D280–D282 *Nucleic Acids Research, 2010, Vol. 38, Database issue* *Published online 30 October 2009*
doi:10.1093/nar/gkp892

NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure

Douglas H. Turner[1] and David H. Mathews[2,*]

[1]Department of Chemistry and Center for RNA Biology, Box 0216, University of Rochester, Rochester, NY 14627-0216 and [2]Department of Biochemistry and Biophysics and Center for RNA Biology, Box 712, University of Rochester Medical Center, Rochester, NY 14642, USA

Received August 17, 2009; Revised October 4, 2009; Accepted October 6, 2009

ABSTRACT

The Nearest Neighbor Database (NNDB, http://rna.urmc.rochester.edu/NNDB) is a web-based resource for disseminating parameter sets for predicting nucleic acid secondary structure stabilities. For each set of parameters, the database includes the set of rules with descriptive text, sequence-dependent parameters in plain text and html, literature references to experiments and usage tutorials. The initial release covers parameters for predicting RNA folding free energy and enthalpy changes.

INTRODUCTION

Nearest neighbor approaches were developed to predict the folding stabilities of nucleic acid secondary structures (1). These parameter sets utilize empirical rules, generally derived from optical melting experimental data, as the basis of the predictions. For RNA, rules exist for predicting both free energy and enthalpy change of Watson–Crick helices, GU pairs and loops (2–5). Parameters for DNA have also been assembled for predicting Watson–Crick pair free energy and enthalpy change and free energy changes of loops (6,7). These parameter sets are the basis of computer programs that predict low free energy secondary structures. Such programs include Mfold/UnaFold (8,9), the Vienna RNA package (10), RNA structure (2), RNAsoft (11) and Sfold (12). Additional approaches that use statistical learning of parameters for RNA folding have also used the rules from the nearest neighbor methods and derived new parameter values (13,14).

Nearest neighbor parameter sets include both a set of rules, called either equations or features, for predicting stability and a set of parameter values used by the equations (14). For RNA, separate rules exist for predicting stabilities of helices, hairpin loops, small internal loops, large internal loops, bulge loops, multi-branch loops, exterior loops and pseudoknots. Given the number of rules and constraints on the length of journal publications, it is difficult to assemble all the parameters in one publication and provide meaningful tutorials for using the parameters. This is a barrier to software development for novel algorithms that could take advantage of the parameters. For example, many software packages that use RNA parameters still implement the set of parameters assembled in 1999 (4), in spite of the fact the RNA parameters were updated in 2004 (2) based on experimental results.

The Nearest Neighbor Database (NNDB) is a web-based tool for assembling and archiving complete nearest neighbor sets, including rules and values. It is available online at http://rna.urmc.rochester.edu/NNDB. It provides documentation of parameter sets and tutorials on how to apply the parameters. Currently, the 1999 and 2004 sets of RNA folding parameters are provided (2–5).

WEBSITE ORGANIZATION

The NNDB is built using a set of static html, specifically XHTML 1.0 transitional pages with a page hierarchy shown in Figure 1. Text is encoded in Unicode (utf-8) to facilitate display of equations in pages with diverse browsers running on diverse operating systems. The top-level page provides access to a help page, available parameter sets and a page of references to RNA optical melting experiments. Additionally, links provide downloading of the whole database in either zip or gzipped tar format. The help page introduces the purpose of the database and defines basic terms, including the set of structural features defined by secondary structures. For example, Figure 2, from the help page, shows an RNA secondary structure that illustrates the loop features covered by nearest neighbor parameter sets. The basic equations for utilizing the parameters to extrapolate folding free energy

*To whom correspondence should be addressed. Tel: +1 585 275 1734; Fax: +1 585 275 6007; Email: david_mathews@urmc.rochester.edu

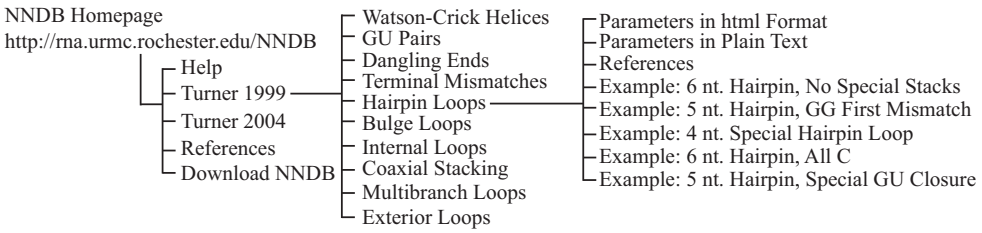

Figure 1. The webpage hierarchy of the NNDB. This figure illustrates the page hierarchy by following the linked pages down through the 1999 parameters and down to the hairpin loop pages. Note that there are five example calculations for hairpin loops to illustrate the separate sequence-dependent rules that are used depending on the specific loop.

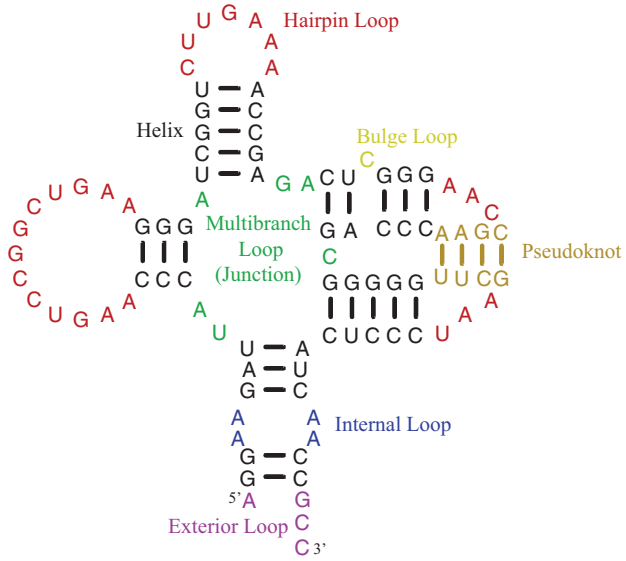

Figure 2. An RNA secondary structure illustrating the types of features included in nearest neighbor parameter sets. This figure appears on the help page of the website. Loops are composed of nucleotides not in canonical pairs. Hairpin loops have one exiting helix. Internal and bulge loops have two exiting helices. Internal loops have nucleotides not in canonical pairs on each of two strands, but bulge loops have nucleotides not in canonical pairs on only one strand. Multibranch loops, also called helical junctions, have three or more exiting helices. Exterior loops contain the ends of sequences and one or more exiting helices. Pseudoknots are canonical pairs connecting loop regions closed by other helices. Formally, a pseudoknot occurs when there are at least two pairs, with indices i paired to j and i' paired to j', that satisfy the condition $i < i' < j < j'$. The pseudoknot helix is often considered to be composed of the fewest pairs that need to be removed to relieve the pseudoknot (19). In this structure, the tan nucleotides are in pairs that could be removed to relieve the pseudoknot.

changes to temperatures other than 37°C and to predict melting temperatures are also provided.

For each set of parameters, a first page introduces the available parameters, which vary from set to set. For example, the 1999 RNA rules predict only folding free energy changes (4), but the 2004 rules can be used to predict both folding free energy and enthalpy changes (2,5). For each structural feature, a page defines the basic equations and provides links to parameter values (in plain text and html), references and tutorial pages (e.g. Figure 3). The number of tutorials varies from feature to feature; the set of tutorials is designed to cover each type of rule that can be encountered in practice. For example, the Watson–Crick helix parameters are covered with two tutorials, one for self-complementary and one for non-self-complementary strands. These two tutorials also demonstrate the difference in the calculation when there are terminal AU base pairs, which receive a free energy and enthalpy change penalty (3), because the self-complementary duplex example has two terminal AU pairs and the non-self-complementary case has no terminal AU pairs.

The individual pages are designed for ease of navigation and clarity. Individual pages above the level of value tables have top banner, a left navigation bar that allows the user to navigate back up the hierarchy to any level above and a bottom bar with the date of last editing. For pages edited after the database has gone online, previous versions of the page are available using this bottom content bar. To facilitate indexing by search engines, all pages have a descriptive title, including the set of parameters to which it belongs (if applicable).

WEBSITE CONTENT

The first release of the NNDB contains the RNA folding rules assembled in 1999 and 2004 (2–5). These rules represent the most recent set of parameters and a prior set that is widely used in software packages. Because folding rules are derived to work as a set, the two versions of rules and values should not be mixed and the website hierarchy reinforces this.

The website is designed to be expandable to additional sets of parameters. It is anticipated, for example, that additional pages will be written to include nearest neighbors for DNA folding (6,7) and for predicting RNA pseudoknot stabilities (15–18). Additionally, the values derived from the re-estimation of the values of the 1999 parameter set using the set of known RNA secondary structures will also be included (14).

DISCUSSION

The NNDB is designed to provide a convenient location for assembling parameter sets for predicting the stability of nucleic acid secondary structures. It is modular in design, which facilitates its future expansion to contain additional parameter sets. Furthermore, the web format

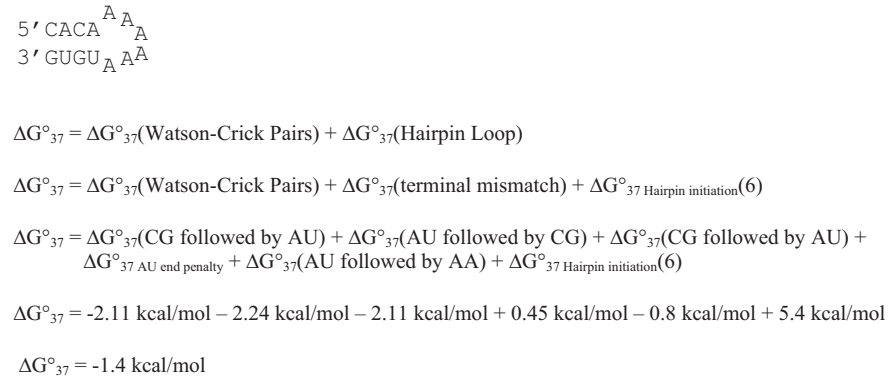

Figure 3. An example tutorial from the database. This tutorial demonstrates the prediction of folding free energy change for a hairpin loop of six unpaired nucleotides using the 2004 parameters (2,3).

makes it feasible to provide extensive tutorials for utilizing the parameters, which is generally not possible in print.

FUNDING

The creation of the NNDB was supported by United States National Institutes of Health grants GM076485 to D.H.M. and GM22939 to D.H.T. Funding for open access charge: United States National Institutes of Health.

REFERENCES

1. Tinoco,I. Jr, Borer,P.N., Dengler,B., Levin,M.D., Uhlenbeck,O.C., Crothers,D.M. and Gralla,J. (1973) Improved estimation of secondary structure in ribonucleic acids. *Nat. New Biol.*, **246**, 40–41.
2. Mathews,D.H., Disney,M.D., Childs,J.L., Schroeder,S.J., Zuker,M. and Turner,D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl Acad. Sci. USA*, **101**, 7287–7292.
3. 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 pairs. *Biochemistry*, **37**, 14719–14735.
4. Mathews,D.H., Sabina,J., Zuker,M. and Turner,D.H. (1999) Expanded sequence dependence of thermodynamic parameters provides improved prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.
5. Lu,Z.J., Turner,D.H. and Mathews,D.H. (2006) A set of nearest neighbor parameters for predicting the enthalpy change of RNA secondary structure formation. *Nucleic Acids Res.*, **34**, 4912–4924.
6. SantaLucia,J. Jr (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA*, **95**, 1460–1465.
7. SantaLucia,J. and Hicks,D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 415–440.
8. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
9. Zuker,M., Mathews,D.H. and Turner,D.H. (1999) In Barciszewski,J. and Clark,B.F.C. (eds), *RNA Biochemistry and Biotechnology*. Kluwer Academic Publishers, Boston, pp. 11–43.
10. Hofacker,I.L., Fontana,W., Stadler,P.F., Bonhoeffer,L.S., Tacker,M. and Schuster,P. (1994) Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, **125**, 167–168.
11. Andronescu,M., Aguirre-Hernandez,R., Condon,A. and Hoos,H.H. (2003) RNAsoft: a suite of RNA secondary structure prediction and design software tools. *Nucleic Acids Res.*, **31**, 3416–3422.
12. Ding,Y., Chan,C.Y. and Lawrence,C.E. (2004) Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.*, **32**, W135–W141.
13. Do,C.B., Woods,D.A. and Batzoglou,S. (2006) CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics*, **22**, e90–e98.
14. Andronescu,M., Condon,A., Hoos,H.H., Mathews,D.H. and Murphy,K.P. (2007) Efficient parameter estimation for RNA secondary structure prediction. *Bioinformatics*, **23**, i19–i28.
15. Dirks,R. and Pierce,N. (2003) A partition function algorithm for nucleic acid secondary structure including pseudoknots. *J. Comput. Chem.*, **24**, 1664–1677.
16. Gultyaev,A.P., van Batenburg,F.H.D. and Pleij,C.W.A. (1999) An approximation of loop free energy values of RNA H-pseudoknots. *RNA*, **5**, 609–617.
17. Cao,S. and Chen,S.J. (2006) Predicting RNA pseudoknot folding thermodynamics. *Nucleic Acids Res.*, **34**, 2634–2652.
18. Cao,S. and Chen,S.J. (2009) Predicting structures and stabilities for H-type pseudoknots with interhelix loops. *RNA*, **15**, 696–706.
19. Smit,S., Rother,K., Heringa,J. and Knight,R. (2008) From knotted to nested RNA structures: a variety of computational methods for pseudoknot removal. *RNA*, **14**, 410–416.