Published in final edited form as:

Curr Protoc Bioinformatics. 2006 March ; 0 12: Unit–12.6. doi:10.1002/0471250953.bi1206s13.

RNA Secondary Structure Analysis Using RNAstructure

David H. Mathews

Department of Biochemistry & Biophysics and Center for RNA Biology University of Rochester
Medical Center 601 Elmwood Avenue, Box 712 Rochester, New York 14642 Phone:
585-275-1734 FAX: 585-275-6007 David_Mathews@urmc.rochester.edu

Abstract

RNAstructure is a user-friendly program for the prediction and analysis of RNA secondary structure. It is available as a web server, as a program with a graphical user interface, or as a set of command line tools. The programs are available for Microsoft Windows, Macintosh OS X, or Linux. This unit provides protocols for RNA secondary structure prediction (using the web server or the graphical user interface) and prediction of high affinity oligonucleotide binding sites to a structured RNA target (using the graphical user interface).

Keywords

RNA Secondary Structure Prediction; Free Energy Minimization; Thermodynamics

RNAstructure is a user-friendly program for the prediction and analysis of RNA secondary structure (Bellaousov et al., 2013; Reuter and Mathews, 2010). It is available for free and can be run using a web server or downloaded to run locally on Windows, Macintosh OS X, or Linux. The program includes several algorithms, including secondary structure prediction by free energy minimization or maximum expected accuracy structure prediction (Lu et al., 2009; Mathews et al., 2004), a partition function for predicting base pair probabilities (Mathews, 2004), ProbKnot for predicting structures including pseudoknots (Bellaousov and Mathews, 2010), stochastic sampling from the Boltzmann ensemble (Ding and Lawrence, 2003), OligoWalk for predicting binding affinity of oligonucleotides to a complementary RNA target (Lu and Mathews, 2007; Lu and Mathews, 2008a; Lu and Mathews, 2008b; Mathews et al., 1999a), methods for predicting the structure of interacting sequences (Piekna-Przybylska et al., 2009), and methods for predicting conserved structures common to two or more sequences (Harmanci et al., 2007; Harmanci et al., 2008; Harmanci et al., 2009; Harmanci et al., 2011; Mathews, 2005; Mathews and Turner, 2002; Ustilov et al., 2006; Xu and Mathews, 2011).

Basic Protocol 1 provides instruction for predicting RNA secondary structure with the RNAstructure web server. Alternative protocol 1 provides instructions for using the graphical interface to predict lowest free energy structures and base pairing probabilities. An

Internet Resources

<http://rna.urmc.rochester.edu> is the location of the Mathews lab website, where RNAstructure can be downloaded and the RNAstructure web servers can be accessed.

example is provided for both, using a tRNA sequence (Sprinzl and Vassilenko, 2005). Basic Protocol 2 covers the use of OligoWalk to predict binding affinities of complementary oligonucleotides to an RNA target. The same tRNA sequence is used as an example. Predicting a conserved secondary structure common to two or more sequences is covered in unit 12.4 of this series.

RNA secondary structure prediction is available in other software packages. The Vienna RNA package can be used to predict secondary structures in either a Unix or Windows environment and is covered in unit 12.2 (Hofacker et al., 1994). The use of the mfold web server for secondary structure prediction is covered in *Current Protocols in Nucleic Acid Chemistry* in unit 11.2 (Zuker et al., 1999). The commentary at the end of the unit compares the available packages.

Basic Protocol 1: Predicting RNA Secondary Structure Using the RNAstructure Web Server

This protocol details the use of the RNAstructure webserver to predict an RNA secondary structure. It assumes basic familiarity with using the World Wide Web and web browsers. An example is provided with a tRNA sequence (Sprinzl and Vassilenko, 2005).

Necessary Resources

Software—A web browser is required for accessing the RNAstructure web servers.

Connect to the web server and submit sequences: 1. Point a browser to <http://rna.urmc.rochester.edu/RNAstructureWeb/>, the RNAstructure web server.

2. Choose the link “Predict a Secondary Structure”.

The RNAstructure web servers are designed around two types of themes. The first is biological problem and the second is program. The “Predict a Secondary Structure” server is problem-themed server, and it runs four programs, Fold (a structure prediction program that finds lowest free energy structures) (Mathews et al., 1999b), partition (a program that predicts base pairing probabilities) (Mathews, 2004), MaxExpect (a maximum expected accuracy structure prediction method) (Lu et al., 2009), and ProbKnot (a program that can predict structures with pseudoknots) (Bellaousov and Mathews, 2010). Farther down the main server page, Fold, partition, MaxExpect, and ProbKnot servers can be chosen. The problem-themed servers are generally more popular and are convenient to use because they provide alternative hypotheses for the structure. The program servers, however, complete their calculations in less time.

Enter the sequence

3. Sequences are either uploaded using a FASTA-formatted text file (Figure 12.6.1) or by pasting the sequence into the browser. Figure 12.6.2 shows a screen shot of the web server form.

To upload a file, click browse and choose a filename to upload. To paste a sequence, enter a title for the sequence in the Sequence Title field and paste the sequence in the Sequence field. The Sequence field should contain only A, C, G, U, T, and X. T and U are equivalent. X is a nucleotide that cannot pair or stack.

An example sequence can be displayed by clicking on the link directly above the text box labeled “click here to add example sequences to the box.” This pastes the tRNA sequence RA7680 (Sprinzl and Vassilenko, 2005) into the form. This sequence is used for the example shown here.

For the webserver, there is a limit to the length of sequences. The limit is detailed at: <http://rna.urmc.rochester.edu/RNAstructureWeb/Information/Limitations.html>. As of this writing, sequences must be 2,500 nucleotides or shorter. The limit is designed to provide users with a reasonable rate of throughput, and it might change as dictated by server demand or as hardware is upgraded.

Note that lowercase nucleotides are not allowed to pair in structure prediction. It is therefore important that most nucleotides be uppercase. For the tRNA sequence used in this example, lowercase nucleotides are modified nucleotides that cannot be accommodated in a helix (Mathews et al., 1999b). [*Figures 1 and 2 near here] 4. Select the nucleic acid backbone. RNA is the default, and this will treat any T in the sequence as uracil. DNA can also be chosen, and this will treat any U in the sequence as thymine.

The stability of forming a specific secondary structure is estimated using a nearest neighbor model. Parameters are available for RNA (Lu et al., 2006; Mathews et al., 2004; Xia et al., 1998) and for DNA (Reuter and Mathews, 2010). Because the stabilities are backbone-specific, the predicted structures for RNA and DNA of the same sequence (but differing in the use of uracil and thymine), are often different.

Select parameters

5. The first adjustable option is the absolute temperature for structure prediction. By default, structure prediction is performed for folding at 37 °C, i.e. 310.15 K.

The nearest neighbor parameters are most accurate at 310.15 K. For RNA, they are known to be accurate between 293 and 333 K (Lu et al., 2006). For most calculations, it is reasonable to choose the default temperature.

6. Next, the maximum loop size can be set.

To reduce the calculation time, the maximum size of internal and bulge loops can be limited. Traditionally, the limit has been set at 30 unpaired nucleotides. It is unlikely that structure prediction for a biologically relevant sequence would require a change to this parameter.

7. Parameters can now be adjusted to control the production of suboptimal structures by Fold (MFE) and MaxExpect (MEA). These include the Maximum % Energy Difference, Maximum Number of Structures, and Window Size.

For most calculations, these parameters can be kept at their default settings. They control the number and diversity of suboptimal structures, which serve as alternative hypotheses for the predicted structure. The default parameters were chosen to provide a small set of diverse structures that can be manually reviewed. Maximum % Energy Difference and Maximum Number of Structure place limits on the total number of suboptimal structures, more structures can be generated by setting these parameters to larger values. Window Size ensures that structures are substantially different. To generate more structures that are more similar, a smaller integer (as low as 0) can be used. Setting Window to a larger integer will generate fewer, more dissimilar structures.

A complete explanation of the parameters is available through the online help, which can be reached by clicking the link at “If you need specific help using the Predict a Secondary Structure server, please click here.” This appears immediately above the sequence entry portion of the web form.

8. Parameters that alter the functioning of the MaxExpect and ProbKnot algorithms can also be changed.

For most calculations, the default parameters are the best choice.

Select optional data

9. Constraints on the folding can be uploaded as plain text files (Figure 12.6.3 for the file format). To upload a restraint file, click Browse at “Select Folding Constraints File.”

Constraints include information that can be gleaned by enzymatic probing of structure, including nucleotides that cannot pair and nucleotides that must pair (Knapp, 1989). Constraints can also be determined by chemical modification probing of structure, which determines nucleotides that cannot be buried in a helix (Ehresmann et al., 1987). These forms of constraints have both been shown to improve the accuracy of structure prediction (Mathews et al., 2004; Mathews et al., 1999b). Other constraints include specifying Uridines in GU base pairs, pairs that must occur, and specific pairs that cannot occur.

10. SHAPE mapping data can also be uploaded to restrain structure prediction (Merino et al., 2005). To upload the data, click the Browse button at “Select SHAPE Constraints File.” The data are uploaded in a plain text file, and the format is specified in Figure 12.6.4. The SHAPE Intercept and SHAPE Slope are used to interpret SHAPE mapping data, and these parameters can be changed.

SHAPE data can dramatically increase the accuracy of structure prediction (Deigan et al., 2009). The SHAPE Intercept and SHAPE Slope default values were shown to provide the best accuracy (Hajdin et al., 2013). For most calculations, these should be kept at the default values.

11. The base pair probabilities are shown in a probability dot plot. The range of $-\log_{10}$ of pairing probability to be displayed can be set.

The default values can be used for most calculations. By default, the Minimum is blank, meaning that there is no minimum. The Maximum defaults to 2, which means only base pairs of greater than 1% pairing probability will be displayed.

Enter an email address and start the calculation

12. Optionally, an email address can be provided. If an email address is provided, email is sent when the calculation is complete, and this email contains a link to the location of the results. If an email address is not provided, it is imperative that the browser window remain open to the web server after the calculation is started to make sure the results remain accessible. 13. Start the calculation by clicking “Submit Query.” The web server will now move to the results page. This page refreshes automatically until the calculation is complete, and the results can be displayed.

Examine and download the results

14. On the results page, the results are displayed for Fold, MaxExpect, ProbKnot, and partition from top to bottom on the page. A screen shot of the output page is shown in Figure 12.6.5.

The three structure prediction methods can predict different results, which should be interpreted as alternative hypotheses for the structure. The Fold prediction is predicted to be the single most likely structure at equilibrium. MaxExpect assembles structures of highly probably pairs. On average, it makes fewer false predictions of pairs than Fold (Lu et al., 2009). ProbKnot is able to predict structures that contain pseudoknots (Bellaousov and Mathews, 2010). It, however, is prone to predicting pairs that are incorrect. In the example, a tRNA, the optimal predicted structures for all programs are the same.

15. Predicted structures are displayed for each program. Using the web example sequence, a screen shot of the output is shown in Figure 12.6.5.

For structure display, an SVG graphic is displayed to indicate nucleotides in canonical base pairs. When there are alternative structures, called suboptimal structures, buttons labeled “Previous” and “Next” are available to change the currently displayed structure. Immediately beneath the structure display, all structures are available for download as Adobe pdf, Adobe Postscript (PS), or ct format, which is a plan text format that indicates the locations of base pairs (Figure 12.6.6). Then, individual structures are available for download as Adobe pdf, SVG, Adobe Postscript (PS), jpeg, or ct format.

Structures are colored according to probabilities. For base paired nucleotides, the probabilities are the probability of being in the specific base pair that is drawn. For unpaired nucleotides, probabilities are the probability of being unpaired. The color annotation key in the panel details the bands of pairing probability. The highest probabilities are red (99%), then orange (99% > probability 95%), yellow (95% > probability 90%), dark green (90% > probability 80%), light green (80% > probability 70%), light blue (70% > probability 60%), dark blue (60% > probability 50%), and purple (50%).

Color annotation provides information about the confidence in the prediction of a specific pair. It has been demonstrated that more probable pairs are more likely to be correctly predicted (Mathews, 2004).

Adobe pdf and Adobe Postscript are vector formats, and can therefore be edited using drawing tools, such as Inkscape or Adobe Illustrator. SVG is another vector graphic, but is generally designed for web display. Jpeg is an image format, and it cannot be easily modified or scaled.

For Fold, the ENERGY value displayed at the bottom of the drawing is the predicted free energy change in kcal/mol. For MaxExpect, the ENERGY value is a score for the structure. With the default parameters, the score is equal to twice the sum of the pair probabilities for all pairs, plus the probability for being unpaired for all single-stranded nucleotides. The ProbKnot structure is not scored.

16. Also provided after each program is the command line that was executed on the server to run the calculation. Clicking the name of the program links to the help pages for the command line program. This information is useful for learning how to run the programs on the command line.

17. A probability dot plot is also provided to summarize the base pairing probabilities (Figure 12.6.7). All pairs within the user-specified pairing probability range appear in this plot. This range defaults to 1% or higher pairing probability (see step 11).

In the probability dot plot, the x and y axes are nucleotide index. A dot indicates two nucleotides that can pair, and the color of the dot indicates the pairing probability. For example, for RA7680, the red dot in the upper right-hand corner indicates that nucleotides 1 and 72 are predicted to pair with probability between 96.5% ($-\log_{10}$ of 0.0154) and 38.7% ($-\log_{10}$ of 0.412). From highest to lowest probability, dots are red, dark green, light green, aqua, and blue. The plot can be downloaded as SVG, jpeg, Adobe PDF, Adobe PS, or as a binary save file (PFS), which can be used by other RNAstructure programs.

Alternate Protocol 1: Predicting Secondary Structure and Predicting Base Pair Probabilities with the RNAstructure Graphical User Interface

This protocol describes the basic use of the RNAstructure graphical user interface to predict a secondary structure by free energy minimization. Many other options are available, and these are described in detail in the online help manual, which can be accessed from the program by choosing the “Help Topics” item from the “Help” menu option. For example, secondary structures could also be predicted by Maximizing Expected Accuracy (MaxExpect) or by ProbKnot. These structures could provide alternative hypotheses for the structure. This protocol requires basic familiarity with point-and-click interfaces.

Necessary Resources

The software package, RNAstructure, can be downloaded from the World Wide Web at <http://rna.urmc.rochester.edu/RNAstructure.html>. Registration is required for download, so

that a count of those using the software can be maintained. The list of registered users is not shared with others and not used for any other purpose.

The RNAstructure graphical user interface (GUI) is available for Windows, both 32 and 64 bit, and is known to run on Windows XP, Windows Vista, Windows 7, and Windows 8. The graphical interface is also available in Java for use on Macintosh OS X (Leopard or later) or Linux, in 32 or 64 bit.

Download and install RNAstructure

1. Download RNAstructure from <http://rna.urmc.rochester.edu/RNAstructure.html>.

2a. For Windows, download either the 32 bit Windows Native Interface or the 64 bit Windows Native Interface. If in doubt as to whether the version of Windows being used is 32 bit or 64 bit, download the 32 bit because this will work in both environments. The 64 bit version, run in a 64 bit environment, is capable of predicting structures for longer sequences than the 32 bit version because it can address more memory. Double-click on the zip file (RNAstructure.zip or RNAstructure64bit.zip). Then, double-click on setup.exe and install the software. RNAstructure can then be run from the start list or from the start screen on Windows 8.

2b. For Macintosh OS X, download the Mac OS X interface by clicking the link, “JAVA Mac OS-X Interface as tarball.” Double-click RNAstructureForMac.tgz to extract the files. Double-click on the RNAstructure directory. RNAstructure can now be launched by double-clicking the RNAstructure icon. If desired, the whole RNAstructure directory can be dragged to the application folder.

Note that on OS X Mountain Lion or later, an error message may appear that states the executable is “damaged.” This is a feature that prevents software downloaded from the internet from running. To run RNAstructure, click on “system Properties.” Under “Security & Privacy,” click the lock to be able to make changes and select “Anywhere” under “Allow applications downloaded from...” Then when RNAstructure is run, a prompt will appear to query if it is OK to run RNAstructure.

2c. For Linux, download either “JAVA 32-bit Linux Interface as tarball” or “JAVA 64-bit Linux Interface as tarball.” If in doubt as to whether the operating system is 32 bit or 64 bit, download the 32 bit because this will work in both environments. The 64 bit version run in a 64 bit environment is capable of predicting structures for longer sequences because it can address more memory. Extract the files using tar -xzf RNAstructureForLinux.tgz or tar -xzf RNAstructureForLinux64bit.tgz. RNAstructure can now be started by executing RNAstructureScript, found in RNAstructure/exe/, which sets environment variables and launches the JAVA virtual machine.

Note that RNAstructure requires the Oracle JAVA, which can be found at: http://java.com/en/download/linux_manual.jsp.

3. Online help for installing and launching RNAstructure is available online at: <http://rna.urmc.rochester.edu/Overview/index.html>.

Enter the sequences

4. Start the RNAstructure GUI.

5. To enter sequences in RNAstructure, use the sequence editor, which can be opened either by selecting “New Sequence” from the “File” menu or by clicking the “New Sequence” icon at the far left of the toolbar. The sequence editor will open as shown in Figure 12.6.8.

Note that items in the toolbar, located directly under the menu bar on Windows and Linux and located at the top of the window on OS X, identify themselves with a pop-up label if the mouse pointer is placed over that icon.

6. Enter a title in the Title field at the top. This title will be used to label the output from calculations. The Comment field provides an opportunity to save comments with the sequence, but can be left blank. Finally, enter the sequence, from 5' to 3', in the Sequence field.

The sequence should consist of A (adenine), C (cytidine), G (guanine), U (uridine), T (treated as uridine in RNA folding), and X (a nucleotide that neither base pair nor stack). Note that nucleotides entered as lower case characters are forced single stranded in the structure prediction and therefore most nucleotides should be upper case. Spaces and carriage returns are ignored during structure prediction. Sequences can be entered manually or pasted from other programs by cut and paste. A sequence copied to the clipboard can be pasted into the Sequence field by first clicking in the Sequence field to place the cursor and then choosing Paste from the Edit menu item or typing Ctrl-V.

Several tools are available in the sequence editor. Clicking the Format Sequence button places the sequence in columns of five nucleotides with fifty nucleotides per line. If the Read Sequence button is clicked, the sequence is read aloud (over the computer's speakers) from 5' to 3'. This can be canceled by clicking the same button, which has its label changed to Cancel Read. Also, the sequence can be read while it is typed by choosing Read Sequence While Typing on the Read menu item. The secondary structure can be predicted for a single sequence by clicking Fold as RNA or Fold as DNA buttons.

The example structure prediction illustrated in this unit is for tRNA sequence RA7680 (Sprinzl and Vassilenko, 2005). This sequence is provided with RNAstructure as an example, and therefore does not need to be entered. On Microsoft Windows, the default location for the examples is the user's documents folder, e.g. C:\Users\dhm\Documents\RNAstructure_examples. On Linux and Macintosh, the files are in the examples folder in the RNAstructure folder.

7. Save the sequence either by choosing Save Sequence under the File menu item or by clicking the disk icon on the toolbar. This sequence can later be accessed by opening the file with the sequence editor, either by choosing Open Sequence under the File menu item or by clicking the Open Sequence icon on the toolbar.

Predict the secondary structure

8. After saving the inputted sequence (above), next choose the “Fold RNA Single Strand” under the “RNA” menu option. The RNA secondary structure prediction form will open as shown in Figure 12.6.9.

9. First choose the name of the sequence file by clicking the “Sequence File” button. This will open a standard open file dialog for selecting the file. After the file has been selected, the name of the file will appear next to the “Sequence File” button and default values will have been entered in all other fields. The output of the calculation (a predicted secondary structure and a set of low free energy structures) will be stored in a ct (connection table) file (Figure 12.6.6). The default ct file name is the name of the sequence file chosen, but with a .ct file extension. The default name can be changed by clicking the “CT File” button.

10. Three parameters control the generation of suboptimal structures, Max % energy Difference, Max Number of Structures, and Window Size. Suboptimal structures are low free energy structures that represent alternative hypotheses for the structure. The parameters have been chosen to provide a diverse set of alternative structures. The default parameter values will work well for most calculations.

Max % energy Difference is the maximum percent energy difference of suboptimal structures as compared to the lowest free energy structure. Larger percent energy differences allow the prediction of more suboptimal structures, whereas zero allows only prediction of the lowest free energy structure. The maximum percent energy difference can be changed from the default by manually typing into the text box adjacent to “Max % Energy Difference.” The second parameter, Max Number of Structures, is the maximum number of structures. This places an absolute limit on the number of suboptimal structures and can be changed manually by typing in the text box adjacent to “Max Number of Structures.” The third parameter is the Window Size. The window size specifies how different each suboptimal structure must be as compared to all other predicted structures. A window size of zero does not place any restriction. Larger window sizes result in structures with greater difference, but also result in fewer predicted structures. The default window parameter can be manually changed by typing an integer in the text box adjacent to “Window Size.”

11. The checkbox next to “Generate Save File” is checked by default. This will generate a save file that can be used to show energy dot plots or to predict a different set of suboptimal structures using different suboptimal structure parameters. The online help contains entries labeled “Dot Plot” and “Refolding a Saved Sequence” that explain these functions. The name of the save file is the same as the ct file, except that the file has a .sav extension instead of .ct.

12. RNAstructure can predict secondary structures with user-specified constraints. These are entered by choosing the appropriate menu item under the “Force” menu option. “Base Pair” is used to specify required base pairs in the structure. “Chemical Modification” is used to specify nucleotides that are accessible to chemical modification, i.e. single stranded, at the end of a helix, or in or adjacent to a GU pair (Ehresmann et al., 1987; Mathews et al., 2004). “Double Stranded” is used to specify nucleotides that must be base paired, without

specifying to which nucleotides they are paired. “FMN Cleavage” is used to indicate Us that are in GU base pairs. “Single Stranded” is used to indicate unpaired nucleotides. “Prohibit Basepairs” is used to indicate specific base pairs that are not allowed. Each of these options opens a dialog box for entering the specified constraints. In the dialog box, “OK” can be clicked to keep the dialog box open to enter more constraints, “Cancel” can be clicked to not record the constraints, or “OK and Close” can be clicked when all constraints are entered into the dialog box. The entered constraints are displayed on the screen if “Current” is chosen under the “Force” menu item. “Reset” removes all entered constraints. “Save Constraints” can be used to save all constraints to a file (.con file). Constraints can be read from a file by choosing “Restore Constraints.”

Nucleotides can be determined to be paired or unpaired by enzymatic mapping (Knapp, 1989). Both enzymatic mapping and chemical modification data can improve the accuracy of RNA secondary structure prediction (Mathews et al., 2004; Mathews et al., 1999b).

13. RNA secondary structure prediction can also be restrained using data from SHAPE mapping (Merino et al., 2005). The data are read from a plain text file (Figure 12.6.4) using the “Read SHAPE Reactivity – Pseudo-Energy Constraints” option under the “Force” menu item. This opens a dialog box from which the SHAPE data file name can be specified by clicking the button “SHAPE Datafile.” The Slope and Intercept parameters are used to translate the data into folding restraints, and the default values are best for most calculations. SHAPE data improve the accuracy of RNA secondary structure prediction (Deigan et al., 2009).

14. The temperature for folding can be changed by selecting the “Temperature” menu item. Temperatures are entered in K, i.e. in absolute temperature. By default, structure prediction is performed for folding at 37 °C, i.e. 310.15 K. The nearest neighbor parameters are most accurate at 310.15 K. For RNA, they are known to be accurate between 293 and 333 K (Lu et al., 2006). For most calculations, it is reasonable to choose the default temperature.

15. Click the button labeled “Start” to begin the secondary structure prediction calculation. A window will open with a progress indicator. When the calculation ends, a dialog box opens with the options “Draw Structures” and “Cancel.” Click “Draw Structures” to display the predicted secondary structures.

Predict base pair probabilities with a partition function calculation

16. Open the partition function window by choosing “Partition Function RNA” under the “RNA” menu option. Choose the sequence name by clicking the “Sequence Name” button. For this example, select RA7680. After the calculation, the base pair probability data are saved to disk in a save (.pfs) file. A default save file name automatically appears in the field next to the “Save File” button after a sequence has been chosen. The default name can be changed by clicking the “Save File” button.

17. Start the calculation by clicking the button labeled “Start.” A window will open to show the progress of the calculation. When it is completed, a probability dot plot will be displayed that indicates the probability of all valid canonical (AU, GC, and GU) base pairs. First,

reduce the probability range of base pairs, by choosing “Plot Range” under the “Draw” menu option. This will open a dialog box in which a min and max range are entered. Dots are registered by $-\log_{10}$ (Base Pair Probability). Set the maximum to 2, so that all base pairs with pairing probability greater than 0.01 (1%) will be displayed. Click “OK.” The plot will now resemble the plot in Figure 12.6.10.

The probability dot plot window provides several features for analyzing the dot plot information. By clicking on a dot, the message window at the bottom of RNAstructure provides the identity of the base pair and $-\log_{10}$ of the base pair probability. The size of the plot in the window can be zoomed by choosing “Zoom” under the “Draw” menu option. Alternatively, pressing the control and left arrow keys zooms out and pressing the control and right arrow keys zooms in. The base pair probabilities can be written to a tab-delimited text file for analysis with other programs by choosing “Output Text File” under the “Output” menu option. The text file contains the $-\log_{10}$ of base pair probability for all base pairs and not just the pairs that are currently displayed on the screen. Secondary structures composed of only probable base pairs can be outputted by choosing “Output Probable Structure” under the “Output” menu option.

The probability dot plot window is created from the data stored in the save (.pfs) file. To draw the probability dot plot again at a later time, choose “Dot Plot Partition Function” under the “File” menu option. This will open an open file dialog box with which the save file can be chosen.

Color annotate the predicted secondary structures according to pairing probabilities

18. Return to the structure drawing window that contains the predicted secondary structures for the tRNA sequence. If the window has been closed, the secondary structure can be redrawn using the data stored in the ct file. A new drawing window can be opened by choosing “Draw” under the “File” menu option. An open file dialog box will appear, from which the ct file can be chosen.

19. To color annotate a predicted secondary structure for which a partition function calculation has also been performed, choose “Add Color Annotation” under the “Annotations” menu option with the drawing window open. This will launch an open file dialog box from which the save (.pfs) file can be selected. The secondary structure will then have color annotation. Nucleotides in base pairs are colored according to the probability that the drawn pair is formed. Unpaired nucleotides are colored with the probability that the nucleotide is unpaired. The most probable items are in red and the least probable are in violet. To show a key that indicates the association between color and probability range, choose “Show Color Annotation Key” under the “Annotations” menu option. Figure 12.6.11 shows the predicted lowest free energy structure for the RA7680 tRNA with color annotation.

The drawing window has several functions to facilitate the analysis of secondary structures. When suboptimal structures are included in the ct file, the displayed structure can be changed by choosing “Structure Number” under the “Draw” menu option. This will open a window that indicates the current structure number and allows that number to be changed.

The lowest free energy structure is structure 1 and folding free energy increases with the structure number. Alternatively, pressing the control and up arrow keys increases the number of the currently displayed structure and pressing the control and down arrow keys lowers the number of the currently displayed structure. The number of the currently displayed structure is indicated in the upper left hand corner of the window. As indicated in Figure 12.6.11, there are a total of four low free energy secondary structures predicted for the RA7680 tRNA using the default parameters. The size of the structures on the screen can be zoomed in and out by choosing “Zoom” under the “Draw” menu option. Alternatively, pressing the control and left arrow keys zooms out and pressing the control and right arrow keys zooms in.

Base pairing probabilities are an indication of the quality of a predicted pair. Highly probably pairs are more likely to be correctly predicted than low probability pairs (Mathews, 2004).

The drawing window also shows the predicted folding free energy change for each structure. This is provided in the drawing window as the ENERGY, and this is in units of kcal/mol.

Basic Protocol 2: Predicting Binding Affinities of Oligonucleotides Complementary to an RNA Target with OligoWalk

RNAstructure includes the OligoWalk program for predicting the binding affinity of complementary oligonucleotides to an RNA target (Mathews et al., 1999a). For an RNA sequence of N nucleotides, OligoWalk predicts an overall free energy change of binding of all N-L+1 oligonucleotides of length L that are complementary. Hence, the binding region is walked down the length of the sequence. The overall free energy change of binding, G_{37}° overall includes the effects of self-structure in the target and self-structure in the oligonucleotides.

Necessary Resources

The software package, RNAstructure, can be downloaded from the World Wide Web at <http://rna.urmc.rochester.edu/RNAstructure.html>. Registration is required for download, so that a count of those using the software can be maintained. The list of registered users is not shared with others and not used for any other purpose.

The RNAstructure graphical user interface (GUI) is available for Windows, both 32 and 64 bit, and is known to run on Windows XP, Windows Vista, Windows 7, and Windows 8. The graphical interface is also available in Java for use on Macintosh OS X (Leopard or later) or Linux, in 32 or 64 bit.

Install RNAstructure

1. Follow steps 1 to 3 in Alternative Protocol 1 to install the RNAstructure graphical user interface (GUI).

Enter the sequence and predict the secondary structure

2. Follow steps 4 to 7 in Alternative Protocol 1 to enter the sequence of the target RNA.
3. Follow steps 8 to 15 in Alternative Protocol 1 to predict the secondary structure of the target RNA.

Start the OligoWalk calculation

4. Open the OligoWalk input window (Figure 12.6.12) by choosing “OligoWalk” under the “RNA” menu option. First click on the button labeled “CT File” and choose the target secondary structure, stored in ct format, using the open file dialog box. A default name is then chosen for output of the thermodynamic estimates. This name is the same as the ct file, but with the .ct extension changed to .rep. The default name can be changed by clicking on the button labeled “Report File.” The report file stores the calculated parameters as tab-delimited text that can be opened in most spreadsheet programs, such as OpenOffice or Microsoft Excel. [*Figure 12 near here]
5. Next choose a Mode for the calculation by clicking the button adjacent to one of the three options: “Break Local Structure,” “Refold Whole RNA for Each Sequence,” or “Do Not Consider Target Structure.” “Do Not Consider Target Structure” is the fastest mode, but is not recommended because the target RNA secondary structure is neglected. “Refold Whole RNA for Each Sequence” predicts a new lowest free energy structure after oligonucleotide binding by predicting a structure with the nucleotides that are bound by the oligonucleotide forced to be unpaired. This mode is the slowest, but best approximates equilibrium. “Break Local Structure” is much faster than “Refold Whole RNA for Each Sequence” and it calculates the secondary structure formation free energy change of the original structure with any pairs that involve the oligonucleotide-bound nucleotides broken. For most applications, “Break Local Structure” is a good compromise between accuracy and calculation time.
6. A checkbox is labeled “Include Target Suboptimal Structures in Free Energy Calculation.” When checked, the cost of opening base pairs in the target structure is calculated for each suboptimal secondary structure in the “Break Local Structure” and “Refold Whole RNA for Each Sequence” modes. The contribution made by each suboptimal structure to the total cost of opening target self-structure is weighted according the folding free energy change. In general, it is recommended to check this box to account for alternative possible secondary structures. This checkbox serves no purpose in the “Do Not Consider Target Structure” mode.

Input information about the oligonucleotides

7. The oligonucleotide length is entered in the text box to the right of “Oligo Length.” Oligonucleotides can be either RNA or DNA and the selection is indicated by the button in the “Oligomer Chemistry” box. Also, choose an oligonucleotide concentration. The concentration units can be changed between mM, μ M, nM, and pM by clicking the down arrow to the right of the displayed units. For the example shown in Figure 12.6.12, DNA oligonucleotides of 18 nucleotides are considered at a concentration of 1 μ M.

Change the focus to a region of the target RNA

8. Next, the region for oligonucleotide binding can be reduced by adjusting the start and stop locations. Start refers to the target RNA nucleotide bound to the 3' end of the first oligonucleotide and stop refers to the nucleotide bound to the 5' end of the last oligonucleotide in the walk. By default, these limits are set to the 5' and 3' ends of the target sequence. To adjust the limits, use the up and down arrows next to the value. Limiting the area of interest on the target RNA strand reduces the calculation time.

Start the Calculation

9. Start the calculation by clicking the “Start OligoWalk” button. A window will open to show the progress of the calculation. When the calculation is complete, RNAstructure will open the OligoWalk output window as shown in Figure 12.6.13.

Navigating the OligoWalk output window

10. The OligoWalk output window provides an interactive method for displaying the calculated thermodynamic parameters. The target sequence is drawn left to right across the window in a 5' to 3' direction. Red nucleotides are predicted to be base paired in the lowest free energy structure. Black nucleotides are predicted to be single-stranded. The currently displayed oligonucleotide is above the target sequence in a 3' to 5' direction. The position along the target of the currently displayed oligonucleotide is indicated in the upper left-hand corner of the display. Oligonucleotides are numbered according to the 5'-most binding nucleotide in the target, therefore the oligonucleotides are numbered from 1. Below the current oligonucleotide is the backbone chemistry (RNA or DNA) and concentration of the oligonucleotide.

The thermodynamic parameters at the top of the screen are free energy changes at 37 °C in kcal/mol. “Overall ΔG_{37}° ” is the total free energy change of binding for a structured target and structured oligonucleotide. “Duplex ΔG_{37}° ” is the free energy change of duplex formation between the oligonucleotide and target, without the cost of opening self-structure. “Break Targ. ΔG_{37}° ” is the free energy cost of opening target secondary structure. “Oligo Self ΔG_{37}° ” and “Oligo-Oligo ΔG_{37}° ” are the free energy costs of opening unimolecular and bimolecular self-structure in the oligonucleotide, respectively. The T_m is the melting temperature of duplex formation in °C, not accounting for self-structure in target or oligonucleotide.

The graph, by default, shows the “Overall ΔG_{37}° ” and “Duplex ΔG_{37}° ” profile along the target sequence. The graph color is identical to the text color above, except that the color for “Overall ΔG_{37}° ” for the current oligonucleotide is in red. The free energy term that is graphed can be changed by selecting “Free Energy” under the “Graph” menu option. Each of the free energy terms can be graphed and a check on the menu shows the current selection.

11. The currently displayed oligonucleotide can be changed in several ways. The left or right arrow keys move the displayed oligonucleotide 5' and 3' respectively. This can also be done by clicking the buttons labeled “<” or “>.” The currently displayed nucleotide is skipped ten

nucleotides by clicking the buttons labeled “<<” or “>>.” By clicking the “Go...” button, a navigation window is reached. In the navigation window, a specific oligonucleotide for display can be indicated or a button, labeled “Most Stable” can be clicked to display the oligonucleotide predicted to have the highest affinity to the target.

12. For oligonucleotides with self-structure, the self-structure can be drawn on the screen by double-clicking the oligonucleotide sequence. For oligonucleotides with both bimolecular and unimolecular structure, a window opens to allow user selection of the structure type to display.

GUIDELINES FOR UNDERSTANDING RESULTS

The RNAstructure computer program, on average, predicts 73% of known base pairs in the lowest free energy structure for a diverse set of sequences shorter than 700 nucleotides and with known secondary structure (Mathews et al., 2004). Using a different set of known structures, however, RNAstructure only predicted 56% of known base pairs (Dowell and Eddy, 2004). Therefore, secondary structure prediction should be viewed as a method for developing structure hypotheses. Suboptimal structures are thus alternative hypotheses for the secondary structure. Recently developed methods for RNA secondary structure prediction can also be used to develop alternative hypotheses, including maximum expected accuracy structure prediction (Lu et al., 2009) and ProbKnot (Bellaousov and Mathews, 2010), which can predict pseudoknots.

Using RNAstructure, constraints or restraints on the possible structures can be specified. It has been shown that the use of constraints based on experimental data improve the accuracy of secondary structure prediction (Deigan et al., 2009; Mathews et al., 2004; Mathews et al., 1999b). RNAstructure can use constraints based on enzymatic cleavage (revealing paired or unpaired nucleotides) (Knapp, 1989), FMN cleavage (revealing uracils in GU pairs) (Burgstaller et al., 1997), chemical modification (revealing nucleotides that are unpaired, at the ends of helices, or in or adjacent to GU pairs) (Ehresmann et al., 1987), quantified DMS reactivity data (Cordero et al., 2012), and SHAPE data (Deigan et al., 2009; Hajdin et al., 2010).

Base pair probabilities can be used to estimate confidence in a predicted base pair (Mathews, 2004). On average, 66% of predicted base pairs in the lowest free energy structure are in the known structure for a diverse set of sequences. When only base pairs with predicted pairing probability at or above 0.90 are considered, however, 83% of predicted pairs are in the known structure. For a probability threshold of 0.99, this accuracy increases to 91%. Nearly one quarter of predicted base pairs, on average, in the lowest free energy structure have pairing probability of at least 0.99.

OligoWalk provides an estimate of binding affinity of structured oligonucleotides to a structured RNA target (Mathews et al., 1999a). For an oligonucleotide to bind tightly, not only should the duplex free energy change be low (more negative), the magnitude of the cost of opening target structure should also be minimized. It has been shown that the duplex formation free energy and oligonucleotide self-structure terms correlate with antisense oligonucleotide efficacy (Lu and Mathews, 2008a; Matveeva et al., 2003). Avoiding self-

structure in mRNA targets is also important in siRNA design (Bohula et al., 2003; Far and Sczakiel, 2003; Heale et al., 2005; Lu and Mathews, 2007; Petch et al., 2003; Tafer et al., 2008). An RNAstructure web server for siRNA design is available, and this uses target accessibilities determined by a modified OligoWalk (Lu and Mathews, 2007; Lu and Mathews, 2008b).

COMMENTARY

RNAstructure predicts secondary structures on the basis of thermodynamics. The lowest free energy structure is the structure that is most likely to occur at equilibrium, and predicting the lowest free energy structure is the traditional method for predicting RNA secondary structure. The secondary structure formation free energy change is estimated using a set of empirical nearest neighbor parameters, determined from optical melting experiments on model systems (Mathews et al., 2004; Mathews et al., 1999b; Xia et al., 1998). The partition function is likewise built from free energy changes for structure formation and implicitly considers all possible secondary structures when calculating base pair probabilities.

For free energy minimization, RNAstructure uses a dynamic programming algorithm that guarantees the predicted lowest free energy structure will be found. Essentially, the structure prediction problem is divided into smaller problems and recursion builds the complete secondary structure. Two reviews are available that explain dynamic programming in detail (Eddy, 2004; Mathews and Zuker, 2004). The partition function is also calculated with a dynamic programming algorithm (McCaskill, 1990).

The dynamic programming algorithms used in RNAstructure, however, cannot predict pseudoknotted (non-nested) base pairs. On average, only 1.4% of base pairs are pseudoknotted in a database of diverse RNA structures, but this percentage can be much higher for some classes of RNA structures, such as RNase P (Brown, 1999) and tmRNA (Williams and Bartel, 1996). A review of the thermodynamics and prediction of pseudoknots is available (Liu et al., 2010). The ProbKnot component of RNAstructure can predict pseudoknots, but the accuracy of pseudoknot prediction by this and other freely available tools is relatively low, but can be much higher when SHAPE mapping data are used to restrain the prediction (Bellaousov and Mathews, 2010; Hajdin et al., 2013).

The structure prediction algorithms presented in this unit scale $O(N^3)$, where N is the sequence length. This means that doubling the sequence length would make the calculation time approximately eight times longer. This scaling is considered costly, but in practice it does not limit most calculations. Table 1 shows sample calculation times for predicting lowest free energy structures and for partition functions. For sequences up to 2,900 nucleotides long, these calculations take less than 11 minutes. For long sequences, partition function calculations in RNAstructure can be performed on graphics processor units (GPUs), but this requires running on the command line (Stern and Mathews, 2013). OligoWalk also poses little difficulty in structure prediction times (Table 2). In “Break Local Structure” mode, it scales $O(NL^3)$, where N is the target length and L is the oligonucleotide length. It scales linearly with length of the target after the target structure

has been predicted ahead of time; therefore doubling the target sequence length roughly doubles the calculation time.

Several other software packages are available for predicting low free energy RNA secondary structures. A well-maintained list of secondary structure prediction programs is available on Wikipedia at http://en.wikipedia.org/wiki/List_of_RNA_structure_prediction_software. Two of the more popular packages are mfold (Zuker, 1989; Zuker, 2003) and the Vienna RNA package (Hofacker, 2003; Lorenz et al., 2011). The packages differ slightly in the implementation of the nearest neighbor parameters for multibranch loops and exterior loops (loops that contain the ends of the sequence) as compared to RNAstructure. For example, RNAstructure explicitly considers both coaxial stacking of adjacent helices and helices separated by a single mismatch. These interactions are known to stabilize RNA structures (Kim et al., 1996; Lescoute and Westhof, 2006; Tyagi and Mathews, 2007; Walter et al., 1994). The Vienna Package 2.0 considers coaxial stacking in free energy minimization, but does not include coaxial stacking in the partition function prediction of base pair probabilities (Lorenz et al., 2011). Mfold does not consider coaxial stacking in the dynamic programming algorithm, but a second step, efn2, recalculates the free energy change of folding for each structure including coaxial stacking of adjacent helices and helices separated by a single mismatch (Mathews et al., 1999b). Because of these differences in the energy model, the programs are not guaranteed to predict the same lowest free energy structure. Benchmarks, however, showed that these programs have similar average accuracy (Dowell and Eddy, 2004; Lorenz et al., 2011).

Suggestions for Further Analysis

XRNA is a program that can make publication quality structure drawings. It is available from the UC Santa Cruz RNA Center at <http://rna.ucsc.edu/rnacenter/xrna/xrna.html>. A second useful program for manipulating structure drawings is VARNA, which allows interactive manipulation of the drawing (Darty et al., 2009).

Acknowledgments

Continued development and support of RNAstructure and the writing of these protocols were supported by National Institutes of Health grant R01 GM076485.

Literature Cited

- Bellaousov S, Mathews DH. ProbKnot: Fast prediction of RNA secondary structure including pseudoknots. *RNA*. 2010; 16:1870–1880. [PubMed: 20699301]
- Bellaousov S, Reuter JS, Seetin MG, Mathews DH. RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res*. 2013; 41:W471–474. [PubMed: 23620284]
- Bohula EA, Salisbury AJ, Sohail M, Playford MP, Riedemann J, Southern EM, Macaulay VM. The efficacy of small interfering RNAs targeted to the type 1 insulin-like growth factor receptor (IGF1R) is influenced by secondary structure in the IGF1R transcript. *J. Biol. Chem*. 2003; 278:15991–15997. [PubMed: 12604614]
- Brown JW. The ribonuclease P database. *Nucleic Acids Res*. 1999; 27:314. [PubMed: 9847214]
- Burgstaller P, Hermann T, Huber C, Westhof E, Famulok M. Isoalloxazine derivatives promote photocleavage of natural RNAs at GU base pairs embedded within helices. *Nucleic Acids Res*. 1997; 25:4018–4027. [PubMed: 9321652]

- Cordero P, Kladwang W, VanLang CC, Das R. Quantitative dimethyl sulfate mapping for automated RNA secondary structure inference. *Biochemistry*. 2012; 51:7037–7039. [PubMed: 22913637]
- Darty K, Denise A, Ponty Y. VARNAs: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*. 2009; 25:1974–1975. [PubMed: 19398448]
- Deigan KE, Li TW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.* 2009; 106:97–102. [PubMed: 19109441]
- Ding Y, Lawrence CE. A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res.* 2003; 31:7280–7301. [PubMed: 14654704]
- Dowell RD, Eddy SR. Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. *BMC Bioinformatics*. 2004; 5:71. [PubMed: 15180907]
- Eddy SR. How do RNA folding algorithms work? *Nat. Biotechnol.* 2004; 22:1457–1458. [PubMed: 15529172]
- Ehresmann C, Baudin F, Mougél M, Romby P, Ebel J, Ehresmann B. Probing the structure of RNAs in solution. *Nucleic Acids Res.* 1987; 15:9109–9128. [PubMed: 2446263]
- Far RK, Sczakiel G. The activity of siRNA in mammalian cells is related to structural target accessibility: a comparison with antisense oligonucleotides. *Nucleic Acids Res.* 2003; 31:4417–4424. [PubMed: 12888501]
- Hajdin CE, Bellaousov S, Huggins W, Leonard CW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. U.S.A.* 2013; 110:5498–5503. [PubMed: 23503844]
- Hajdin CE, Ding F, Dokholyan NV, Weeks KM. On the significance of an RNA tertiary structure prediction. *RNA*. 2010; 16:1340–1349. [PubMed: 20498460]
- Harmanci AO, Sharma G, Mathews DH. Efficient pairwise RNA structure prediction using probabilistic alignment constraints in Dynalign. *BMC Bioinformatics*. 2007; 8:130. [PubMed: 17445273]
- Harmanci AO, Sharma G, Mathews DH. PARTS: Probabilistic Alignment for RNA joint Secondary structure prediction. *Nucleic Acids Res.* 2008; 36:2406–2417. [PubMed: 18304945]
- Harmanci AO, Sharma G, Mathews DH. Stochastic sampling of the RNA structural alignment space. *Nucleic Acids Res.* 2009; 37:4063–4075. [PubMed: 19429694]
- Harmanci AO, Sharma G, Mathews DH. TurboFold: Iterative probabilistic estimation of secondary structures for multiple RNA sequences. *BMC Bioinformatics*. 2011; 12:108. [PubMed: 21507242]
- Heale BS, Soifer HS, Bowers C, Rossi JJ. siRNA target site secondary structure predictions using local stable substructures. *Nucleic Acids Res.* 2005; 33:e30. [PubMed: 15722476]
- Hofacker IL. Vienna RNA secondary structure server. *Nucleic Acids Res.* 2003; 31:3429–3431. [PubMed: 12824340]
- Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.* 1994; 125:167–168.
- Kim J, Walter AE, Turner DH. Thermodynamics of coaxially stacked helices with GA and CC mismatches. *Biochemistry*. 1996; 35:13753–13761. [PubMed: 8901517]
- Knapp G. Enzymatic approaches to probing RNA secondary and tertiary structure. *Methods Enzymol.* 1989; 180:192–212. [PubMed: 2482414]
- Lescoute A, Westhof E. Topology of three-way junctions in folded RNAs. *RNA*. 2006; 12:83–93. [PubMed: 16373494]
- Liu B, Mathews DH, Turner DH. RNA pseudoknots: folding and finding. *F1000 Biol. Rep.* 2010; 2:8. [PubMed: 20495679]
- Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 2011; 6:26. [PubMed: 22115189]
- Lu ZJ, Gloor JW, Mathews DH. Improved RNA secondary structure prediction by maximizing expected pair accuracy. *RNA*. 2009; 15:1805–1813. [PubMed: 19703939]
- Lu ZJ, Mathews DH. Efficient siRNA selection using hybridization thermodynamics. *Nucleic Acids Res.* 2007; 36:640–647. [PubMed: 18073195]

- Lu ZJ, Mathews DH. Fundamental differences in the equilibrium considerations for siRNA and antisense oligodeoxynucleotide design. *Nucleic Acids Res.* 2008a; 36:3738–3745. [PubMed: 18483081]
- Lu ZJ, Mathews DH. OligoWalk: An online siRNA design tool utilizing hybridization thermodynamics. *Nucleic Acids Res.* 2008b; 36:W104–W108. [PubMed: 18490376]
- Lu ZJ, Turner DH, Mathews DH. A set of nearest neighbor parameters for predicting the enthalpy change of RNA secondary structure formation. *Nucleic Acids Res.* 2006; 34:4912–4924. [PubMed: 16982646]
- Mathews DH. Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA.* 2004; 10:1178–1190. [PubMed: 15272118]
- Mathews DH. Predicting a set of minimal free energy RNA secondary structures common to two sequences. *Bioinformatics.* 2005; 21:2246–2253. [PubMed: 15731207]
- Mathews DH, Burkard ME, Freier SM, Wyatt JR, Turner DH. Predicting oligonucleotide affinity to nucleic acid targets. *RNA.* 1999a; 5:1458–1469. [PubMed: 10580474]
- Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A.* 2004; 101:7287–7292. [PubMed: 15123812]
- Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters provides improved prediction of RNA secondary structure. *J. Mol. Biol.* 1999b; 288:911–940. [PubMed: 10329189]
- Mathews DH, Turner DH. Dynalign: An algorithm for finding the secondary structure common to two RNA sequences. *J. Mol. Biol.* 2002; 317:191–203. [PubMed: 11902836]
- Mathews, DH.; Zuker, M. Predictive methods using RNA sequences. In *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*. 3rd Edition. Baxevanis, A.; Ouellette, F., editors. John Wiley & Sons, Inc.; 2004. p. 143-170.
- Matveeva OV, Mathews DH, Tsodikov AD, Shabalina SA, Gesteland RF, Atkins JF, Freier SM. Thermodynamic criteria for high hit rate antisense oligonucleotide design. *Nucleic Acids Res.* 2003; 31:4989–4994. [PubMed: 12930948]
- McCaskill JS. The equilibrium partition function and base pair probabilities for RNA secondary structure. *Biopolymers.* 1990; 29:1105–1119. [PubMed: 1695107]
- Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* 2005; 127:4223–4231. [PubMed: 15783204]
- Petch AK, Sohail M, Hughes MD, Benter I, Darling J, Southern EM, Akhtar S. Messenger RNA expression profiling of genes involved in epidermal growth factor receptor signalling in human cancer cells treated with scanning array-designed antisense oligonucleotides. *Biochem. Pharmacol.* 2003; 66:819–830. [PubMed: 12948863]
- Piekna-Przybylska D, DiChiacchio L, Mathews DH, Bambara RA. A sequence similar to tRNA³Lys gene is embedded in HIV-1 U3/R and promotes minus strand transfer. *Nat. Struct. Mol. Biol.* 2009; 17:83–89.
- Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics.* 2010; 11:129. [PubMed: 20230624]
- Sprinzl M, Vassilenko KS. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* 2005; 33:D139–140. [PubMed: 15608164]
- Stern HA, Mathews DH. Accelerating calculations of RNA secondary structure partition functions using GPUs. *Algorithms Mol. Biol.* 2013; 8:29. [PubMed: 24180434]
- Tafer H, Ameres SL, Obernosterer G, Gebeshuber CA, Schroeder R, Martinez J, Hofacker IL. The impact of target site accessibility on the design of effective siRNAs. *Nat. Biotechnol.* 2008; 26:578–583. [PubMed: 18438400]
- Tyagi R, Mathews DH. Predicting helical coaxial stacking in RNA multibranch loops. *RNA.* 2007; 13:939–951. [PubMed: 17507661]
- Uzilov AV, Keegan JM, Mathews DH. Detection of non-coding RNAs on the basis of predicted secondary structure formation free energy change. *BMC Bioinformatics.* 2006; 7:173. [PubMed: 16566836]

- Walter AE, Turner DH, Kim J, Lyttle MH, Müller P, Mathews DH, Zuker M. Coaxial stacking of helices enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl. Acad. Sci. USA*. 1994; 91:9218–9222. [PubMed: 7524072]
- Williams KP, Bartel DP. Phylogenetic analysis of tmRNA secondary structure. *RNA*. 1996; 2:1306–1310. [PubMed: 8972778]
- Xia T, SantaLucia J Jr. Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick pairs. *Biochemistry*. 1998; 37:14719–14735. [PubMed: 9778347]
- Xu Z, Mathews DH. Multalign: an algorithm to predict secondary structures conserved in multiple RNA sequences. *Bioinformatics*. 2011; 27:626–632. [PubMed: 21193521]
- Zuker M. On finding all suboptimal foldings of an RNA molecule. *Science*. 1989; 244:48–52. [PubMed: 2468181]
- Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 2003; 31:3406–3415. [PubMed: 12824337]
- Zuker, M.; Mathews, DH.; Turner, DH. Algorithms and thermodynamics for RNA secondary structure prediction: A practical guide.. In: Barciszewski, J.; Clark, BFC., editors. *RNA Biochemistry and Biotechnology*. Kluwer Academic Publishers; Boston: 1999. p. 11-43.

>Title of Sequence

AAA GCGG UUTGTT UTCUTaaTCTXXXXUCAGG

UUA GCCG UUTGTT UTCUTaaTCTGGG

Figure 12.6.1.

The FASTA file format for the RNAstructure web servers. Sequences can be uploaded to the RNAstructure webserver in FASTA format. For FASTA, the first line, a title line, needs to start with “>”. Subsequent lines should only contain sequence and whitespace, which is ignored. Lowercase nucleotides will not be forced single stranded in structure prediction. X can also be used to indicate a nucleotide that neither pairs nor stacks.

Firefox

Welcome to the Predict a Secondary Structure...

ma.umc.rochester.edu/RNAstructureWeb/Servers/Predict1/Example.php

RNAstructure

Web Servers for RNA Secondary Structure Prediction

[Servers Home](#) [Contributors](#) [Mathews Lab Home](#) [RNAstructure Downloads](#) [Help / Contact Us](#)

Welcome to the Predict a Secondary Structure Web Server

The **Predict a Secondary Structure** server combines four separate prediction and analysis algorithms: **calculating a partition function, predicting a maximum free energy (MFE) structure, finding structures with maximum expected accuracy, and pseudoknot prediction.** This server takes a sequence, either RNA or DNA, and creates a highly probable, probability annotated group of secondary structures, starting with the lowest free energy structure and including others with varied probabilities of correctness. Other structures are included because the minimum free energy structure may not be the correct one. Note that if SHAPE constraints are specified, the SHAPE constraints are applied to the probability annotated structures. In addition, a second group of SHAPE constrained, SHAPE annotated structures will be generated. This SHAPE structure group is distinct from the probability annotated structure group, and is not probability annotated itself.

If you plan to leave this page while waiting for calculations to complete, please **make sure you have entered a valid email address** below. Leaving this page without entering a valid email address will render results inaccessible.

If you need specific **help using the Predict a Secondary Structure server**, please click [here](#).

Select Sequence Input:

Select Sequence File:

No file selected.

– OR –

Enter your sequence title and content below (replaces upload if present).
Please enter nucleotides only, no headers or comments in content.
Valid nucleotides are A, C, G, T, U, and X (unknown nucleotide). Lower case nucleotides are forced single stranded in calculation.

Sequence Title:

> RA7680

Sequence: (Example sequence has been displayed.)

```
GGGGGCGUAG CUCAGAUSSU AGAGCGCUCG CUUGGCGUGU GAGAGGUACC
GGGAUCCGaUA CCGGGCGCCU CCACCA
```

Select Nucleic Acid Type:

DNA: ☐

RNA: ☒

Select Predict a Secondary Structure Options:

Default Data

Figure 12.6.2.

A screen shot of the input form for the RNAstructure “Predict a Secondary Structure” web server. The example sequence for tRNA RA7680 (Sprinzl and Vassilenko, 2005) was inserted by clicking on the link labeled “Click here to add an example sequence to the box.”

A:	DS :	B:	DS :
	XA		15
	-1		25
	SS :		76
	XB		-1
	-1		SS :
	Mod :		17
	XC		18
	-1		20
	Pairs :		35
	XD1 XD2		-1
	-1 -1		Mod :
	FMN :		2
	XE		15
	-1 -1		-1
	Forbids :		Pairs :
	XF1 XF2		16 26
	-1 -1		-1 -1
			FMN :
			-1
			Forbids :
			15 27
			-1 -1

Figure 12.6.3.

The RNAstructure constraint file format. Folding constraint files are plain text files. These can be manually edited. For multiple entries of a specific type of constraint, entries are each listed on a separate line. Note that all specifiers, followed by “-1” or “-1 -1”, are expected by RNAstructure. For all specifiers that take two arguments, it is assumed that the first argument is the 5’ nucleotide. Panel A shows the specification of the fields. The constraints are XA, nucleotides that will be double-stranded; XB, nucleotides that will be single-stranded (unpaired); XC, nucleotides accessible to chemical modification; XD1 and XD2,

forced base pair between XD1 and XD2; XE, nucleotides accessible to FMN cleavage (U in GU pair); and XF1 and XF2, a base pair prohibited between nucleotides XF1 and XF2. All nucleotide indexes are from numbering 5' to 3', with the nucleotide at the 5' end with index of 1. Panel B shows an example.

9	-999
10	-999
11	0.042816
12	0
13	0.15027
14	0.16201

Figure 12.6.4.

The SHAPE data file format. SHAPE mapping data are provided in a plain text file. The file format comprises two columns. The first column is the nucleotide number, and the second column is the reactivity. Nucleotides for which there is no SHAPE data can either be left out of the file, or the reactivity can be entered as less than -500. Columns are separated by any white space. By default, RNAstructure looks for SHAPE data files to have the file extension .SHAPE, but any plain text file can be read. Note that there is no header information in the file. In this example, nucleotides 1 through 10 have no reactivity information. Nucleotide 11 has a normalized SHAPE reactivity of 0.042816. Nucleotide 12 has a normalized SHAPE reactivity of 0, which is NOT the same as having no reactivity when using the pseudo-energy constraints.

Figure 12.6.5.

A screen shot of the RNAstructure “Predict a Secondary Structure” output form. This view shows the lowest free energy structure predicted by Fold for the example sequence, RA7680, which appears at the top of the results page. The predicted structure is color annotated according to probabilities. Base paired nucleotides are colored according to pair probabilities, and unpaired nucleotides are colored according to the probability of being single stranded. The next structure can be displayed by clicking the button labeled “Next.”

```

76  ENERGY = -31.8      RA7680
    1  G          0      2   72      1
    2  G          1      3   71      2
    3  G          2      4   70      3
    4  G          3      5   69      4
    5  G          4      6   68      5
    6  C          5      7   67      6
    7  G          6      8   66      7
    8  U          7      9    0      8
    9  A          8     10    0      9
   10  G          9     11   25     10
   11  C         10     12   24     11
   12  U         11     13   23     12
...
   70  U         69     71    3     70
   71  C         70     72    2     71
   72  C         71     73    1     72
   73  A         72     74    0     73
   74  C         73     75    0     74
   75  C         74     76    0     75
   76  A         75     0      0     76
76  ENERGY = -31.8      RA7680
    1  G          0      2   72      1
    2  G          1      3   71      2
    3  G          2      4   70      3
    4  G          3      5   69      4
    5  G          4      6   68      5
...

```

Figure 12.6.6.

The ct file format. A ct (connectivity table) file contains secondary structure information for a sequence. The format used by RNAstructure is as follows. The start of the first line is the number of nucleotides in the sequence. The remainder of the first line is the title of the structure. Each of the subsequent lines provides information about a given base in the sequence. Each base has its own line, with these elements in order: nucleotide number (starting with 1), base (A, C, G, T, U, X), the nucleotide connection in the 5' direction, the nucleotide connection in the 3' direction, number of the base to which the current nucleotide

is paired (no pairing is indicated by 0, zero), and natural numbering (this will be the nucleotide index repeated for the calculations described in this unit). The ct file may hold multiple structures for a single sequence. This is done by repeating the format for each structure without blank lines between structures. The example shown here is the structure predicted for RA7680 by the “Predict a Secondary Structure Common to Three or More Sequences” webserver, as illustrated in Protocol 1. “...” indicates text from the file that is not shown in the figure.

Figure 12.6.7.

A screen shot of the probability dot plot predicted by the RNAstructure “Predict a Secondary Structure” web server. This shows the probability dot plot from the bottom of the results page. This summarizes the set of base pairs predicted to have pairing probability of 1% or higher. Note that the color key indicates the negative base 10 logarithm of pairing probability, therefore lower values are more probable pairs.

Figure 12.6.8.

A screen shot of the RNAstructure sequence editor. This is the RNAstructure GUI as it appears on Microsoft Windows 7. The windows are similar on other operating systems, with the exception that the menu items appear at the top of the screen on OS X, as expected. The RA7680 sequence, which is available as a sample file, was opened from disk.

Figure 12.6.9.

A screen shot of the RNAstructure Fold RNA Single Strand input form. This is the RNAstructure GUI as it appears in Microsoft Windows 7. The Sequence File button was clicked to select RA7680.seq. The remaining fields show their default values.

Figure 12.6.10.

A screen shot of the base pairing probability window, as it appears in Microsoft Windows 7, for the calculation with RA7680.seq. The plot range was reduced to a maximum of 2 in $-\log_{10}$ of the pairing probability, hence only pairs with 1% or higher pairing probability are displayed. The user clicked on the dot in the upper-right hand corner, which is the pair between nucleotides 1 and 72. The text at the bottom of the screen shows that $-\log_{10}$ of the pairing probability is 0.195, or 63.8%.

Figure 12.6.11.

A screen shot of the drawing Window for the RNAstructure GUI, showing the lowest free energy structure predicted for RA7680 with probability color annotation. This is the drawing windows as it appears on Microsoft Windows 7. The probability color annotation key window also appears.

Figure 12.6.12.

A screen shot of the RNAstructure GUI OligoWalk input form as it appears in Microsoft Windows 7. The windows are similar on other operating systems, with the exception that the menu items appear at the top of the screen on OS X, as expected. The structure for tRNA RA7680 was selected, and the default values appear on the screen.

Figure 12.6.13.

A screen shot of the RNAstructure GUI OligoWalk results for tRNA RA7680. This shows the results as they appear on Microsoft Windows 7. The bar graph shows the default plot of overall free energy of binding (Overall $G_{37.0}^{\circ}$) and duplex binding free energy (Duplex $G_{37.0}^{\circ}$). Folding free energy changes are in kcal/mol. The red bar at position 1 indicates that the first oligonucleotide (targeted to the 5' end of the sequence) is selected, and its free energy changes are indicated at the top of the screen.

Table 1

Sample Structure Prediction Times. These calculations were performed using the RNAstructure graphical interface on Microsoft Windows 7. The hardware was a machine with a 3.4 GHz Intel I7-2600K, 4 core processor and 8 GB of memory. These calculations multithread across multiple cores; therefore using a 4 core computer cuts the calculation time by almost a factor of 4.

Sequence:	Length:	Calculation Time (min:sec):	
		Fold (lowest free energy prediction):	partition (partition function calculation):
<i>Tetrahymena thermophila</i> Group I Intron	433	0:03	0:02
<i>Escherichia coli</i> small subunit ribosomal RNA	1542	1:49	1:35
<i>Escherichia coli</i> large subunit ribosomal RNA	2904	10:35	11:02

Table 2

Sample OligoWalk Calculation Times for the “Break Local Structure” mode. These calculations were performed using the RNAstructure graphical interface on Microsoft Windows 7. The hardware was a machine with a 3.4 GHz Intel I7-2600K, 4 core processor and 8 GB of memory. These calculations execute on a single core only.

Sequence:	Length:	Oligonucleotide Length:	Time (min:sec):
RA7680 tRNA	76	18	0:01
<i>Escherichia coli</i> small subunit ribosomal RNA	1542	12	0:04
<i>Escherichia coli</i> small subunit ribosomal RNA	1542	18	0:07
<i>Escherichia coli</i> large subunit ribosomal RNA	2904	18	0:33