Nucleic Acids Research, 1997, Vol. 25, No. 22 **4679–4684**

RnaViz, a program for the visualisation of RNA secondary structure

Peter De Rijk and Rupert De Wachter*

Departement Biochemie, Universiteit Antwerpen (UIA), Universiteitsplein 1, B-2610 Antwerpen, Belgium

Received May 9, 1997; Revised and Accepted September 30, 1997

ABSTRACT

RnaViz is a user-friendly, portable, windows-type program for producing publication-quality secondary structure drawings of RNA molecules. Drawings can be created starting from DCSE alignment files if they incorporate structure information or from mfold ct files. The layout of a structure can be changed easily. Display of special structural elements such as pseudo-knots or unformatted areas is possible. Sequences can be automatically numbered, and several other types of labels can be used to annotate particular bases or areas. Although the program does not try to produce an initially non-overlapping drawing, the layout of a properly positioned structure drawing can be applied to a newly created drawing using skeleton files. In this way a range of similar structures can be drawn with a minimum of effort. Skeletons for several types of RNA molecule are included with the program.

INTRODUCTION

RNA molecules form a structure of helical regions interspersed with single stranded areas. This structure is important in the function of these molecules, and its knowledge has already contributed to the understanding of processes such as the splicing of group I (1) and group II (2) introns, the functional role of rRNA in protein synthesis (3) and the function of RNase P (4). In the case of rRNA, secondary structure features can be helpful to fine-tune the alignment of sequences for phylogenetic studies.

The secondary structure of RNA molecules can be studied using experimental, thermodynamic and comparative methods. Programs that calculate the most thermodynamically favorable structure such as mfold (5) produce connection data: a list of bases and of numbers indicating secondary structure interactions. In DCSE (6) the structural information is incorporated in the alignment by interspersing the sequence with special symbols denoting the start and end of structural features. A special 'helix numbering line' contains the names for the helix strands, and indicates which are complementary. Although these forms of structural information are very useful, they cannot be used for publications as they are difficult to evaluate. Since the classical 2D drawing of the secondary structure is easier to grasp and more aesthetically pleasing, it is the preferred visualization for publications.

Although several programs (7–12) exist that produce 2D structure drawings, they share some of the following problems: Most are too tightly coupled to an energy minimization prediction program to be of general use. Furthermore, the user cannot easily change the produced layout; much effort has been put into automatically producing a layout where none of the helices overlap, but this often does not properly emphasize similarities in structure because of insertions or deletions in less conserved areas. Other common problems are limitations in the size of molecule that can be displayed, and the inability to handle complex structural elements such as pseudo-knots.

RnaViz is a program for producing publication-ready secondary structure drawings starting from the connection data in the ct format as produced by mfold (5) or alignments with extra structure information in the DCSE format (6). It does not try to produce non-overlapping drawings, so the first drawing produced for a new molecule might show considerable overlap. However, this structure can be easily arranged interactively according to the user's wishes. As illustrated in Figures 1 and 2, RnaViz is capable of producing large as well as complicated structures. The layout or skeleton of a structure can be saved to a file, and used as a template to automatically arrange similar structures in the same layout. Skeletons for several molecules are included in the package. The program also incorporates many options for labeling the structure or emphasizing special features in it.

MATERIALS AND METHODS

RnaViz is implemented using a combination of C (13) and Tcl/Tk (14). Tcl is a high level scripting language originating from the University of Berkeley and now being developed at SUN labs (Mountain View, CA). Tk is an extension to Tcl, which can be used to create portable interfaces. C is used for parts where speed is critical. The use of a combination of Tcl/Tk and C has several advantages. Although C is portable, the libraries to create an interface are not. Tcl/Tk does provide the possibility to produce interface code that can be ported to MS Windows, MacOS and a wide variety of Unix systems. As a bonus, a Tcl/Tk interface can be easily customized and extended by the user.

*To whom correspondence should be addressed. Tel: +32 3820 2319; Fax: +32 3820 2248; Email: dwachter@uia.ua.ac.be

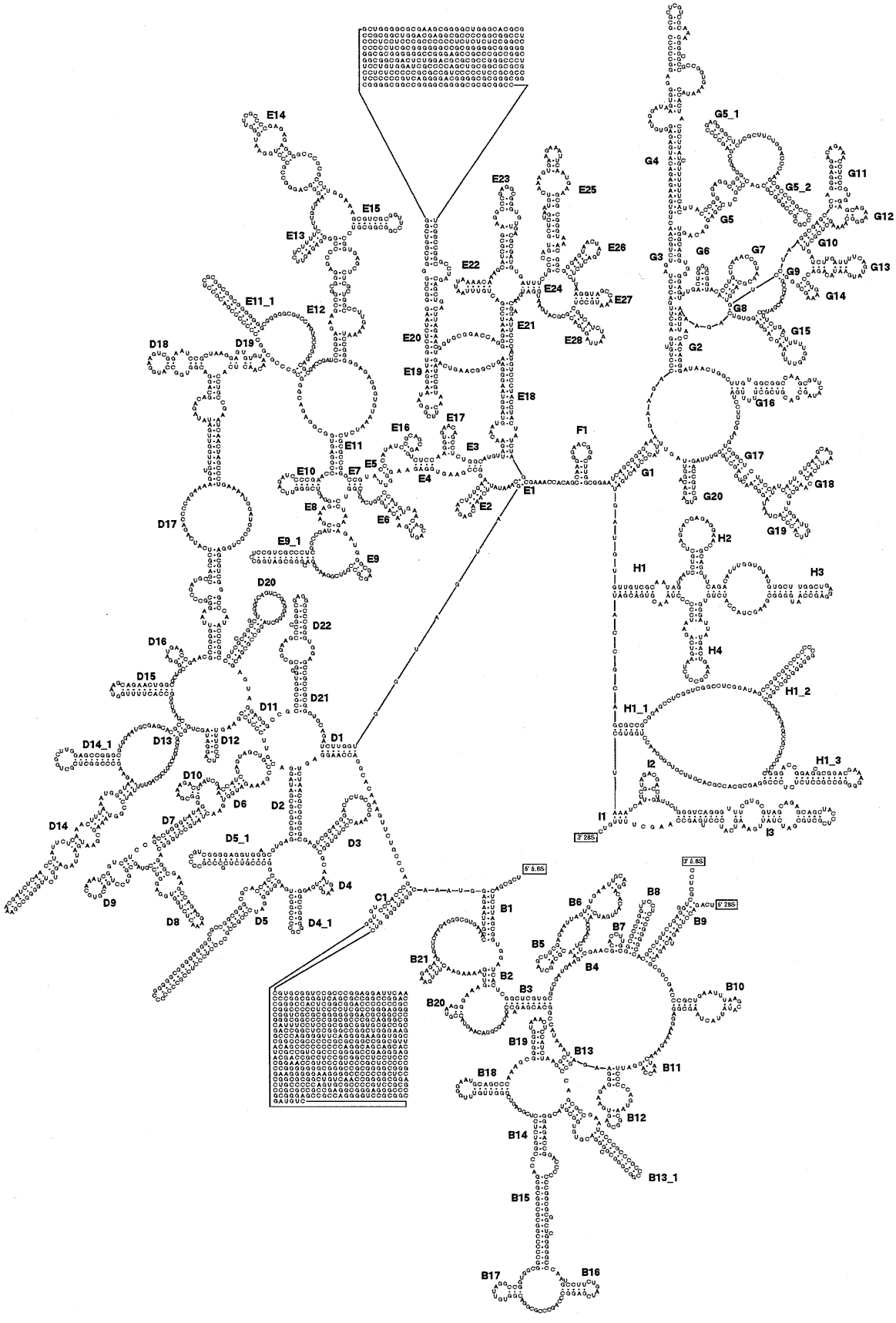

Figure 1. Secondary structure model of the large subunit ribosomal RNA of *Xenopus laevis*. The areas enclosed by helices C1 and E20 have been drawn unstructured.

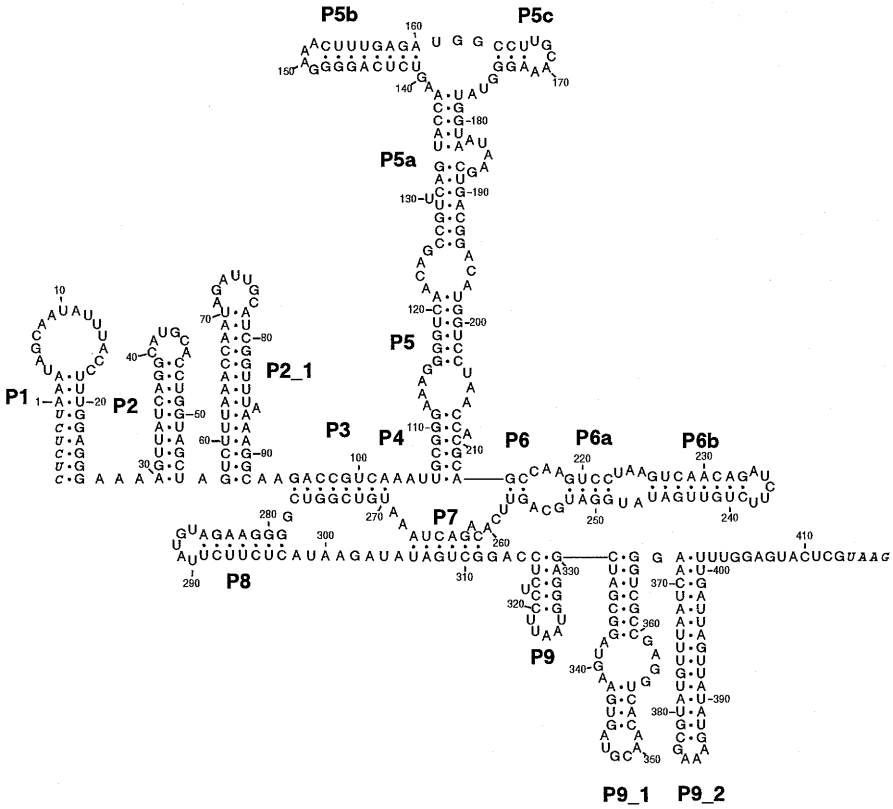

Figure 2. Drawing of the group I intron in the large subunit rRNA precursor of *Tetrahymena thermophila* (3). Helices P3 and P7 form a pseudoknot structure. Helices P7, P8, P9 P9_1 and P9_2 have been flipped in order to draw this structure properly. The bases drawn in italics are part of the exons bordering the intron.

RESULTS

The interface

RnaViz is intended to be easy to use, so the native look and feel of the operating system it is run on is largely followed. Therefore the interface of RnaViz will differ slightly on different platforms. In Figure 3 an example of the interface is given for the MS Windows 95 version. The largest part of the window is occupied by the display of the page containing the structures. The page can be displayed and edited at different zoom levels. A user customizable menu bar and pop up dialog boxes control the program, but customizable key shortcuts can be used throughout the program. A context sensitive help system can be invoked from the menu or the dialog boxes.

Files can be selected using a file selection box. Since RnaViz can contain several structures on one page, structures already on the page are not automatically deleted when a new file is opened. However, the page can be cleared before a new structure is loaded. Individual structures on the page can also be deleted.

The type of supported file formats is automatically detected. Opening a file in the RnaViz structure format will cause the structures in the file to be loaded directly onto the current page. When a DCSE alignment file or an mfold ct file is opened, the program will prompt for a skeleton file. If one is given, the program will produce drawings of the structure(s) in the file with the layout given in the skeleton file. If no skeleton is given, the structure drawings produced will probably contain overlapping areas. However, this can be fixed interactively. The program distribution contains examples and skeletons of several types of RNA molecules, e.g., tRNA, 5S, SSU and LSU rRNA and group I introns. When a DCSE or ct file contains more than one structure, one or more of these can be selected. The user can choose to either draw all selected structures on the current page, or to create several structure files, each containing a drawing of one of the selected structures.

Several user-definable parameters control how a newly created structure will be drawn. Among others, the general distance between bases in single stranded areas, the distance between bases in a helix and between the bases in a base pair can be set. By default, bases in a base pair are connected by a dot according to the IUPAC convention (15), but both width and length of the connections between the bases of standard and non-standard base pairs can be changed independently. The bases of non-standard base pairs can also be made to bulge slightly out of the helix. These settings can be changed for a drawn structure, but will only have effect when the structure or parts of it are redrawn. It is also possible to scale a structure.

Under MS Windows, the drawings can be printed directly using the standard Windows printer drivers, or exported to the clipboard for further processing in other packages. On Unix systems PostScript files are produced that can either be printed directly to a PostScript printer, or to other printers using GhostScript.

Arranging the layout of a structure

Each structure on the page consists of a number of individual objects, such as bases, base pair connections or helix names. An object can be selected by clicking on it using the first mouse button. Selecting an object will make the structure containing the

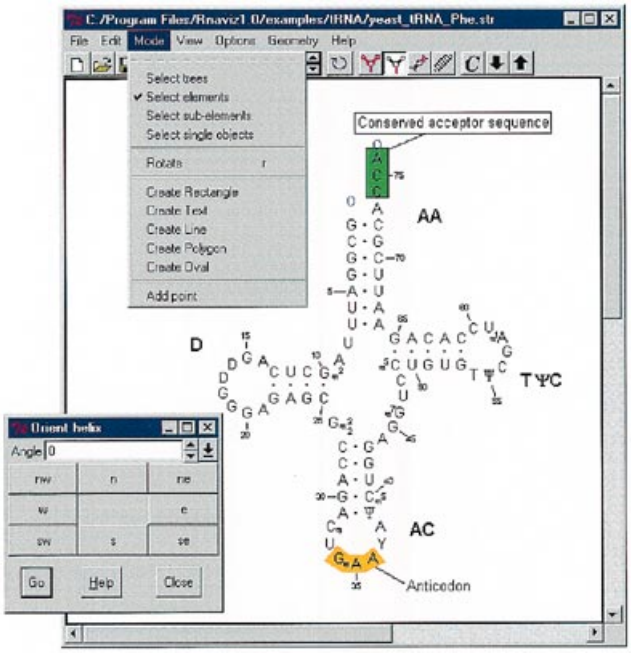

Figure 3. The interface of RnaViz running on MS Windows 95. All RnaViz functions can be invoked from the menu bar at the top of the Window. The window shows the structure of yeast phenylalanine tRNA. Several labels indicate base numbers and special areas in this molecule. The blue Os at the termini are used for the orientation of the terminal single strands, but do not appear in print. Contrary to the structures in Figures 1 and 4, the sequence is drawn anti clockwise, as is customary for tRNAs.

object the current structure. The current structure can be moved as a whole by clicking outside the structure and dragging. Extra objects can be added to or removed from the selection by clicking on them with the 'Control' key pressed.

Usually RnaViz is used in the 'element selection' mode. In this mode, clicking on a base that belongs to a helix will select the entire helix. A structure on the page can be rearranged quickly by clicking on a base or on the selection and dragging it to a different position. When the selection is released, the bases connecting the selection to the rest of the structure will be rearranged so as to maintain a correct structure drawing. If the distance between the bases reaches a given threshold, they will be automatically connected by a line. The threshold and the width of the line can be chosen by the user. The alternate mouse button can be used to select the apical portion of a helix, starting from the segment clicked on, rather than a complete helix. In DCSE files helix segments are defined as the parts of a helix separated by internal or bulge loops. Using the 'Select tree' entry in the 'Edit' menu or its key shortcut the entire area enclosed by the two strands of the selected helix can be added to the selection, e.g. the tree starting from helix D1 in Figure 1 contains all helices from D2 to D22. The selected part of a structure can be rotated by dragging with the 'Shift' key pressed. The center of rotation is indicated by a gray circle, and can be repositioned by clicking with the alternate mouse button while holding the Shift key. The selection can also be oriented into a specific direction using the 'Orient helix' or 'Orient' options in the 'Geometry' menu. When more than one helix is selected, the last helix added will determine the orientation. The 'Geometry' menu also offers options to easily straighten or bend the single stranded areas, or to flip a helix. The latter is often necessary to create a clear drawing of pseudoknots as shown in Figure 2. As illustrated in Figure 1, areas with unknown structure can be drawn unformatted.

In contrast to the previous mode, the 'single select' mode allows individual selection and positioning of objects. Only the selected objects will be moved, without redrawing the objects connecting them to the rest of the structure. This makes any special arrangement of objects possible. Other selection modes are the 'select tree' and 'select sub-element', which automatically select a tree or a set of segments of a helix.

Labeling a structure

All helices are automatically labeled with their helix name. This label will move together with the helix. The position of the label relative to the helix can be changed by selecting and dragging the helix name. Another type of label is the base numbering. Base numbers can be added automatically at a specified interval starting from a specific base. Base numbers can also be individually added or removed.

RnaViz also contains a limited drawing component. Several types of objects such as texts, rectangles, ovals, lines and polygons can be created and edited. Any of these objects can be used as a label by linking them to a certain base. An example of several types of labels is shown in Figure 3 of a tRNA.

Configuring objects

Each object has properties such as font, text, color, line width and position. The 'Configure Objects' dialog offers a versatile interface to change these properties for any object or groups of objects. Several parameters can be set that limit property changes to objects fulfilling certain criteria. This way the properties of either the currently selected objects, the objects of the current drawing or all objects can be changed. In addition, the changes can be limited to a specific type of objects such as bases, base pairings, base numbers, helix names or labels.

Tags are a very powerful feature; they can be used to indicate any special feature of specific bases, such as the variability of their position (16) or their use in a certain analysis. Since object configuration can be limited to objects with certain tags, it is easy to display such bases using different colors or fonts as demonstrated in Figure 4. A tag is actually a short text that is attached to an object. Every object can have several tags. The tags attached to an object can be changed individually by selecting the object and invoking the 'Edit tags' dialog. Tags can be added to or removed from groups of objects using the object configuration dialog. If the structure was created starting from a DCSE alignment that contained a line named 'mask' each base is tagged with the character on the corresponding position in the 'mask' line. It is also possible to add a list of tags to the sequence.

DISCUSSION

Producing clear, publication-ready 2D drawings of the secondary structure of RNA molecules is not a simple task. Most attempts have focused on producing a non-overlapping layout of the structure without user intervention. This focus has several drawbacks. The methods used are usually quite computing-intensive, limiting their use to smaller molecules and workstations. Another problem is the lack of post-production editing; the drawings produced cannot be easily changed or annotated. This makes it difficult to emphasize

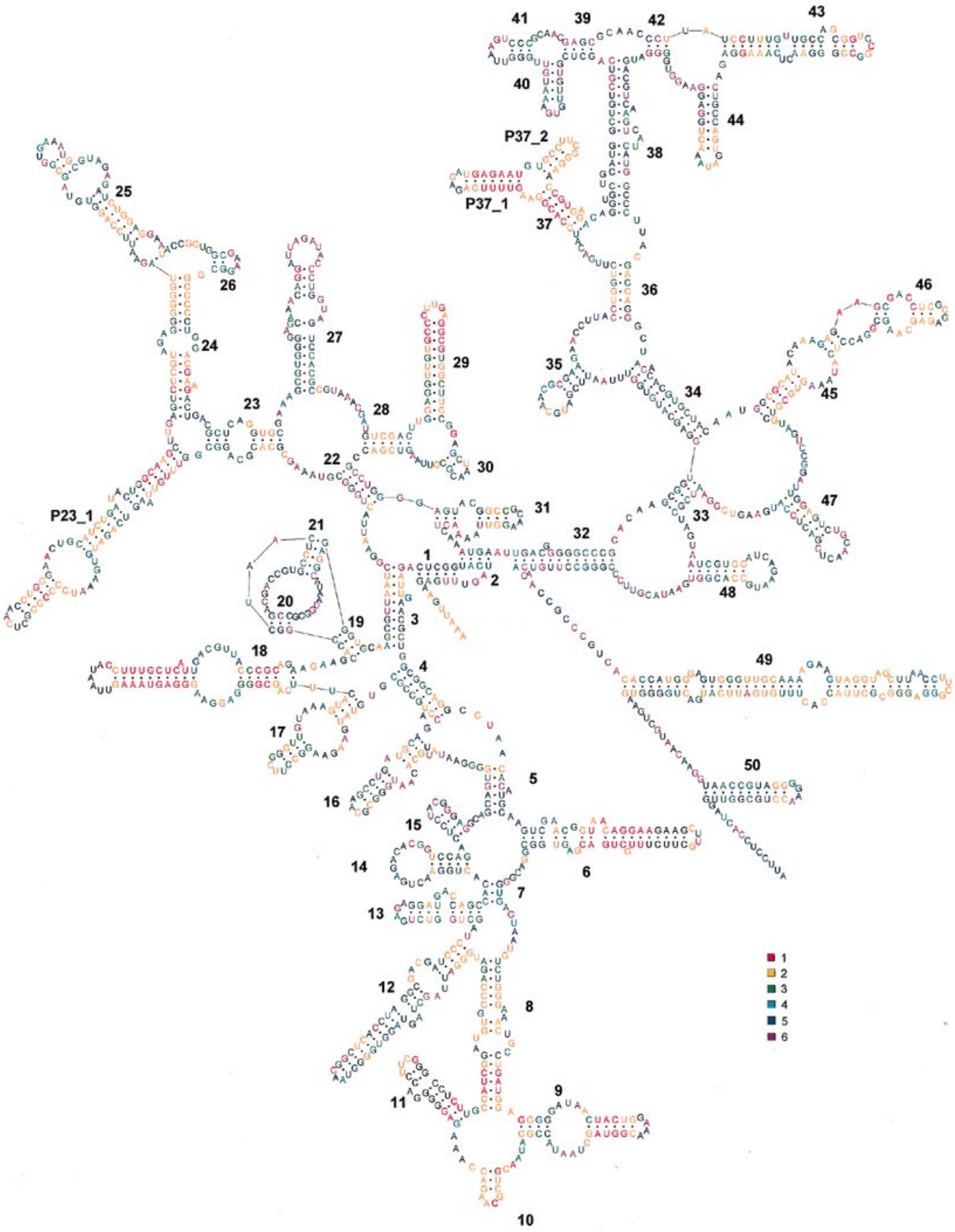

Figure 4. Secondary structure model of *Escherichia coli* SSU rRNA, where the variability of each position (16) is indicated by the color of the base according to the scale at the bottom of the page. The bases at the most variable positions are colored red, while those at the most conserved positions are indicated in purple. Black is used for positions where variability could not be measured.

structural similarities between different molecules, or to indicate peculiar areas. A final general problem is the inability to draw special structural features such as pseudo-knots or unformatted areas. A different direction was taken by the programs CARD (17) and XRNA (18), which give control to the user, at the expense of being very labor intensive. In CARD, the sequences for every structural element have to be typed in. XRNA does allow reading in the structure from a file, but arranging large structures is very laborious. Furthermore, the program is also only available for Unix workstations and difficult to use.

RnaViz solves these problems. Secondary structure drawings can be produced from data produced by other programs, without the need to enter the sequences from the keyboard. Rearrangement of the structure is straightforward, and several methods for annotating or labeling structures are present. The use of Tcl/Tk for the interface makes the program highly portable and extensible, thus more generally useful. In the future the algorithm for creating the initial layout could be improved, as drawings created without skeleton usually contain overlapping areas. However, the easy way structures can be rearranged and the use of skeleton files make this a minor issue.

Availability

RnaViz needs a modified version of Tcl and several extensions. Binary distributions of the modified Tcl and the RnaViz package are available for Linux and MS Windows 95 on the rRNA server at URL http://rrna.uia.ac.be. The sources are also available there for people who want to port the code to other systems. More information is present at the RnaViz home page (http://rrna.uia.ac.be/rnaviz/).

ACKNOWLEDGEMENTS

Our research was supported by the Fund for Scientific Research and by the Special Research Fund of the university of Antwerp. Peter De Rijk is a Research Assistant of the Fund for Scientific Research.

REFERENCES

1 Cech,T.R. (1988) *Gene* **73**, 259–271.
2 Michel,F., Umesono,K. and Ozeki,H. (1989) *Gene* **82**, 5–30.
3 Dahlberg,A.E. (1989) *Cell* **57**, 525–529.
4 Pace,N.R., Smith,D.K., Olsen G.J. and James,B.D. (1989) *Gene* **82**, 65–75.
5 Jacobson,A.B. and Zuker,M. (1993) *J. Mol. Biol.* **233**, 261–269.
6 De Rijk,P. and De Wachter,R. (1993) *Comput. Appl. Biosci*. **9**, 735–740.
7 Cedergren,R., Gautheret,D., Lapalme,G. and Major,F. (1988) *Comput. Appl. Biosci.* **4**, 143–146.
8 Gautheret,D., Major,F. and Cedergren,R. (1990) *Methods Enzymol.* **183**, 318–330.
9 Martinez,H.M. (1988) *Nucleic Acids Res.* **16**, 1789–1798.
10 Muller,G., Gaspin,C., Etienne,A. and Westhof E. (1993) *Comput. Appl. Biosci.* **9**, 551–561.
11 Shapiro,B.A., Maizel,J., Lipkin,L.E., Currey,K. and Whitney C. (1984) *Nucleic Acids Res.* **12**, 75–88.
12 Yamamoto,K., Sakurai,N. and Yoshikura H. (1987) *Comput. Appl. Biosci.* **3**, 99–103.
13 Kernighan,B. and Ritchie,D. (1988) *The C Programming Language*, 2nd Edn, Prentice Hall, Englewood Cliffs, NJ, USA.
14 Ousterhout,J. (1994) *Tcl and the Tk Toolkit*, Addison-Wesley, Reading, MA, USA.
15 IUPAC-IUB commision on biochemical nomenclature (1970) *Eur. J. Biochem.* **15**, 203–208.
16 Van de Peer,Y., Chapelle,S. and De Wachter,R. (1996) *Nucleic Acids Res.* **24**, 3381–3391.
17 Winnepenninckx,B., Van de Peer,Y., Backeljau,T. and De Wachter R. (1995) *BioTechniques* **18**, 1060–1063.
18 Weiser,B. and Noller,H.F. University of California, Santa Cruz, CA, USA.