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

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

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

| Type of Organism or RNA | Function | Method   | PDB #     | Ref. | Size, nts | Stem 1, bps | Loop 1, nts | Stem 2, bps | Loop 2, nts | Loop 3, nts | Coaxial stack or bend |
|------------------------|----------|----------|-----------|------|-----------|-------------|-------------|-------------|-------------|-------------|-----------------------|
| Mammalian retrovirus   | Ribosomal frameshifting | NMR | IE95 [51] | 36   | 6         | 1           | 6           | 0           | 12          |            | Coaxial stack          |
| Mammalian retrovirus   |          | NMR | IKPD [56] | 32   | 5         | 2           | 6           | 1           | 8           |            |                       |
| Mammalian retrovirus   | Mouse mammary tumor virus (MMTV) | NMR | IRNK [58] | 46   | 6         | 1           | 6           | 0           | 12          |            |                       |
| Plant Luteoviridae     | Beetal western yellow virus (BWYY) | X-ray | 437D [5] | 26   | 5         | 2           | 3           | 1           | 7           |            | Bend                  |
| Plant Luteoviridae     | Sugarcane yellow leaf virus (ScYLV) | NMR | 2AP0 [8] | 28   | 5         | 2           | 3           | 1           | 9           |            |                       |
| Plant Luteoviridae     | Potato leaf roll virus (PLRV) | X-ray | 1YG4 [2] | 26   | 4         | 2           | 3           | 1           | 9           |            |                       |
| Plant Luteoviridae     | Pea enation mosaic virus (PehV) | NMR | 1KPY [6] | 33   | 5         | 2           | 3           | 1           | 7           |            |                       |
| Human                  | Telomerase RNA | Telomere maintenance | NMR | IYMO [3] | 46   | 6         | 8           | 9           | 0           | 8           | Slight bend            |
| Plant virus (tymovirus) | Turnip yellow mosaic virus (TYMV) | NMR | 2K96 [43] | 23   | 3         | 4           | 5           | 0           | 3           |            | Coaxial stack          |
| Plant virus (tymovirus) | 3' NCR | NMR | IYMO [43] | 23   | 3         | 4           | 5           | 0           | 3           |            |                       |
| Bacteria               | Aquifex aeolicus tmRNA | Trans-translation | NMR | 2GIW [61] | 21   | 4         | 1           | 3           | 1           | 5           | Bend                  |

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. Future work can also resolve current ambiguities. For example, there is disagreement about the role of pseudoknot stability in human telomerase RNA. The pseudoknot domain may act as a molecular switch [52], in which the pseudoknot and the stem 1 hairpin have nearly equal stability [34], or only the pseudoknot conformation may be important for function [53]. A two-base mutation destabilizing the human telomerase pseudoknot is found in some patients with the inherited disease dyskeratosis congenita [54]. A theoretical calculation suggests that folding kinetics of the pseudoknot may determine activity [55]. There is much more to be revealed about the roles of pseudoknots and their modus operandi.

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
1. Brierley I, Pennell S, Gilbert Rj: Viral RNA pseudoknots: versatile motifs in gene expression and replication. Nat Rev Microbiol 2007, 5:598-610.
2. Cornish PV, Hennig M, Giedroc DP: A loop 2 cytidine-stem 1 minor groove interaction as a positive determinant for pseudoknot-stimulated -1 ribosomal frameshifting. Proc Natl Acad Sci U S A 2005, 102:12694-9.
3. Theimer CA, Blois CA, Feigon J: Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. Mol Cell 2005, 17:671-82.
4. Taufer M, Licon A, Araiza R, Mireles D, van Batenburg FH, Gultyaev AP, Leung MY: PseudoBase++: an extension of PseudoBase for easy searching, formatting and visualization of pseudoknots. Nucleic Acids Res 2009, 37:D127-35.
5. Su L, Chen L, Egli M, Berger JM, Rich A: Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. Nat Struct Biol 1999, 6:285-92.
6. Nixon PL, Rangan A, Kim YG, Rich A, Hoffman DW, Hennig M, Giedroc DP: Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot. J Mol Biol 2002, 322:621-33.
7. Pallan PS, Marshall WS, Harp J, Jewett FC 3rd, Wawrzak Z, Brown BA, 2nd, Rich A, Egli M: Crystal structure of a luteoviral...
RNA pseudoknot and model for a minimal ribosomal frameshifting motif. Biochemistry 2005, 44:11315-22.
8. Cornish PV, Stammler SN, Giedroc DP: The global structures of a wild-type and poorly functional plant luteoviral RNA pseudoknot are essentially identical. RNA 2006, 12:1959-69.
9. Egli M, Minaszov G, Su L, Rich A: Metal ions and flexibility in a viral RNA pseudoknot at atomic resolution. Proc Natl Acad Sci U S A 2002, 99:4302-7.
10. Chen G, Chang KY, Chou MY, Bustamante C, Tinoco I Jr: Triplex structures in an RNA pseudoknot enhance mechanical stability and increase efficiency of -1 ribosomal frameshifting. Proc Natl Acad Sci U S A 2009, 106:12706-11.
11. Gilbert SD, Rambo RP, Van Tyne D, Batey RT: Structure of the SAM-II riboswitch bound to S-adenosylmethionine. Nat Struct Mol Biol 2008, 15:177-82.
12. Klein DJ, Edwards TE, Ferre-D’Amare AR: Cocystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. Nat Struct Mol Biol 2009, 16:343-4.
13. Spitale RC, Torelli AT, Krucinska J, Bandaranaike WN, Wedekind JE: The structural basis for recognition of the PreQ0 metabolite by an unusually small riboswitch aptamer domain. J Biol Chem 2009, 284:10121-6.
14. Klein DJ, Ferre-D’Amare AR: Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. Science 2006, 313:1752-6.
15. Toor N, Keating KS, Taylor SD, Pyle AM: Crystal structure of a self-spliced group I intron. Science 2008, 320:77-82.
16. Adams PL, Stanley MR, Kosek AB, Wang J, Strobel SA: Crystal structure of a self-splicing group I intron with both exons. Nature 2004, 430:45-50.
17. Golden BL, Kim H, Chae E: Crystal structure of a phage Twort group I ribozyme-product complex. Nat Struct Mol Biol 2005, 12:82-9.
18. Pfingsten JS, Costantino DA, Kieft JS: Structural basis for ribosome recruitment and activation by a viral IRES RNA. Science 2006, 314:1450-4.
19. Ferre-D’Amare AR, Zhou K, Doudna JA: Crystal structure of a hepatitis delta virus ribozyme. Nature 1998, 395:567-74.
20. Guo F, Gooding AR, Cech TR: Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. Mol Cell 2004, 16:351-62.
21. Cao S, Chen SJ: Predicting RNA pseudoknot folding thermodynamics. Nucleic Acids Res 2006, 34:2634-52.
22. Cao S, Chen SJ: Predicting structures and stabilities for H-type pseudoknots with interhelic loop. RNA 2009, 15:696-706.
23. Akutsu T: Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots. Discrete Appl Math 2000, 104:45-62.
24. Lyngso RB, Pedersen CN: RNA pseudoknot prediction in energy-based models. J Comput Biol 2000, 7:409-27.
25. van Bentum FH, Gultuyse AF, Pleij CW: An APL-programmed genetic algorithm for the prediction of RNA secondary structure. J Theor Biol 1995, 174:269-80.
26. Isambert H, Sigga ED: Modeling RNA folding paths with pseudoknots: application to hepatitis delta virus ribozyme. Proc Natl Acad Sci U S A 2000, 97:6515-20.
27. Dawson WK, Fujikawa K, Kawai G: Prediction of RNA pseudoknots using heuristic modeling with mapping and sequential folding. PLoS One 2007, 2:e905.
28. Rivas E, Eddy SR: A dynamic programming algorithm for RNA structure prediction including pseudoknots. J Mol Biol 1999, 285:2053-68.
29. Reeder J, Giegerich R: Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics. BMC Bioinformatics 2004, 5:104.
30. Dirks RM, Pierce NA: A partition function algorithm for nucleic acid secondary structure including pseudoknots. J Comput Chem 2003, 24:1664-77.
31. Condon A, Davy B, Rastegari B, Zhao S, Tarrant F: Classifying RNA pseudoknotted structures. Theor Comput Sci 2004, 320:153-50.
32. Hart JM, Kennedy SD, Mathews DH, Turner DH: NMR-assisted prediction of RNA secondary structure: identification of a probable pseudoknot in the coding region of an R2 retrotransposon. J Am Chem Soc 2008, 130:10233-9.
33. Nixon PL, Giedroc DP: Energetics of a strongly pH dependent RNA tertiary structure in a frameshifting pseudoknot. J Mol Biol 2000, 296:659-71.
34. Theimer CA, Finger LD, Trantirek L, Feigon J: Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase RNA. Proc Natl Acad Sci U S A 2003, 100:449-54.
35. Soto AM, Misra V, Draper DE: Tertiary structure of an RNA pseudoknot is stabilized by “diffuse” Mg2+ ions. Biochemistry 2007, 46:2973-83.
36. Liu B, Shankar N, Turner DH: Fluorescence competition assay measurements of free energy changes for RNA pseudoknots. Biochemistry 2009. [Epub ahead of print].
37. Xia T, SantaLucia J Jr., Burkard ME, Sorensen MA: Characterization of the mechanical unfolding of RNA pseudoknots. J Mol Biol 2008, 375:511-28.
38. Nam Y, Moran SJ, Stuart DI, Gilbert RJ, Brierley I: A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. Nature 2006, 441:244-7.
42. Chen G, Wen JD, Tinoco I: Single-molecule mechanical unfolding and folding of a pseudoknot in human telomerase RNA. RNA 2007, 13:2175-88.

43. Kim NK, Zhang Q, Zhou J, Theimer CA, Peterson RD, Feigon J: Solution structure and dynamics of the wild-type pseudoknot of human telomerase RNA. J Mol Biol 2008, 384:1249-61.

44. Meyer IM, Miklos I: SimulFold: Simultaneously inferring RNA structures including pseudoknots, alignments, and trees using a Bayesian MCMC framework. PLoS Comput Biol 2007, 3:1441-54.

45. Tabaska JE, Cary RB, Gabow HN, Stormo GD: An RNA folding method capable of identifying pseudoknots and base triples. Bioinformatics 1998, 14:691-9.

46. Witwer C, Hofacker IL, Stadler PF: Prediction of consensus RNA secondary structures including pseudoknots. IEEE/ACM Trans Comput Biol Bioinform 2004, 1:66-77.

47. Ren J, Rastegari B, Condon A, Hoos HH: HotKnots: heuristic prediction of RNA secondary structures including pseudoknots. RNA 2005, 11:1494-504.

48. Ruan J, Stormo GD, Zhang W: An iterated loop matching approach to the prediction of RNA secondary structures with pseudoknots. Bioinformatics 2004, 20:58-66.

49. ten Dam E, Brierley I, Inglis S, Pleij C: Identification and analysis of the pseudoknot-containing gag-pro ribosomal frameshift signal of simian retrovirus-1. Nucleic Acids Res 1994, 22:2304-10.

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

51. Michiels P, Versleijen AA, Verlaan PW, Pleij CW, Hilbers CW, Heus HA: Solution structure of the pseudoknot of SRV-I RNA, involved in ribosomal frameshifting. J Mol Biol 2001, 310:1109-23.

52. Comolli LR, Smirnov I, Xu L, Blackburn EH, James TL: A molecular switch underlies a human telomerase disease. Proc Natl Acad Sci U S A 2002, 99:16998-7003.

53. Chen JL, Greider CW: Functional analysis of the pseudoknot structure in human telomerase RNA. Proc Natl Acad Sci U S A 2005, 102:8080-5; discussion 8077-89.

54. Chen JL, Greider CW: Telomerase RNA structure and function: implications for dyskeratosis congenita. Trends Biochem Sci 2004, 29:183-92.

55. Cao S, Chen SJ: Biphasic folding kinetics of RNA pseudoknots and telomerase RNA activity. J Mol Biol 2007, 367:909-24.

56. Kang H, Tinoco I Jr: A mutant RNA pseudoknot that promotes ribosomal frameshifting in mouse mammary tumor virus. Nucleic Acids Res 1997, 25:1943-9.

57. Kang H, Hines JV, Tinoco I Jr: Conformation of a non-frameshifting RNA pseudoknot from mouse mammary tumor virus. J Mol Biol 1996, 259:135-47.

58. Shen LX, Tinoco I Jr: The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. J Mol Biol 1995, 247:963-78.

59. Kolk MH, van der Graaf M, Fransen CT, Wijmenga SS, Pleij CW, Heus HA, Hilbers CW: Structure of the 3'-hairpin of the TYMV pseudoknot: preformation in RNA folding. EMBO J 1998, 17:7498-504.

60. Kolk MH, van der Graaf M, Wijmenga SS, Pleij CW, Heus HA, Hilbers CW: NMR structure of a classical pseudoknot: interplay of single- and double-stranded RNA. Science 1998, 280:434-8.

61. Nonin-Lecomte S, Felden B, Dardel F: NMR structure of the Aquifex aeolicus tmRNA pseudoknot PK1: new insights into the recoding event of the ribosomal trans-translation. Nucleic Acids Res 2006, 34:1847-53.