are calculated according to the algorithm of Bruccoleri and Heinrich (1988).

Program description

Standard Windows features, such as online help, drag-and-drop, toolbars, status bars, high-quality printing and clipboard support are implemented. In particular, extensive use is made of the right mouse button to access menus and functions, allowing the user to operate RNAdraw with high efficiency and ease. A user-customizable RNA data library can be maintained from within the program, allowing easy access to commonly used data files.

Figure 1 shows a sample RNAdraw session, illustrating, row-wise from the top (left to right):

1. The *Data List*: displays open files and their data items in a hierarchy-like manner. It is mainly via a data-items right button popup menu that different functions in RNAdraw are activated.
2. The *Sequence Editor*: allows the user to edit RNA sequences prior to calculations. Sequences can either be entered directly or imported from text files, the Windows clipboard or GenBank search result files.
3. The *2D coordinate view*: displays RNA secondary structures which have either been calculated with

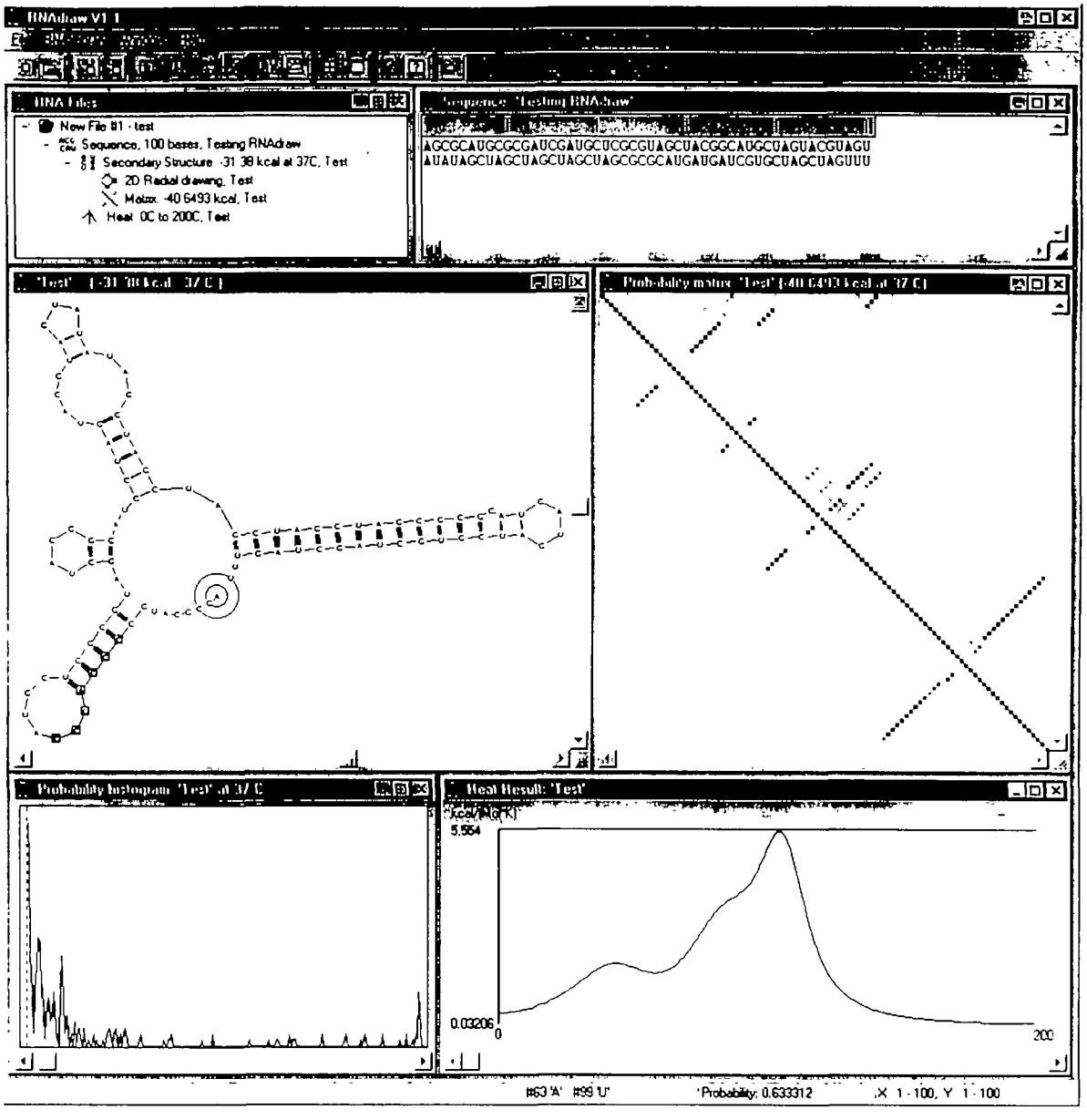

Fig. 1. RNAdraw main window with six sample data windows as described in the text.

RNAdraw or imported from external data files (such as Mfold connect files). It is possible to set many viewing parameters, zoom/rotate/move the structure interactively and search for and mark individual bases. Basepairing probabilities can be indicated by thicker basepair lines for basepairs with higher probabilities.

4. The *Basepair-matrix view*: shows secondary structure basepairs in the lower left triangle and a basepair probability matrix (if calculated) in the upper right triangle. Basepair probabilities are displayed in the status bar as the cursor is moved over the matrix. It is possible to pair/unpair bases interactively by pointing at them in the lower left triangle. Basepair changes are directly applied to open two-dimensional coordinate views and structure energy, allowing the user to

Table I. Memory requirements for structure and matrix calculation

Sequence length (bases)	Calculation structure and matrix (kbytes)
250	390
500	1530
1000	6100
1500	13570

Table II. (Structure calculation)/(structure and basepair probability matrix calculation) times (s) are shown relative to sequence length

Sequence length	RNAdraw		RNAFold	
	486-66DX2 20 Mbytes RAM	Pentium-90 32 Mbytes RAM	AS 6000 32 Mbytes RAM	Sun SPARC 10 (dual processor) 512 MB RAM
250	16/59	7/22	14/38	9/25
500	93/364	36/136	74/277	49/142
750	253/1073	102/394	185/530	131/403
1000	510/2310	213/921	346/1050	251/841

interactively monitor structure and energy changes. It is also possible to extract structures from the probability matrix, resulting in structures of different probability levels instead of energetically favorable ones.

5. The *Probability Histogram view*: displays a histogram over all probabilities in the underlying probability matrix. A cutoff value can be adjusted, filtering out basepairs below a certain probability from the probability matrix triangle.
6. The *Heat Curve view*: shows a calculated heat capacity curve for an RNA sequence. A marker can be used to see energy values at different temperatures.

The contents of the views 3–6 described above can all be printed or exported for incorporation in, for example, word-processing documents.

Energy parameters

It is possible to adjust all energy parameters used by the underlying algorithms. This opens several possibilities, e.g. (i) monitoring structure/matrix changes with different energy parameters; (ii) entering new experimentally derived energies; (iii) 'ruling out' certain motifs by assigning these very high/low energy values.

Introduced changes will affect all calculation algorithms in RNAdraw. The modified energy parameters can easily be saved and loaded from disk, allowing the user to work with a variety of energy parameter schemes.

Memory requirements and calculation performance

The memory requirements for RNAdraw depend on the sequence length and on the type of calculation to be done. As shown in Table I, sequences of 1000 bases can easily be analyzed on a computer with 16 Mbytes of memory. The program requires 800 kbytes of free disk space and will run on any of the Win32 platforms mentioned above.

The issue of calculation performance has been one of the main arguments against RNA structure calculation on Intel x86 computers. Today, Pentium computers provide substantial calculation power, as shown in Table II. A calculation of the secondary structure of 1000 bases takes <4 min on a Pentium-90 system. The ported version of NAVIEW (Bruccoleri and Heinrich, 1988) is extremely fast: the coordinate calculation for a structure with 7250 bases takes <4 s on a 486-66DX2 system. Running RNAdraw on Windows 95 or Windows NT also gives the possibility of doing multiple calculations simultaneously in the background, allowing the user to continue working with RNAdraw and other applications as the calculations are performed.

Availability

RNAdraw is freeware. The latest information about RNAdraw and updates is available on the RNAdraw WWW homepage (see below). The latest version of RNAdraw is available through the following channels: (i) via anonymous ftp at broccoli.mfn.ki.se in the pub/rnadraw directory; (ii) via the RNAdraw WWW homepage, located at http://mango.mef.ki.se/~ole/rnadraw/rnadraw.html.

To receive a detailed printed manual and direct information on future updates, it is possible to register oneself as a RNAdraw user, as instructed in the RNAdraw help files.

Acknowledgements

We would like to thank Professor Alexander von Gabain and Professor Rudolf Rigler for their encouragement and support. We are also grateful to Ivo L.Hofacker for helpful discussions and advice on the Vienna RNA package.

References

Bruccoleri,R. and Heinrich,G. (1988) An improved algorithm for nucleic acid secondary structure display. *Comp Applic. Biosci.*, **4**, 167–173.

Hofacker,I.L., Fontana,W., Stadler,P.F., Bonhoeffer,L.S., Tacker,M. and Schuster,P. (1994) Fast folding and comparison of RNA secondary structures. *Chem. Monthly*, **125**, 167–188.

McCaskill,J.S. (1990) The equilibrum partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105–1119.

Theil,E.C. (1993) The IRE (iron regulatry element) family: structures which regulate mRNA translation or stability. *Biofactors*, **4**, 87–93.

Zuker,M. and Stiegler,P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxillary information. *Nucleic Acids Res.*, **9**, 133–148.

Received on June 30, 1995, revised and accepted on May 29, 1996