RNA pseudoknots: folding and finding
Biao Liu¹, David H Mathews²,³ and Douglas H Turner¹,³*

Addresses: ¹Department of Chemistry, 120 Trustee Road, University of Rochester, Rochester, NY 14627, USA; ²Department of Biochemistry and Biophysics, School of Medicine and Dentistry, 601 Elmwood Ave, University of Rochester, Rochester, NY 14642, USA; ³Center for RNA Biology, School of Medicine and Dentistry, University of Rochester, Rochester, NY 14642, USA

¹ Corresponding author: Douglas H Turner (turner@chem.rochester.edu)

Abstract
RNA pseudoknots are important for function. Three-dimensional structural information is available, insights into factors affecting pseudoknot stability are being reported, and computer programs are available for predicting pseudoknots.

Introduction and context
RNA pseudoknots are important for many functions [1-3]. Pseudoknots are formed by pairing between bases in a loop region and complementary bases outside the loop (Figure 1). That is, for any base pairs i-j and k-l, i<j, k<l, and i<k, there are cases in which i<k<j<l. A two-loop nomenclature was initially used, but 30.9% of pseudoknots now listed in PseudoBase [4] have an additional loop, so the three-loop nomenclature from Brierley et al. [1] is more generally applicable (Figure 1). Loop 2 typically has zero or one nucleotide, whereas loops 1 and 3 and the stems are more variable (Table 1).

Major recent advances
Folding of pseudoknots
Quite a few three-dimensional (3D) structures have been determined for isolated pseudoknots of fewer than 50 nucleotides (Table 1). Some tertiary interactions are conserved in particular classes and are essential for biological activity [2,3]. Examples include a quadruple-base interaction in pseudoknots from Luteoviridae viruses [2,5-9] and triplexes in telomerase RNA [3] and in viral mRNA that undergo -1 frameshifting [10]. Structures of natural pseudoknots bound to small molecules are being reported [11-13], and structures of larger RNAs are revealing long-range pseudoknots [14-20].

The stability of RNA pseudoknots is a key factor determining structure-function relationships [10] and is important for predicting RNA structure [21-32]. Thermodynamic measurements have started appearing [3,33-36].

A statistical polymer model for loops [21,22] coupled with the INN-HB (Individual Nearest Neighbor-Hydrogen Bonding) model for stems [37] allows estimates of the stability of small pseudoknots. Coaxial stacking of the two stems can be included, although this is not always observed in 3D structures [5]. Contributions from tertiary interactions between the loops and stems are neglected because little is known about their thermodynamics.

Mechanical unfolding of single molecules by optical tweezers [10,38,39] reveals that frameshifting efficiency is highly related to the mechanical stability of pseudoknots, as suggested from cryoelectron microscopy [40] and prediction by a statistical polymer model [41]. Single-molecule experiments also indicate that pseudoknot folding and unfolding at low forces are stepwise [42] and that the presence of Mg²⁺ stabilizes the pseudoknot more than hairpins [39]. A better understanding of pseudoknot thermodynamic and mechanical stability and of folding dynamics will help reveal structure-function relationships [43].

Finding pseudoknots
Finding pseudoknots by computationally folding RNA sequences is a difficult problem. Because of
computational cost, most of the popular dynamic programming algorithms for predicting the lowest free energy structure do not allow pseudoknots. With state-of-the-art knowledge, finding the lowest free energy structure with pseudoknots takes an exponentially increasing amount of time as the sequence gets longer; that is, the problem is NP-complete [23,24].

To predict low free energy structures with pseudoknots, roughly four different practical approaches are used by available programs. The first approach is to apply stochastic methods either to simulate folding pathways or to sample structures [25,26,44]. With these algorithms, structures are revised according to an element of chance and new pairs that are pseudoknotted with existing pairs can be added. A variation on this theme follows a folding pathway to find low free energy structures but is deterministic in its choices of stems [27]. The second approach is to use a dynamic programming algorithm in which the possible topologies of the predicted structures are limited [28-30]. The possible topologies predicted by a number of different programs have been examined [31]. A third approach is to assemble structures from component base pairs using a graph-theoretic approach [45,46]. A fourth approach is to iteratively build structures using algorithms that cannot predict pseudoknots with a single iteration [47,48]. The Nuclear magnetic resonance (NMR)-Assisted Prediction of RNA Secondary Structure (NAPSS) algorithm is an iterative approach that includes constraints from simple NMR experiments to improve predictions [32].

A number of the programs cited above either require multiple homologous sequences or are capable of using them to find a conserved structure [44-46,48]. These programs should be more accurate at structure prediction than those that use a single sequence because of the additional information available in the multiple

Table 1. Three-dimensional structures of short pseudoknots

<table>
<thead>
<tr>
<th>Type of Organism or RNA</th>
<th>RNA Function</th>
<th>Method</th>
<th>PDB #</th>
<th>Ref.</th>
<th>Size, nts</th>
<th>Stem 1, bps</th>
<th>Loop 1, nts</th>
<th>Stem 2, bps</th>
<th>Loop 2, nts</th>
<th>Loop 3, nts</th>
<th>Coaxial stack or bend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammalian retrovirus</td>
<td>Simian retrovirus-1 (SRV-1) Ribosomal frameshifting</td>
<td>NMR</td>
<td>IE95 [51]</td>
<td>36</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>12</td>
<td>Coaxial stack</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mouse mammary tumor virus (MMTV)</td>
<td>NMR</td>
<td>IKPD [56]</td>
<td>32</td>
<td>5</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plant Luteoviridae</td>
<td>Beet western yellow virus (BWYY)</td>
<td>X-ray</td>
<td>437D [5]</td>
<td>26</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>Bend</td>
<td></td>
</tr>
<tr>
<td>viruses</td>
<td>Sugarcane yellow leaf virus (SqYLV)</td>
<td>NMR</td>
<td>1L2X [9]</td>
<td>28</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Potato leaf roll virus (PLRV)</td>
<td>X-ray</td>
<td>2AP0 [2]</td>
<td>26</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pea enation mosaic virus (PEMV)</td>
<td>NMR</td>
<td>1IY4 [7]</td>
<td>33</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>Telomerase RNA Ribonuclease maintenance</td>
<td>NMR</td>
<td>1YMO [3]</td>
<td>46</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>0</td>
<td>8</td>
<td>Slight bend</td>
<td></td>
</tr>
<tr>
<td></td>
<td>t-RNA like</td>
<td>2K96 [43]</td>
<td>23</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td></td>
<td>Coaxial stack</td>
<td></td>
</tr>
<tr>
<td>Plant virus (tymovirus)</td>
<td>Turnip yellow mosaic virus (TYMV)</td>
<td>Telomere</td>
<td>3PHP [59]</td>
<td>21</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>Bend</td>
<td></td>
</tr>
<tr>
<td>3' NCR</td>
<td>Aquifex aeolicus tmRNA</td>
<td>Trans-translation</td>
<td>1A60 [60]</td>
<td>2G1W [61]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

bps, base pairs; NCR, non-coding region; NMR, nuclear magnetic resonance; nts, nucleotides; PDB, Protein Data Bank; Ref., reference.
sequences. For example, pairs that are not conserved in the set of sequences can be excluded from the final structure.

A recent benchmark of the accuracy of structure prediction using single sequences provides guidance for choosing a program [47]; Table 2 lists programs available for free. For 12 RNA sequences of 210 nucleotides or longer, a pathway folding algorithm [25] had the highest accuracy. The iterative approaches also performed similarly and were time-efficient [47,48].

**Future directions**

Predictions of the occurrence and stabilities of pseudoknots can be improved. The best-performing program tested by Ren et al. [47] predicted only 57% of known canonical base pairs, and only 39% of the predicted pairs were in the known structures. New insights into the structures and stabilities will suggest computational simplifications for existing strategies and improve approximations of stabilities. New methods to find pseudoknots conserved in multiple sequences and to incorporate more experimental data will restrain folding space. Finding more pseudoknots will likely expand the set of sequences can be excluded from the final structure.

This table provides a list of programs that are available for free and can make predictions on a desktop computer for sequences that are up to at least 200 nucleotides long. Note that hXmatch is designed to find conserved structures for sequences in a multiple alignment and that ILM can be run with one sequence or with an alignment of multiple sequences. SimulFold works on a set of homologous sequences and infers the sequence alignment. KineFold and vsfold5 are available as web servers. pknotsRG is available for use on a web server or for download.

### Abbreviations

3D, three-dimensional; INN-HB, individual nearest neighbor-hydrogen bonding; NAPSS, nuclear magnetic resonance (NMR)-assisted prediction of RNA secondary structure; NMR, nuclear magnetic resonance; SRV-1, simian retrovirus-1.

### Competing interests

The authors declare that they have no competing interests.

### Acknowledgments

This work was supported by National Institutes of Health grants GM22939 (DHT) and R01GM076485 (DHM).

### References


F1000 Factor 3.0 Recommended Evaluated by Kathleen Hall 22 Feb 2008


F1000 Factor 8.0 Exceptional Evaluated by Robert Batey 08 Apr 2008, Douglas Turner 14 Jul 2008


F1000 Factor 9.6 Exceptional Evaluated by Douglas Turner 09 Jul 2004, David Mj Lilley 12 Jul 2004


F1000 Factor 6.0 Must Read Evaluated by John Burke 06 Dec 2004


F1000 Factor 3.2 Recommended Evaluated by Niles Lehman 05 Jun 2009, Douglas Turner 24 Jun 2009


F1000 Factor 6.0 Must Read Evaluated by P Shing Ho 05 Apr 2007


F1000 Factor 6.0 Must Read Evaluated by Samuel Butcher 17 May 2007


50. ten Dam EB, Verlaan PW, Pleij CW: Analysis of the role of the pseudoknot component in the SRV-1 gag-pro ribosomal frameshift signal: loop lengths and stability of the stem regions. RNA 1995, 1:46-54.


