*S*fold web server for statistical folding and rational design of nucleic acids

Ye Ding*, Chi Yu Chan and Charles E. Lawrence

Bioinformatics Center, Wadsworth Center, New York State Department of Health, 150 New Scotland Avenue, Albany, NY 12208, USA

ABSTRACT

The *S*fold web server provides user-friendly access to *S*fold, a recently developed nucleic acid folding software package, via the World Wide Web (WWW). The software is based on a new statistical sampling paradigm for the prediction of RNA secondary structure. One of the main objectives of this software is to offer computational tools for the rational design of RNA-targeting nucleic acids, which include small interfering RNAs (siRNAs), antisense oligonucleotides and *trans*-cleaving ribozymes for gene knock-down studies. The methodology for siRNA design is based on a combination of RNA target accessibility prediction, siRNA duplex thermodynamic properties and empirical design rules. Our approach to target accessibility evaluation is an original extension of the underlying RNA folding algorithm to account for the likely existence of a population of structures for the target mRNA. In addition to the application modules *S*irna, *S*oligo and *S*ribo for siRNAs, antisense oligos and ribozymes, respectively, the module *S*rna offers comprehensive features for statistical representation of sampled structures. Detailed output in both graphical and text formats is available for all modules. The *S*fold server is available at http://sfold.wadsworth.org and http://www.bioinfo.rpi.edu/applications/sfold.

INTRODUCTION

The prediction of RNA secondary structure for a single sequence is a classic problem in computational biology. Free energy minimization has been one of the most popular methods for addressing this problem. The established algorithm computes the optimal folding and a set of suboptimal foldings (1,2). A more recent algorithm computes all suboptimal foldings within any specified increment of the minimum free energy (3). However, neither method computes a statistically valid set of foldings. Toward a better characterization of the ensemble of probable RNA secondary structures, McCaskill developed an algorithm for the calculation of equilibrium partition functions and base-pairing probabilities (4). Although these algorithms are computationally efficient, each nevertheless has inherent limitations (5). Perhaps the most important limitation stems from the fact that these methods were developed primarily for structural RNAs that may have unique structures. Messenger RNAs (mRNAs), on the other hand, may exist as a population of structures (6).

We recently developed an algorithm to sample rigorously and exactly from the Boltzmann ensemble of secondary structures (5), based on Turner thermodynamic parameters (7,8). This statistical sampling algorithm guarantees the generation of a statistically representative sample of structures. In addition, this algorithm enables the development of unique tools for a number of important applications that include the rational design of RNA-targeting nucleic acids.

Single-stranded regions in RNA secondary structure are likely to be accessible for RNA-targeting nucleic acids through base-pairing interactions. Target accessibility has long been established as an important factor for the potency of antisense oligonucleotides (oligos) and *trans*-cleaving ribozymes. Recently, the importance of target structure and accessibility for the function of siRNAs has been demonstrated using a number of experimental approaches that include oligo library (9), oligo array (10), antisense evaluation of accessibility (11), and targeting the same sequence in both structured and unstructured sites (12).

Based on the RNA structure-sampling algorithm, we also developed a probability profiling method for the prediction of target accessibility (5,13). A stochastic approach to accessibility evaluation may be essential to account for the likely existence of a population of structures for mRNAs (6). The probability profiling approach reveals target sites that are commonly accessible in a high proportion of statistically representative structures for the target RNA. Through assignment of statistical confidence in predictions, this approach bypasses the long-standing difficulty in accessibility evaluation due to limited representation of probable structures.

*To whom correspondence should be addressed. Tel: +1 518 486 1719; Fax: +1 518 402 4623; Email: yding@wadsworth.org

Based on these novel algorithms (5,13), the first version of *S*fold, a nucleic acid folding and design software package, was completed during spring 2002. The package currently consists of four application modules. Modules *S*irna, *S*oligo, and *S*ribo provide computational tools for target accessibility prediction and for the rational design of siRNAs, antisense oligos and *trans*-cleaving ribozymes, respectively. General statistical folding features and output are available from the fourth module, *S*rna. Since April 2003, the software has been available to the scientific community through web servers at http://sfold.wadsworth.org and http://www.bioinfo.rpi.edu/applications/sfold. A manual and frequently asked questions (FAQs) for the software are provided on the websites.

In this article, we highlight the main features of the software and the web service. Users are encouraged to consult the online manual for more information on the software, and to examine sample output located at http://sfold.wadsworth.org/demo.

INPUT

A user can submit jobs to the web server by first clicking on the application module of interest on the server's front page and then filling out the job submission form. A job can run in either interactive mode or batch mode. Current limits are 200 bases for an interactive job and 5000 bases for a batch job. For a batch job, a correct email address is required for notification of job completion. Strategies for sequences >5000 bases in length are described in the online manual. Sequences in raw format, in FASTA format, or in GenBank format are accepted. A sequence can be entered by copying and pasting into the sequence input window; alternatively, a sequence file in the user's file directory (folder) can be selected for uploading. Any character other than A, C, G, T or U will be edited out. An option is provided for cases in which the RNA sequence to be folded is the reverse complement of the input sequence. For *S*oligo, the user has the option to set the length of the oligos. The default length is 20 nt. For *S*ribo, the user can specify a specific NUH cleavage triplet for hammerhead ribozymes. The default triplet is GUC.

OUTPUT

The output page for each module includes both graphical representations and relevant text files. Many images are presented through interactive graphics applications via the user's web browser. The user can download a colored or black-and-white plot in PDF or PostScript format. For monochrome printers, the black-and-white version is recommended. All of the graphs, with the exception of RNA structure diagrams, are first generated by the public domain software Gnuplot and then post-processed by Perl scripts to provide additional features and to improve visualization. Links to output from other modules with default input settings are provided. Output can be accessed on the server for up to 72 h after job completion. With the exception of siRNA internal stability profiles, all of the output in Zip or compressed tar (tar.gz) format is available for download. After the compressed file has been uncompressed by the user, a directory with the job ID as its name is created under the user's current directory. Under the job ID

directory, there are seven subdirectories and a file readme.txt to describe the files in the subdirectories. Below, we highlight the main output features for the four modules.

*S*irna module

The *S*irna module provides tools for the rational design of siRNAs. The design methodology is based on target accessibility evaluation (5,13), typical design rules and published empirical rules, and stability rules for siRNA duplex ends and for the region of the cleavage site (14,15). Recently, it has been reported that functional siRNA duplexes tend to have lower stability on the 5′-antisense end (∼4 bp) than on the 5′-sense end (14,15). It is proposed that both the absolute and relative stabilities of the siRNA duplex ends determine the degree to which each strand participates in the RNAi pathway (14). Furthermore, functional siRNAs tend to have relative instability at the cleavage site, which may facilitate product release and multiple turnovers (15). However, these rules on siRNA duplex stabilities do not guarantee siRNA function (15), perhaps because they do not address the structure of the target mRNA. By integrating rules on target accessibility, empirical rules and thermodynamic properties for siRNA duplexes, *S*irna provides a unique combination of tools for siRNA design.

Probability profiling for target accessibility prediction. For prediction of target accessibility, a complete probability profile of single-stranded regions is generated for the entire target RNA. Sites with high probabilities of being single-stranded are predicted to be accessible. At nucleotide position i, the profile shows the probability that nucleotides $i, i+1, i+2, \ldots, i+W-1$ are all unpaired. In other words, the profile is for consecutive fragments with a width W. Although the profile can be generated for any value of W, the default width $W = 4$ has been found to be particularly useful (13), perhaps because a minimum of four unpaired bases is thought to be required for the initiation of antisense binding (16). The profile probabilities are computed from a representative statistical sample of target secondary structures, with a default sample size of 1000 structures. A regional probability profile allows the user to examine any region of the complete profile. In the interactive graphic window for regional profiling, a new region of 200 bases can be selected either by specifying the starting position or by clicking ≪ or ≫ for the upstream or the downstream region. The text file sstrand.out contains output for probability profiling. As an example, the regional profile for the nt 201–400 region of rabbit β-globin mRNA (GenBank accession no. V00879) is illustrated in Figure 1. In addition, loop-specific profiles for $W = 1$ (5) are also available. However, we do not know whether a certain type of loop is more favorable than other types for binding by complementary nucleic acids.

Empirical rules. The following basic design rules are widely used: (i) siRNA duplexes should be composed of 21 nt sense and 21 nt antisense strands, paired so as to each have a 2 nt 3′ dTdT overhang; (ii) the siRNA sequence should have low to moderate GC content; (iii) sequences with more than three Gs or three Cs in a row should be avoided, because polyG and polyC sequences can form quartets (17) and may interfere in the siRNA silencing mechanism; (iv) AAAA or TTTT should also be avoided for RNA polymerase III mediated promoters
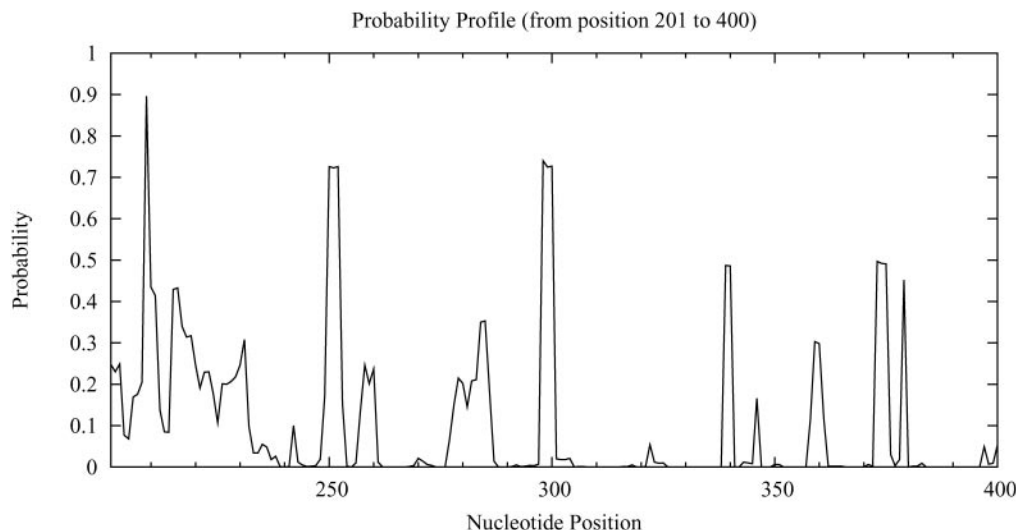
Figure 1. Regional profile (window width $W = 4$) for the region between nt 201 and 400 of the rabbit β-globin mRNA (GenBank accession no. V00879). Sites with high probabilities are predicted to be accessible.

because transcription tends to terminate at these sequences. It is of note that there is a lack of support in the literature for the significance of target patterns such as AA(N19) or NA(N19). Recently, empirical rules on sequence features of functional siRNAs have been reported by several groups (18–20). Based on the largest published siRNA data set, the rules by Reynolds and colleagues (18) are the most comprehensive. However, these rules alone do not guarantee siRNA function. A number of potent siRNAs in studies on effect of target structure and accessibility (9–12) do not meet key empirical rules. However, their function is explained by accessibility.

siRNA duplex thermodynamics. *S*irna computes a number of thermodynamics indexes for the implementation of rules on siRNA duplex stabilities, based on recent RNA thermodynamics parameters (7,8). 5′-antisense stability (AntiS, in kcal/mol) is computed as the sum of free energies for 4 bp stacks and the 3′ dangling T and a penalty for terminal A–U for the 5′ end of the antisense siRNA strand; 5′-sense stability (SS, in kcal/mol) is the sum for the 5′ end of the sense siRNA strand. Differential stability of siRNA duplex ends (DSSE, in kcal/mol) is the difference between the 5′-antisense stability and the 5′-sense stability, i.e. DSSE = AntiS − SS. For each of positions 2–18 of the antisense strand, the internal stability is the sum of 4 bp stacks, starting at this position in the 5′ → 3′ orientation. For position 1, the internal stability is the 5′-antisense stability. For position 19, the internal stability is the 5′-sense stability. The internal stabilities are used for constructing an internal stability profile for each siRNA duplex. For positions 16–19 on the 3′ end of the antisense strand, Khvorova and colleagues (15) extended the target sequence with the target RNA for the purpose of calculation. This treatment can lead to inaccurate information. For example, the profile is not guaranteed to be symmetric when the bases for the ends of the duplex are symmetric. For a correct comparison of the stabilities for the two siRNA duplex ends, we simply reverse the orientation to 3′ → 5′ in the calculation for positions 16–19. Average internal stability at the cleavage site (AIS, in kcal/mol) is

the average of internal stability values for positions 9–14 of the antisense strand (15).

Filters and scores for siRNA duplexes. Based on available rules that we consider to be important, *S*irna currently uses the following filters for screening siRNA candidates:

(i) antisense siRNA binding energy ≤−10 kcal/mol (target accessibility rule);
(ii) duplex feature score of 6 or higher;
(iii) DSSE > 0 kcal/mol (asymmetry rule);
(iv) AIS > −8.6 kcal/mol (cleavage site instability rule);
(v) 30% ≤ GC% ≤ 60%;
(vi) exclusion of target sequence that has at least one occurrence of AAAA, CCCC, GGGG, or UUUU.

The antisense siRNA binding energy is a weighted sum of the RNA/RNA stacking energies (7) for the hybrid formed by the antisense siRNA and the targeted sequence. For a base-pair stack, the weight for the sum is calculated by the probability of the unpaired dinucleotide in the target sequence that is involved in the stack. In addition, an A–U terminal penalty for the hybrid is included and is weighted by the probability of the unpaired terminal base. This weighting scheme accounts for the structural variation at the target site. For example, an siRNA with an antisense binding energy of −15 kcal/mol is predicted to be more effective than an siRNA with a binding energy of −10 kcal/mol. The target accessibility rule is implemented by requiring the siRNA binding energy to be below a threshold value. The current default of the threshold is −10 kcal/mol.

The siRNA duplex feature score is computed with the algorithm by Reynolds and colleagues (18) and has a minimum of −2 points and a maximum of 10 points. The 'asymmetry rule', i.e. that the 5′-antisense end is less stable than the 5′-sense end, is enforced by DSSE > 0 kcal/mol. The rule of relative instability at the cleavage site is enforced by AIS > −8.6 kcal/mol, the midpoint between the minimum (−3.6 kcal/mol) and the maximum (−13.6 kcal/mol). In addition, based on antisense

siRNA binding energy, a target accessibility score is computed and it ranges between 0 and 8 points. A duplex thermodynamics score is computed and it can have value 0, 1 or 2, with 1 point contributed by DSSE > 0 (kcal/mol), and another point by AIS > −8.6 (kcal/mol). The total siRNA score is the sum of accessibility score, duplex feature score and duplex thermodynamics score. The maximum total score is 20 points.

The output file filtered.out presents siRNAs that meet all filter criteria. The file sirna_s.out provides output information for siRNAs with total score greater or equal to a preset threshold. The current threshold is 12 points. The file sirna.out contains output information for all siRNAs. Results for DSSE, AIS and other siRNA duplex stability indices are provided in the output file stability.out.

siRNA internal stability profiling. On the output page for *S*irna, the internal stability profile for every possible siRNA duplex is available through the interactive graphic window for siRNA ends and internal stability profiling. On the profile, the user can use the stabilities for positions 1 and 19 to make a comparison of the stabilities of siRNA duplex ends. These files are not included in the Zip or tar.gz file for *S*fold output, because of the large number of profiles for long mRNAs. However, from the interactive profiling window, the user can download the profiles for the selected target sites. The internal stability profile for target positions nt 44–64 for the rabbit β-globin mRNA is illustrated in Figure 2. The 5′-antisense end is less stable than the 5′-sense end (DSSE = 1.6 kcal/mol), and the siRNA is relatively unstable at the cleavage site, with AIS = −7.8 kcal/mol.

Soligo module

The *S*oligo module provides tools for the rational design of antisense oligos by combining prediction of target secondary structure and accessibility with empirical design rules. Probability profiling for the prediction of accessible regions is the same as described above for the *S*irna module. The file oligo_f.out gives filtered output for antisense oligos of user-specified length. *S*oligo currently uses the following filters:

  (i) antisense oligo binding energy ⩽−8 kcal/mol;
 (ii) 40% ⩽ GC % ⩽ 60%;
(iii) No GGGG in the target sequence.

The antisense oligo binding energy is a weighted sum of the DNA/RNA stacking energies (21) for the hybrid formed by the antisense oligo and the targeted sequence. For a base-pair stack, the weight for the sum is calculated by the probability of the unpaired dinucleotide in the target sequence that is involved in the stack. This weighting scheme accounts for the structural variation at the target site.

Sribo module

This module currently offers tools for the design of hammerhead ribozymes based on target accessibility prediction. For every site of the selected cleavage triplet (e.g. GUC) on the target RNA, the probability profile for individual bases (i.e. window width $W = 1$ for ribozyme applications) is produced for the region that includes the triplet and the two flanking sequences of 15 bases each. The length of 15 bases is sufficient to cover the normal range of lengths for the binding arms of the hammerhead ribozyme. We recommend selection of cleavage sites for which both flanking sequences are at least partially accessible. This is because antisense hybridization is believed to start with nucleation at a site of several unpaired bases, and elongation then occurs by 'unzipping' the adjacent helix on the target (22). As an example, the GUC triplet at positions 87–89 for the rabbit β-globin mRNA meets this selection criterion (Figure 3). We note that profiling of the target RNA only addresses the accessibility of the target. However, it is also important to assess the folding of a designed ribozyme whose binding arms are determined by the cleavage triplet and its flanking sequences. To address the issue of ribozyme folding, the user can run module *S*rna to fold the ribozyme. Diagrams of representative structures and associated sampling frequencies
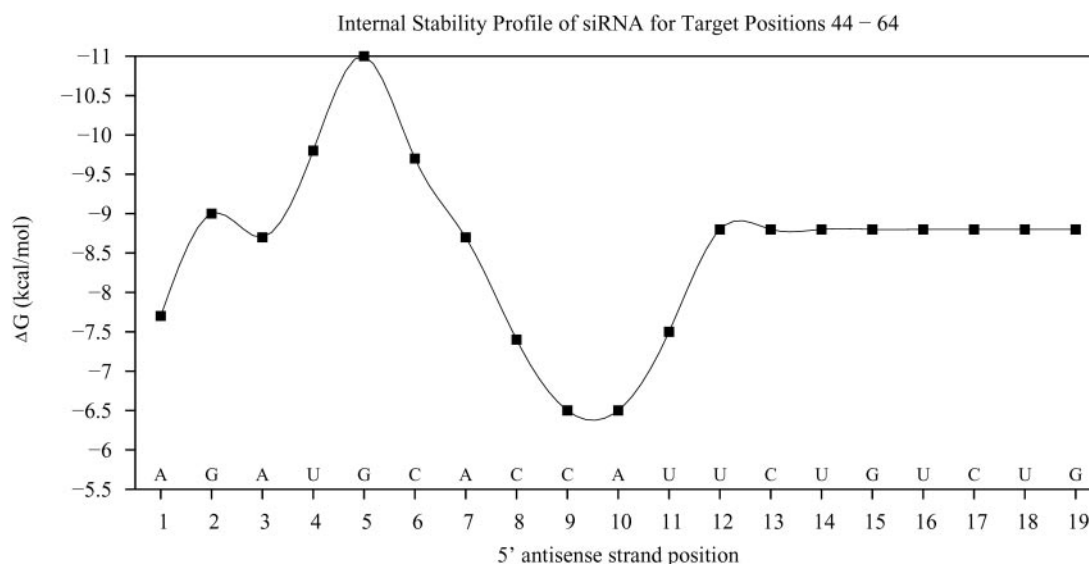

Figure 2. Internal stability profile (15) for the siRNA targeting positions 44–64 of the rabbit β-globin mRNA.
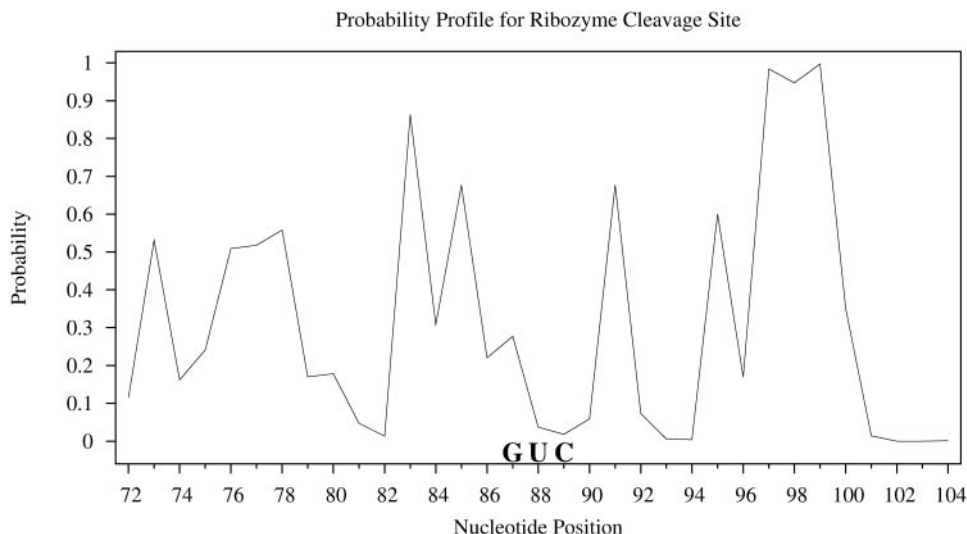
Probability Profile for Ribozyme Cleavage Site

Figure 3. Probability profile (window width $W = 1$ for ribozyme applications) for the region of the cleavage triplet $G^{87}UC^{89}$ and its two flanking sequences of 15 each for the rabbit β-globin mRNA.

from this module are helpful in providing a confidence assessment about the degree of correct ribozyme folding.

Srna module

This module provides tools and sampling statistics to statistically characterize the Boltzmann ensemble of RNA secondary structures. A two-dimensional histogram (2Dhist) displays base pair probabilities computed from a statistical sample with a default size of 1000 structures. In 2Dhist, base pair probabilities are shown by solid squares in the upper left triangle, with the nucleotide positions on both axes. The areas of the solid squares are proportional to the frequencies of the base pairs in the sampled structures. 2Dhist has an option for the display of base pair probabilities. When this option is selected, the probability, and the nucleotide positions of the base pair to which a solid square corresponds, can be shown through mouse pointing (Figure 4). For long RNA sequences, the base pair probabilities take some time to load after the 2Dhist is displayed. We thus have set 'no base pair probabilities' as the default display.

An *ad hoc* representation of the structure sample is given in a table format. First, the minimum free energy of structures in the sample (SMFE) and the largest free energy (LFE) in the sample are computed. The free energy range covering all structures in the sample (i.e. [SMFE, LFE]) is then divided into 10 equally spaced intervals. For each free energy interval, the structure with the lowest free energy is selected as the representative. For the representative, the table presents its associated free energy interval, the frequency with which structures in the sample fall into the energy interval (the frequencies for the 10 intervals sum to 1), the free energy of the structure and the secondary structural diagram (Figure 5). We note that this is a rather crude representation of the structure sample and the Boltzmann ensemble, mainly because structures in a common free energy interval can possess structural features substantially different from one another. Structure diagrams are generated using a modified version of the naview program (24). In our modified version, a number of bugs have been fixed to reduce the chance of overlapping between
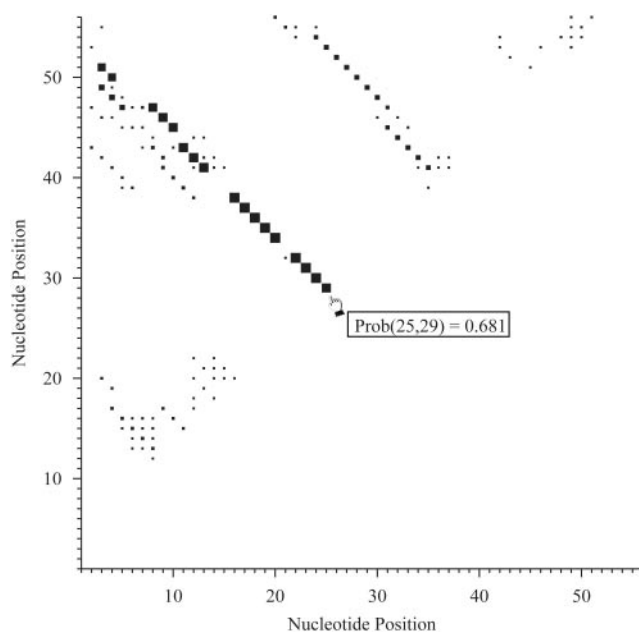

Figure 4. Two-dimensional histograms (2Dhist) for a sample of 1000 structures generated for *Leptomonas collosoma* spliced leader RNA of 56 nt in length (23). As an example of the option of 'with base pair probabilities', the sampling estimate of the probability for the base pair between nt 25 and 29 is shown to be 0.681 when the mouse is pointed at the solid square positioned at (25,29).

diagram components. A Perl script was written to convert the device-independent .plt2 output from naview into a structural diagram in PostScript format. The structural diagram is also available in PNG and PDF formats. The PNG format has capabilities for enlargement and local display. The GCG connect file is also provided.

A text file of base pair frequencies for generating 2Dhist and a file of free energies of sampled structures for plotting free energy distributions (5) are available. These files are described in more detail in the online manual.
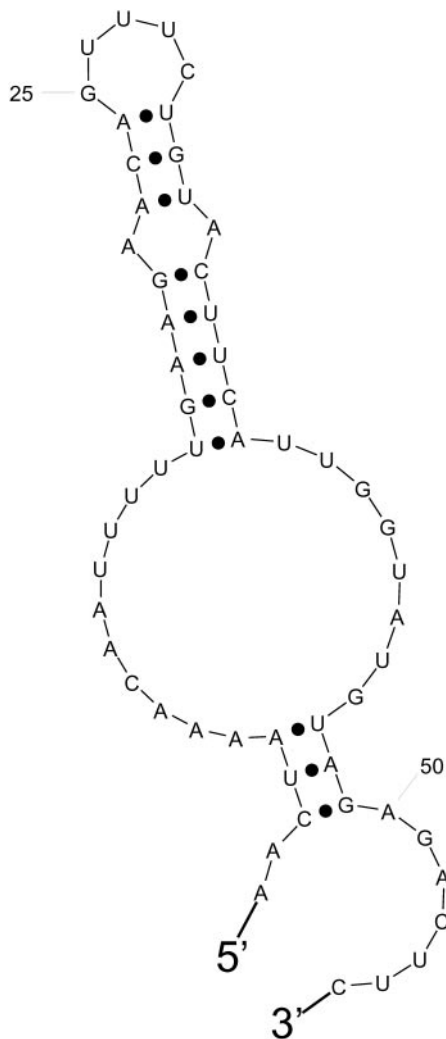
Figure 5. Secondary structure diagram generated by a modified version of the naview program (24) for a predicted structure of *Leptomonas collosoma* spliced leader RNA.

FUTURE PLANS

We have ongoing collaborations with several experimental labs to validate and potentially improve the design methods for RNA-targeting nucleic aids. Additional design tools for modules *S*rna, *S*oligo and *S*ribo may be developed from these studies. For *S*rna, an appealing method for an efficient statistical representation of the Boltzmann ensemble of RNA secondary structures is to cluster the structures in the sample generated by our algorithm (5). We have made substantial progress toward algorithmic determination of distinct clusters. We expect to include cluster representation in the output for the *S*rna module in the foreseeable future. The ability to allow folding constraints is an important feature for future development. DNA folding prediction using DNA parameters (25,26) will enable the development of modules for other applications such as rational design of PCR primers and amplicons. Furthermore, oligo self-interactions (27) may be considered in the development of additional screening filters for *S*oligo. Alternative programs for RNA structure display will be explored for improvement in visualization, colored annotation

based on sampling statistics, and the potential of user interaction through Java. The feasibility of including a standardized RNA structure format to facilitate data exchange, e.g. the RNAML format (28), is under consideration. The Wadsworth Center server is hosted on a web cluster currently consisting of three 1U servers each with dual Athlon MP 2800+ processors. The Grid Engine software is used as the job management system in this web cluster. The master node has 2 GB of memory while each of the two execution nodes has 4 GB of memory. The job load is closely monitored, and the server hardware will be upgraded when this becomes necessary.

CITING THE *S*FOLD WEB SERVER

In research publications, users of *S*fold should cite this article and the papers describing the algorithms (5,13), in additional to including the URL for the main server web site: http://sfold.wadsworth.org. Future relevant articles for citation will be listed on the server's front page.

ACKNOWLEDGEMENTS

REFERENCES

1. Zuker,M. and Stiegler,P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.*, **9**, 133–148.
2. Zuker,M. (1989) On finding all suboptimal foldings of an RNA molecule. *Science*, **244**, 48–52.
3. Wuchty,S., Fontana,W., Hofacker,I.L. and Schuster,P. (1999) Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, **49**, 145–165.
4. McCaskill,J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105–1119.
5. Ding,Y. and Lawrence,C.E. (2003) A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res.*, **31**, 7280–7301.
6. Christoffersen,R.E., McSwiggen,J.A. and Konings,D. (1994) Application of computational technologies to ribozyme biotechnology products. *J. Mol. Structure (Theochem)*, **311**, 273–284.
7. Xia,T., SantaLucia,J.Jr, Burkard,M.E., Kierzek,R., Schroeder,S.J., Jiao,X., Cox,C. and Turner,D.H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry*, **37**, 14719–14735.
8. Mathews,D.H., Sabina,J., Zuker,M. and Turner,D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.
9. Lee,N.S., Dohjima,T., Bauer,G., Li,H., Li,M.J., Ehsani,A., Salvaterra,P. and Rossi,J. (2002) Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. *Nature Biotechnol.*, **20**, 500–505.

10. Bohula,E.A., Salisbury,A.J., Sohail,M., Playford,M.P., Riedemann,J., Southern,E.M. and Macaulay,V.M. (2003) 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.*, **278**, 15991–15997.

11. Far,R.K. and Sczakiel,G. (2003) The activity of siRNA in mammalian cells is related to structural target accessibility: a comparison with antisense oligonucleotides. *Nucleic Acids Res.*, **31**, 4417–4424

12. Vickers,T.A., Koo,S., Bennett,C.F., Crooke,S.T., Dean,N.M. and Baker,B. (2003) Efficient reduction of target RNAs by small interfering RNA and RNase H-dependent antisense agents. A comparative analysis. *J. Biol. Chem.*, **278**, 7108–7118.

13. Ding,Y. and Lawrence,C.E. (2001) Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond. *Nucleic Acids Res.*, **29**, 1034–1046.

14. Schwarz,D.S., Hutvagner,G., Du,T., Xu,Z., Aronin,N. and Zamore,P.D. (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, **115**, 199–208.

15. Khvorova,A., Reynolds,A. and Jayasena,S.D. (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell*, **115**, 209–216.

16. Zhao,J.J. and Lemke,G. (1998) Rules for Ribozymes. *Mol. Cell Neurosci.*, **11**, 92–97.

17. Hardin,C.C., Watson,T., Corregan,M. and Bailey,C. (1992) Cation-dependent transition between the quadruplex and Watson–Crick hairpin forms of d(CGCG3GCG). *Biochemistry*, **31**, 833–841.

18. Reynolds,A., Leake,D., Boese,Q., Scaringe,S., Marshall,W.S. and Khvorova,A. (2004) Rational siRNA design for RNA interference. *Nature Biotechnol.*, **22**, 326–330.

19. Ui-Tei,K., Naito,Y., Takahashi,F., Haraguchi,T., Ohki-Hamazaki,H., Juni,A., Ueda,R. and Saigo,K. (2004) Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res.*, **32**, 936–948

20. Amarzguioui,M. and Prydz,H. (2004) An algorithm for selection of functional siRNA sequences. *Biochem. Biophys. Res. Commun.* **316**, 1050–1058.

21. Sugimoto,N., Nakano,S., Katoh,M., Matsumura,A., Nakamuta,H., Ohmichi,T., Yoneyama,M. and Sasaki,M. (1995) Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry*, **34**, 11211–11216.

22. Milner,N., Mir,K.U. and Southern,E.M. (1997). Selecting effective antisense reagents on combinatorial oligonucleotide arrays. *Nature Biotechnol.*, **15**, 537–541.

23. LeCuyer,K.A. and Crothers,D.M. (1993) The *Leptomonas collosoma* spliced leader RNA can switch between two alternate structural forms. *Biochemistry*, **32**, 5301–5311.

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

25. SantaLucia,J.Jr (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA.*, **95**, 1460–1465.

26. Owczarzy,R., Vallone,P.M., Goldstein,R.F. and Benight,A.S. (1999) Studies of DNA dumbbells VII: evaluation of the next-nearest-neighbor sequence-dependent interactions in duplex DNA. *Biopolymers*, **52**, 29–56.

27. Matveeva,O.V., Mathews,D.H., Tsodikov,A.D., Shabalina,S.A., Gesteland,R.F., Atkins,J.F. and Freier,S.M. (2003) Thermodynamic criteria for high hit rate antisense oligonucleotide design. *Nucleic Acids Res.*, **31**, 4989–4994.

28. Waugh,A., Gendron,P., Altman,R., Brown,J.W., Case, D., Gautheret, D., Harvey, S.C., Leontis,N., Westbrook,J., Westhof,E., Zuker,M. and Major,F. (2002) RNAML: a standard syntax for exchanging RNA information. *RNA*, **8**, 707–717.

29. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.