

Chapter 3

Methods for Predicting RNA Secondary Structure

Kornelia Aigner, Fabian Dreßen, and Gerhard Steger

Abstract The formation of RNA structure is a hierarchical process: the secondary structure builds up by thermodynamically favorable stacks of base pairs (helix formation) and unfavorable loops (non-Watson–Crick base pairs; hairpin, internal, and bulge loops; junctions). The tertiary structure folds on top of the thermodynamically optimal or close-to-optimal secondary structure by formation of pseudoknots, base triples, and/or stacking of helices. In this chapter, we will concentrate on available algorithms and tools for calculating RNA secondary structures as the basis for further prediction or experimental determination of higher order structures. We give an introduction to the thermodynamic RNA folding model and an overview of methods to predict thermodynamically optimal and suboptimal secondary structures (with and without pseudoknots) for a single RNA. Furthermore, we summarize methods that predict a common or consensus structure for a set of homologous RNAs; such methods take advantage of the fact that the structures of noncoding RNAs are more conserved and more critical for their biological function than their sequences.

3.1 Introduction

In this review, we will concentrate on software tools intended for prediction of secondary structure(s) of a given RNA sequence. The first such computational tool available was mfold (Zuker and Stiegler 1981); in the past 30 years, however, it was improved and refined several times (Zuker 2003). It is still commonly used, but it is now replaced by the UNAFold package (Markham and Zuker 2008), which includes several features not available in mfold. The two major alternative packages of

K. Aigner • F. Dreßen • G. Steger (✉)

Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany
e-mail: aigner@biophys.uni-duesseldorf.de; dressen@biophys.uni-duesseldorf.de;
steger@biophys.uni-duesseldorf.de

comparable or even greater scope are the Vienna RNA (Hofacker 2003) and the RNAstructure (Reuter and Mathews 2010) packages. All rely on a simplifying thermodynamic model of nearest-neighbor interaction; we will briefly summarize this model in Sect. 3.2.1. In Sect. 3.2.2, we present some of the available tools.

Because all tools use the same basic thermodynamic model and associated thermodynamic parameters, they “know” about special features of certain loops: for example, parameters of thermodynamically extra-stable hairpin loops (for a review, see Varani 1995) or small internal loops with non-Watson–Crick base pairs are taken into account (e.g., see Xia et al. 1997), but no tool mentions such details in its output. More complex arrangements, for example, stacking of helices in multi-branched loops, are not taken into account, by and large, because of the increased computational complexity and the lack of relevant parameters. Furthermore, all of the abovementioned tools disregard pseudoknots, which are important structural features in many noncoding as well as messenger RNAs. Thus, we will turn to the prediction of pseudoknotted RNA structures in Sect. 3.3.

In those cases where a set of two or more homologous RNA sequences is available, comparative sequence analysis methods can be applied to predict a consensus structure common to all sequences in the set. Such approaches, which we review in Sect. 3.4, are based on the observation that in many cases, RNA secondary and tertiary structures are more conserved than primary sequence and are of greater importance for the biological function.

We apologize to all authors whose methods and tools we have not mentioned in this review for lack of space.

3.2 RNA Secondary Structure Prediction Based on Thermodynamics

3.2.1 Overview of RNA Secondary Structure Formation

A secondary structure of an RNA sequence R consists of base stacks and loops. It is defined—at least in the context of this chapter—as

$$R = r_1, r_2, \dots, r_N,$$

with the indices $1 \leq i \leq N$ numbering the nucleotides $r_i \in \{A, U, G, C\}$ in the $5' \rightarrow 3'$ direction. Base pairs are denoted by $r_i:r_j$ or, for short, $i:j$ with $1 \leq i < j \leq N$. Allowed base pairs are *cis*-Watson–Crick (WC; A:U, U:A, G:C, C:G) and wobble pairs (G:U, U:G). Formation of base pairs belonging to a given secondary structure is restricted by

$$j \geq 4 + i, \tag{3.1}$$

which gives the minimum size of a hairpin loop, and the order of two base pairs $i:j$ and $k:l$ has to satisfy

$$i = k \quad \text{and} \quad j = l, \quad (3.2)$$

or

$$i < j < k < l, \quad (3.3)$$

or

$$i < k < l < j. \quad (3.4)$$

Condition (3.2) allows for neighboring base pairs but disallows any triple strand formation; a base triple $j:k:l$ would force $i = k$ and $j \neq l$. Condition (3.3) allows for formation of several hairpin loops in a structure. Condition (3.4) explicitly disallows “tertiary” interactions; such interactions do, in fact, occur in many RNAs, for example, in pseudoknots (see Sect. 3.3).

Structure formation—from an unfolded, random coil structure, C, into the folded structure, S—is a standard equilibrium reaction with a temperature-dependent equilibrium constant, K :

$$\begin{aligned} C &\rightleftharpoons S, \\ K &= \frac{[S]}{[C]}, \\ \Delta G_T^0 &= -RT \ln K = \Delta H^0 - T \cdot \Delta S^0. \end{aligned}$$

At the denaturation temperature $T_m = \Delta H^0 / \Delta S^0$ (melting temperature or mid-point of transition), the folded structure S has the same concentration as the unfolded structure ($K = 1$; $\Delta G_{T_m}^0 = 0$). This is only true if the structure S denatures in an all-or-none transition. In most cases, however, structural rearrangements and/or partial denaturation take place prior to complete denaturation, as temperature is increased.

The number of possible secondary structures of a single sequence grows exponentially ($\approx 1.8^N$) with the sequence length N (Waterman 1995). Accordingly, all possible structures S_i of a single sequence coexist in solution with concentrations dependent on their free energies $\Delta G^0(S_i)$; that is, each structure is present as a fraction given by (3.5):

$$f_{S_i} = \exp\left(\frac{-\Delta G_T^0(S_i)}{RT}\right) / Q. \quad (3.5)$$

The partition function, Q , for the ensemble of all possible structures, is given by (3.6):

$$Q = \sum_{\text{all structures } S_i} \exp\left(\frac{-\Delta G_T^0(S_i)}{RT}\right). \quad (3.6)$$

The structure of lowest free energy is called the optimal structure or structure of minimum free energy (mfe). It is possible for a single sequence to fold into quite different structures with nearly identical energies. This is of special biological relevance for RNA switches (Garst and Batey 2009; Nagel and Pleij 2002). Thus, one should not assume that an RNA folds into a single, static structure.

The free energy ΔG_T^0 of a single structure S is calculated as a sum over the free energy contributions (enthalpy ΔH^0 and entropy ΔS^0) of all structural elements i and j of S at temperature T :

$$\Delta G_T^0 = \sum_i (\Delta H_{\text{stack}}^0 - T\Delta S_{\text{stack}}^0) + \sum_j (\Delta H_{\text{loop}}^0 - T\Delta S_{\text{loop}}^0). \quad (3.7)$$

In this calculation, a nearest-neighbor model with base-pair stacking and loop formation is assumed to be sufficient. Energetic contributions of adjacent base pairs are favorable ($\Delta G^0 < 0$) due to their stacking on top of each other to form regular helices. Formation of loops is often but not necessarily unfavorable ($\Delta G^0 > 0$); exact values depend on loop type, nucleotides neighboring the loop-closing base pair(s), as well as on the exact sequence of the loop and whether the loop nucleotides form a stable, structured motif. Loop types are classified according to the number of loop-closing base pair(s): a single base pair closes hairpin loops, two base pairs close bulge loops (with no nucleotides in one strand) and interior loops (with symmetric or asymmetric numbers of nominally unpaired nucleotides in both strands), and three or more base pairs close multiloops (also called bifurcations or junctions). Note that for a given interior loop of n nucleotides, there are up to $6 \times 6 \times n^4$ different sequence combinations with possibly different energetic contributions, when taking into account the six possibilities for each of the closing base pairs. The parameter set measured by the group of D. Turner is used almost universally (Mathews et al. 1999, 2004; Xia et al. 1998). Parameters are known only within certain error limits; because these errors are smallest near $T = 37^\circ\text{C}$, mostly $\Delta G_{37^\circ\text{C}}^0$ values are reported.

A loop should not be thought of as a floppy structural element: in many cases, loop nucleotides form distinct structures due to stacking and/or non-Watson–Crick (non-WC; Leontis et al. 2002; Stombaugh et al. 2009) interactions with other loop nucleotides. Famous examples are loop E of eukaryotic 5S rRNA and the multiloop of tRNA. Eukaryal loop E, which is the same as the sarcin/ricin loop, is an asymmetric internal loop of four and five bases in its parts; all nucleotides are involved in non-WC interactions including one triple-base interaction (Wimberly et al. 1993). In tRNA, stacking of multiloop-closing base pairs across the multiloop is a major energetic contribution to the stability of the cloverleaf and is critical for formation of the tRNA tertiary structure.

Compensation of the negatively charged phosphate backbone of nucleic acids by positively charged counter ions M^+ leads to stabilization of structural elements according to



From this expression, a logarithmic dependence between denaturation temperature T_m and salt concentration (ionic strength) follows

$$\frac{dT_m}{d\ln[M]} = -n \frac{RT_m^2}{\Delta H^0}. \quad (3.9)$$

All thermodynamic parameters for RNA structure formation were determined in 1 M NaCl. This is not far from the ionic conditions in cells, except when specific interactions with divalent cations play a role (Draper 2008; Ramesh and Winkler 2010). If necessary, however, values for the ionic strength dependence of a structure or a structural element may be found in the literature, including functions for G:C-content of the RNA, or for dependence upon various types of buffers (e.g., TRIS/borate) and cosolvents like formamide or urea (Klump 1977; Michov 1986; McConaughy et al. 1969; Record and Lohman 1978; Riesner and Steger 1990; Sadhu and Gedamu 1987; Shelton et al. 1999; Steger et al. 1980).

3.2.2 Tools for RNA Secondary Structure Prediction Based on Thermodynamics

Most users seek to predict the mfe structure for a given (single) sequence. For the answer, the most widely used tools (see Table 3.1) rely on (3.7); that is, their basic algorithm solves the optimization problem of finding the mfe structure of a given single sequence in the haystack of thermodynamically possible secondary structures via dynamic programming (Bellman and Kalaba 1960; Nussinov et al. 1978; Zuker and Stiegler 1981). The computational effort grows with the cube of the sequence length N , that is, $O(N^3)$.

All tools except UNAFold rely on the same recent set of thermodynamic parameters, which only allows for calculation of $\Delta G_{37^\circ\text{C}}^0$; UNAFold uses a parameter set of enthalpy and entropy values that makes possible calculations at any relevant temperature.

Knowledge of the mfe structure might not be sufficient due to several reasons:

- Minor errors in the thermodynamic parameter set can lead to incorrect prediction of the mfe structure; nonetheless, the “true” mfe structure may be one of the suboptimal structures close in energy to the calculated mfe structure. While a further improvement of the accuracies of experimentally determined parameters is unlikely, improvements in secondary structure prediction by statistical evaluations of known structures look promising (Andronescu et al. 2007; Wu et al. 2009).
- Quite often, the mfe structure only accounts for a very tiny fraction of all possible structures; that is, a cluster of suboptimal structures, which are nearly identical to each other but different from the mfe structure, might account for a much higher fraction of all possible structures and thus be of higher (biological) relevance.
- Usually, it is assumed that structure formation is a hierarchical process: the tertiary structure builds on top of the most favorable secondary structure(s) (Brion and

Table 3.1 Tools for prediction of RNA secondary structure

Package name	Address ^a	Reference
UNAFold	C: dinamelt.bioinfo.rpi.edu/download.php ^{b,c,d}	Markham and Zuker (2008)
	W: dinamelt.bioinfo.rpi.edu/	Markham and Zuker (2005)
Mfold ^e	C: mfold.bioinfo.rpi.edu/download/ ^{b,c}	Zuker (1989)
	W: mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi	Zuker (2003)
Vienna RNA	C: www.tbi.univie.ac.at/~ivo/RNA/ ^{b,d}	Hofacker et al. (1994)
	W: rna.tbi.univie.ac.at/	Hofacker (2003)
RNAstructure	C: rna.urmc.rochester.edu/RNAstructure.html ^{b,c,d}	Reuter and Mathews (2010)
RNASHAPES	C: bibiserv.techfak.uni-bielefeld.de/rnashapes/ ^{b,c,d}	Steffen et al. (2006)
	W: bibiserv.techfak.uni-bielefeld.de/rnashapes/	Giegerich et al. (2004)
CentroidFold	C: http://www.ncrna.org/centroidfold/	Hamada et al. (2009)
	W: http://www.ncrna.org/centroidfold/	
Sfold	W: http://sfold.wadsworth.org/srna.pl	Chan et al. (2005)

The Vienna RNA, RNAstructure, and UNAFold packages include, for example, programs for prediction of the mfe structure and for partition function folding of a single sequence and for bimolecular structure prediction

^aC, source code and binary are available at given address; W, address of Web service

^bUnix, Linux

^cMacOSX

^dWindows

^eMfold is replaced by UNAFold

Westhof 1997; Cho et al. 2009; Tinoco and Bustamante 1999). But as long as the free energy of the RNA decreases upon tertiary structure formation, this structure may also be able to fold starting from suboptimal secondary structures.

- Site-specific binding of multivalent ions, small molecules, or macromolecules, including proteins or RNAs, might influence the process of structure formation. Note, however, that the thermodynamic stability of even short RNA helices is larger than that of most proteins.

Consequently, the programs of Table 3.1, also predict individual, suboptimal secondary structures (like mfold, which is able to generate the thermodynamically best structure for each admissible base pair) or, alternatively, predict the probability of any base pair possible for a certain sequence by using (3.5) and (3.6) (like RNAfold); the prediction is usually represented as a dot plot (see Fig. 3.1b). This base-pairing probability matrix is easily converted to a plot showing the probability of each nucleotide to be paired or unpaired; this allows, for example, for comparison to chemical or enzymatic mapping data. In case certain bases are experimentally known to pair or to remain unpaired, mfold as well as RNAfold allow imposition of the corresponding constraints, so that the predicted structures and the dot plot satisfy the known constraints (see help and man pages of mfold and RNAfold, respectively, and Steger 2004).

Enumeration of all secondary structures is possible (Waterman and Byers 1985; Wuchty et al. 1999), but one must be aware of the huge number of structures that result, many of which are very similar to each other. In many cases, the algorithm implemented in RNASHAPES (see Table 3.1) is more useful: it classifies all

3.3 Pseudoknots

A pseudoknot is an RNA structure characterized by WC base pairing between nucleotides in a loop with complementary residues outside the loop. In contrast to proteins (Taylor 2007), no knots are known in RNA. Pseudoknots are a tertiary structural motif that occurs widely in RNA. They were first detected nearly 30 years ago as part of tRNA-like structures in plant viral RNAs (Rietveld et al. 1982). Some pseudoknots play a role in ribosomal frameshifting, while others are essential for the three-dimensional topology (and function) of many structured RNAs. In the following, we will give a description of pseudoknots and sequence constraints on their biophysical stability.

Databases on structural, functional, and sequence data related to RNA pseudoknots are maintained by PseudoBase (<http://www.ekevanbatenburg.nl/PKBASE/PKABOUT.HTML>; van Batenburg et al. 2001) and PseudoBase++ (<http://pseudobaseplus.utep.edu/>; Taufer et al. 2008).

3.3.1 Conformation

A classical or H-type (hairpin-type) pseudoknot consists of two helical regions named S1 and S2 (or H1 and H2) and three loop regions L1, L2, and L3 (see Fig. 3.2). In sequence, the serial arrangement of these elements is S1, L1, S2, L2, S1' (complement of S1), L3, and S2' (complement of S2). The crossing order $S1 < S2 < S1' < S2'$ fulfills the definition of base pairs in a tertiary structure [see Sect. 3.2.1, (3.4)]. In many cases, the loop region L2 is absent and the two helices coaxially stack as shown in Fig. 3.2.

In the classical pseudoknot, the loops L1–L3 contain only unpaired nucleotides. There are, however, more complicated pseudoknots, in which these regions contain structured parts, including non-WC pairs and base triples; an example of a double pseudoknot with a stem-loop structure in L3 is shown in Fig. 3.3.

In the absence of L2, the helices S1 and S2 generally stack coaxially forming a structure closely resembling an uninterrupted A-form helix. Consequently, the loops L1 and L3 are not equivalent: L1 crosses the deep (major) groove and L3 the shallow (minor) groove of the double helix (see Fig. 3.2 right and Fig. 3.4 left). In a WC base pair, the distance between the phosphates (P'_0 and P_0 in Fig. 3.4) is about 1.7 nm; this distance can be bridged, for example, by a minimum of three nucleotides in a hairpin loop. In an RNA double helix, the minimal distance between a given phosphate on one strand and a phosphate on the opposite strand is about 1 nm when crossing the deep groove (e.g., P'_0 to P_{-7} ; see Fig. 3.4). The smallest distance bridging the shallow groove is about 1.1 nm, the distance between P'_0 and P_2 or P_3 . These distances fit well to the sizes of small pseudoknots with coaxial helix stacking: 3–7 base pairs in stem regions bridged by loops of at least 2 nucleotides. A loop L2 and smaller L1 and/or L3 tend to introduce a bend in between the two stems. The shallow and wide minor groove

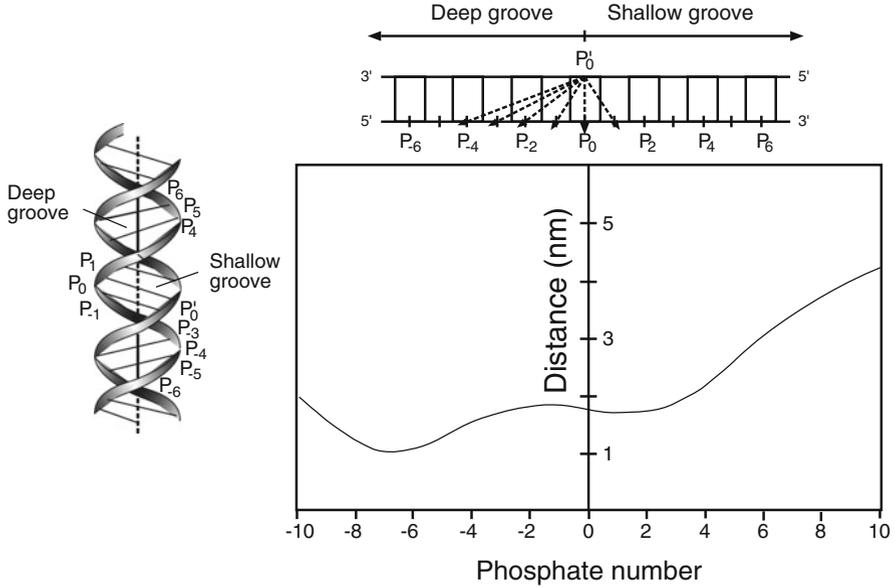


Fig. 3.4 Distance constraints on pseudoknots. Shown are the minimum distances between a certain phosphate P_0' to phosphates P located opposite on the other strand. Indices are negative for phosphates located in 5' direction of the opposing strand and positive in 3' direction. *Left*: Three-dimensional model of a helix. *Top*: Two-dimensional model of a helix; the *arrows* symbolize the distance from P_0' to the corresponding phosphate. *Bottom*: A graph with distances from P_0' to phosphates in the opposing strand. According to Pleij et al. (1985)

the difficulties inherent in evaluating parameters from overlapping and/or coupled unfolding transitions by optical melting or calorimetry (Gluick and Draper 1994; Gultyaev et al. 1999; Nixon and Giedroc 1998, 2000; Qiu et al. 1996; Soto et al. 2007; Theimer and Giedroc 1999, 2000; Theimer et al. 1998; Wyatt et al. 1990).

The total free energy of a pseudoknot is assumed to be the sum over free energies of stems, coaxial stacking, loop lengths and sequences, tertiary interactions, and assembling (Liu et al. 2010):

$$\Delta G_{\text{pseudoknot}}^0 = \sum \Delta G_{\text{stems}}^0 + \Delta G_{\text{coaxial stacking}}^0 - T \sum \Delta S_{\text{loops}}^0 + \Delta G_{\text{loop sequences}}^0 + \Delta G_{\text{tertiary interactions}}^0 + \Delta G_{\text{assemble}}^0.$$

$\Delta G_{\text{stems}}^0$ and $\Delta G_{\text{coaxial stacking}}^0$ can be calculated with experimentally determined nearest-neighbor parameters (Xia et al. 1998) and coaxial stacking parameters (Walter et al. 1994), respectively. Several computational models have been established for the remaining parameters (Cao and Chen 2006, 2009; Dirks and Pierce 2003, 2004; Gultyaev et al. 1999; Rivas and Eddy 1999). The models particularly determine the loop length dependence of $\Delta S_{\text{loops}}^0$. Taking into account volume exclusion effects of the loop strands and considering the loop length with

respect to the length of the associated stem resulted in improvement of the models (statistical polymer model; Cao and Chen 2006, 2009). Details of $\Delta G_{\text{loop sequences}}^0$ are currently neither determined experimentally nor does a computational model exist. This energy does, however, contribute to the total energy, even if the loop sequence is not involved in tertiary interactions, as demonstrated experimentally by Liu et al. (2010). $\Delta G_{\text{tertiary interactions}}^0$ accounts for possible interactions between loops and stems; it is currently neither determined experimentally nor does a computational model exist. In some cases, tertiary interactions are more favorable than the maximum number of canonical base pairs (Liu et al. 2010). $\Delta G_{\text{assemble}}^0$ is assessed to account for the entropy change as the two subunits (the two stems with their associated loops) are assembled into the pseudoknot (Cao and Chen 2006).

3.3.3 Ionic Strength Dependence of Pseudoknots

For their formation, most pseudoknots need a relatively high ionic strength including the presence of divalent cations like Mg^{2+} (Gluick et al. 1997). Considering the structure of a simple pseudoknot, as depicted in Fig. 3.2 right, the reason for this is quite obvious: the stabilizing interactions realized upon formation of stems 1 and 2 and stacking of stem 1 on stem 2 are partly counteracted by the necessary loop formation and the close approach of four negatively charged phosphate backbones. To compensate for this increased charge density, “diffuse” (fully hydrated) Mg^{2+} ions seem generally to be sufficient; binding of dehydrated ions (inner-sphere complexes) to specific positions is not necessary (Soto et al. 2007).

3.3.4 Prediction Methods for Pseudoknots

None of the tools mentioned in Sect. 3.2.2 are capable of predicting pseudoknots or any form of tertiary interactions due to the restrictions in their dynamic programming algorithms. Expanding these algorithms to general pseudoknot prediction is difficult; actually, Lyngsø and Pedersen (2000) have proven that the general problem of predicting RNA secondary structures containing pseudoknots is NP complete for a large class of reasonable models of pseudoknots. Thus, several heuristic approaches were developed. In the following, we will mention several of the recent tools able to detect pseudoknots and other tertiary interactions in structure predictions; an incomplete list of available tools is summarized in Table 3.2.

PKNOTS, developed by Rivas and Eddy (1999), is the first dynamic programming algorithm that finds optimal, pseudoknotted RNA structures. Due to its computational effort of $O(N^6)$, its use is restricted to short sequences.

pknotsRG (Reeder and Giegerich 2004; Reeder et al. 2007) predicts the structure of an RNA sequence, possibly containing pseudoknots. Energies of possible

Table 3.2 Tools for prediction of tertiary interactions

Name	Address ^a	Effort ^b	Reference
ConStruct	C: http://www.biophys.uni-duesseldorf.de/construct3/		Wilm et al. (2008b)
DotKnot	W: http://dotknot.csse.uwa.edu.au		Sperschneider and Datta (2010)
HotKnots	C: http://www.cs.ubc.ca/labs/beta/Software/HotKnots W: http://www.masoft.ca/cgi-bin/RNAsoft/HotKnots/hotknots.pl		Ren et al. (2005)
HPknotter	W: http://bioalgorithm.life.nctu.edu.tw/HPKNOTTER/		Huang et al. (2005)
ILM	C: http://www.cse.wustl.edu/~zhang/projects/rna/ilm/ W: http://cic.cs.wustl.edu/RNA/	$O(N^3) - O(N^4)$	Ruan et al. (2004)
KNetFold	W: http://knetfold.abcc.ncifcrf.gov/	$O(n^2)^c$	Bindewald and Shapiro (2006)
KnotSeeker	C: http://knotseeker.csse.uwa.edu.au/download.html W: http://knotseeker.csse.uwa.edu.au/		Sperschneider and Datta (2008)
NUPACK	W: http://nupack.org/ C: http://nupack.org/downloads	$O(n^5)$	Dirks and Pierce (2003)
PKNOTS	C: ftp://selab.janelia.org/pub/software/pknots/	$\geq O(N^6)$	Rivas and Eddy (1999)
pknotsRG	C: http://bibiserv.techfak.uni-bielefeld.de/download/tools/pknotsrg.html W: http://bibiserv.techfak.uni-bielefeld.de/pknotsrg	$O(N^4)$	Reeder et al. (2007) Reeder and Giegerich (2004)
PSTAG	C: http://phmmts.dna.bio.keio.ac.jp/pstag/download.html W: http://phmmts.dna.bio.keio.ac.jp/pstag/	$O(on^4 + mn^5)^d$	Matsui et al. (2005)
vsfold5	W: http://www.ma.it-chiba.ac.jp/~vsfold/vsfold5	$O(N^{4.7})$	Dawson et al. (2007)

^aW, address of Web service; C, code is available at given address

^bComputing effort; N , length of sequence

^c n , length of alignment

^d n , length of unfolded pair of sequences; m , o , nodes on structure tree

structures are calculated as sum of the energies of the two pseudoknot helices and some not described loop folding energies. Suboptimal foldings up to a user-defined energy threshold can be enumerated, and for large scale analysis, a fast sliding window mode is available.

ILM (Ruan et al. 2004) combines dynamic programming (mainly maximizing number of base pairs) and comparative information to find iteratively high-scoring helices, adds them to the structure, and removes the corresponding sequence segments from the sequence. Due to the removal step, there is no restriction on the type of pseudoknot. The thermodynamic approach uses energy parameters for helix stacking from the Vienna package.

HotKnots (Ren et al. 2005) expands the idea of ILM by considering several alternative secondary structures and returning a fixed number of suboptimal folding scenarios. The program uses Turner parameters (Mathews et al. 1999; Serra and Turner 1995) together with those of Dirks and Pierce (2004) and Cao and Chen (2006) for pseudoknotted loops to determine the energy of a structure. According to its authors, HotKnots outperforms STAR (Gulyaev 1991), PKNOTS, pknotsRG, and ILM.

KnotSeeker was described by Sperschneider and Datta (2008) as capable of detecting pseudoknots in long RNA sequences. The algorithm combines the output of several known programs for prediction in a serial fashion. According to the authors, KnotSeeker has higher sensitivity and specificity in detection of pseudoknots than pknotsRG, ILM, and HPknotter (Huang et al. 2005).

DotKnot (Sperschneider and Datta 2010) predicts a wide class of pseudoknots including bulged stems (not accessible for pknotsRG) by consulting probability dot plots, from which probable stems are inferred. These are assembled to compose pseudoknot candidates by employing bulge-loop and multiloop dictionaries. After free energy calculations with one of three different energy models, chosen according to the length of the interhelix loop, “reliable” pseudoknots are retained. This approach also manages long sequences with complex pseudoknotted structures.

The authors of each of the abovementioned programs tested their programs with restricted datasets, and the programs are not benchmarked by an independent group. However, new experimental results on free energies for specific pseudoknots from Liu et al. (2010) show that (1) it is not sufficient to calculate pseudoknot energies just by summing nearest-neighbor interactions within the component helices; (2) conformational entropy parameters for loops give the best approximation to loop entropies; (3) the lack of parameters for tertiary interactions is best compensated for by building as many *cis* WC base pairs as possible, although crystal structures show that these are sometimes replaced by favorable tertiary interactions.

3.4 Prediction of Consensus Structures

The accuracy of (mfe) secondary structure prediction for a single RNA sequence is relatively low. This is due to several factors including simplifications in the underlying model, uncertainties of the energy parameters (especially with stacking in larger loops and junctions), ignorance of kinetic factors (which are of increasing importance with increasing sequence length), and disregard of energy contributions of tertiary interactions. Values of accuracy for predicting correct base pairs range from as low as $(45 \pm 16)\%$ up to $(83 \pm 22)\%$ mostly depending on the tested sequence families (see Doshi et al. 2004; Mathews et al. 1999; Wilm et al. 2006, 2008b). A formidable improvement in prediction accuracy can be achieved, however, by using the additional information from sufficiently

diverged homologous sequences. This approach is based on the fact that the secondary and tertiary structure of a noncoding RNA changes more slowly than the sequence during evolution. Mutations in base-paired regions are mainly compensated by further mutations that retain the pairing scheme. Due to the isostericity of all WC pairs (and other groups of non-WC pairs; see Leontis et al. 2002; Stombaugh et al. 2009), the structure common to homologous RNAs can easily be conserved while their sequences might differ from each other to a large extent.

The common structure for a set of homologous sequences is called the consensus structure. To find it, one would like to perform simultaneously a sequence and structure alignment, which has a prohibitive computational cost of $O(N^{3m})$ for m sequences of length N (Sankoff 1985). Hence, several simplifying and more pragmatic approaches for consensus structure prediction have been developed (see Table 3.3) that can be classified as follows (Gardner and Giegerich 2004):

1. Align the sequences first and then predict the structure common to the aligned sequences (Bernhart et al. 2008; Bindewald and Shapiro 2006; Wilm et al. 2008b). For the primary alignment step, pure sequence alignment programs or one of the sequence + structure alignment programs (see below) can be used. Dynamic programming (Bernhart et al. 2008; Wilm et al. 2008b) (for secondary structure prediction) or “maximum weighted matching” (for secondary structure prediction including pseudoknots or base triples; Tabaska et al. 1998; Wilm et al. 2008b) might be used in the structure prediction step given the fixed alignment. Several RNA sequence + structure editors are available (e.g., Griffiths-Jones 2004; Jossinet and Westhof 2005; Seibel et al. 2006; Wilm et al. 2008b) that allow a user to refine the initial alignment.
2. Predict structures for all single sequences and then align these structures (Dalli et al. 2006; Höchsmann et al. 2004; Moretti et al. 2008; Xu et al. 2007).
3. Align and predict structures at the same time, using heuristics and/or restrict the alignment to two sequences to lower the computing cost of Sankoff’s algorithm (Bauer et al. 2007; Harmanci et al. 2007, 2008; Hofacker et al. 2004; Holmes 2005; Katoh and Toh 2008; Kiryu et al. 2007; Lindgreen et al. 2007; Perriquet et al. 2003; Torarinsson et al. 2007; Will et al. 2007; Yao et al. 2005).

This separation of approaches should not be taken too strictly; for example, several of the Sankoff-like approaches first restrict the sequence + structure search space by taking into account a sequence alignment and partition functions for the individual sequences. Other approaches do also exist: for example, RNACast predicts an abstract shape common to all sequences (Reeder and Giegerich 2005), where each shape of an RNA molecule comprises a class of similar structures and has a representative structure of minimal free energy within the class. That is, RNACast predicts a consensus structure but does not align the sequences.

In general, all methods for consensus structure prediction outperform the single-sequence methods, but several prerequisites have to be met:

Table 3.3 Tools for prediction of RNA consensus secondary structure

Name	Address ^a	Reference
CARNAC	C: http://bioinfo.lifl.fr/RNA/carnac/index.php	Perriquet et al. (2003)
	W: http://bioinfo.lifl.fr/RNA/carnac/carnac.php	Touzet and Perriquet (2004)
CMfinder	C: http://bio.cs.washington.edu/yzizhen/CMfinder/	Yao et al. (2005)
	W: http://wingless.cs.washington.edu/htbin-post/unrestricted/CMfinderWeb/CMfinderInput.pl	
Consan	C: http://selab.janelia.org/software.html	Eddy and Dowell (2006)
ConStruct	C: http://www.biophys.uni-duesseldorf.de/construct3/	Wilm et al. (2008b)
Dynalign	C: http://rna.urmc.rochester.edu/dynalign.html	Harmanci et al. (2007)
foldalignM	C: http://foldalign.ku.dk/software/index.html	Torarinsson et al. (2007)
KNetFold	C: http://www-lmmb.ncifcrf.gov/~bshapiro/downloader_v1/register.php	Bindewald and Shapiro (2006)
	W: http://knetfold.abcc.ncifcrf.gov/	
LARA	C: https://www.mi.fu-berlin.de/w/LiSA/	Bauer et al. (2007)
LocARNA	C: http://www.bioinf.uni-freiburg.de/Software/LocARNA/	Will et al. (2007)
MAFFT	W: http://align.bmr.kyushu-u.ac.jp/mafft/online/server/	Katoh and Toh (2008)
MASTR	C: http://mastr.binf.ku.dk/	Lindgreen et al. (2007)
Murlet	W: http://murlet.ncrna.org/murlet/murlet.html	Kiryu et al. (2007)
MXSCARNA	C: http://www.ncrna.org/software/mxscarna/download/	Tabei et al. (2008)
	W: http://mxscarna.ncrna.org/mxscarna/mxscarna.html	
PARTS	C: http://rna.urmc.rochester.edu/	Harmanci et al. (2008)
Pfold	W: http://www.daimi.au.dk/~compbio/mfold/	Knudsen and Hein (2003)
PMcomp/ PMmulti	C: http://www.tbi.univie.ac.at/~ivo/RNA/PMcomp/	Hofacker et al. (2004)
	W: http://rna.tbi.univie.ac.at/cgi-bin/pmcgi.pl	
R-Coffee	C: http://www.tcoffee.org/Projects_home_page/r_coffee_home_page.html	Wilm et al. (2008a)
	W: http://www.tcoffee.org/	
RNAalifold	C: http://www.tbi.univie.ac.at/~ivo/RNA/	Hofacker et al. (2002)
	W: http://rna.tbi.univie.ac.at/cgi-bin/RNAalifold.cgi	
RNAcast	C: http://bibiserv.techfak.uni-bielefeld.de/rnecast/	Reeder and Giegerich (2005)
	W: http://bibiserv.techfak.uni-bielefeld.de/mashapes/submission.html	
RNAforester	C: http://bibiserv.techfak.uni-bielefeld.de/rnaforester/	Höchsmann et al. (2004)
	W: http://bibiserv.techfak.uni-bielefeld.de/rnaforester/submission.html	
RNAmine	W: http://rnamine.ncrna.org/rnamine/	Hamada et al. (2006)
RNASampler	C: http://ural.wustl.edu/~xingxu/RNASampler/index.html	Xu et al. (2007)
SCARNA	W: http://www.scarna.org/scarna/	Tabei et al. (2006)
SimulFold	C: http://www.cs.ubc.ca/~irmtraud/simulfold/	Meyer and Miklós (2007)
StemLoc	C: http://biowiki.org/StemLoc	Holmes (2005)

(continued)

Table 3.3 (continued)

Name	Address ^a	Reference
StrAl	C: http://www.biophys.uni-duesseldorf.de/stral/about.php W: http://www.biophys.uni-duesseldorf.de/stral/advancedForm.php	Dalli et al. (2006)
WAR	W: http://genome.ku.dk/resources/war/	Torarinsson and Lindgreen (2008)

^aC, source code is available at given address; W, address of Web service

- The performance of most (iterative) programs improves with an increasing number of input sequences and decreasing identities of sequences. Optimal values might be five sequences with an average pairwise sequence identity (APSI) of 55–70%.
- Only the structure alignment programs (approach 3 or RNACast) might give reasonable results for a sequence set with an APSI below 55%, but most of these programs are very demanding in computer resources.
- While even a single compensating base-pair change might hint to a certain structure, a pure statistical analysis [e.g., via information theory (Chiu and Kolodziejczak 1991; Wilm et al. 2008b); for other methods see Gruber et al. (2008)] needs more than ten sequences and still does not reach the accuracy of thermodynamic-based approaches.

3.5 Conclusions

In concluding this review, we propose the following approach for constructing an RNA alignment for consensus structure prediction:

1. We assume at least one and probably no more than a few closely related sequences are known.
2. First, use pure sequence search methods (like BLAST) to find more homologues of the sequence(s) from *step 1*. Due to the use of pure sequence search, the found homologues will be closely related to the already known sequences. For an overview and benchmark of selected RNA search tools, see Freyhult et al. (2007).
3. Next, create an alignment of the sequences and a consensus structure using an alignment program appropriate for the lengths and number of sequences; for example, MAFFT (in mode Q-INS-i; Katoh and Toh 2008) or StrAl (Dalli et al. 2006) accepts more and longer sequences than StemLoc (Holmes 2005) or LocARNA (Will et al. 2007). This preliminary consensus structure should be checked for consistency (and refined accordingly) by means of ConStruct (Wilm et al. 2008b) or RNAalifold (Bernhart et al. 2008).
4. Use the preliminary consensus structure to create either a pattern (see, e.g., Dsouza et al. 1997; Gautheret and Lambert 2001; Gräf et al. 2006; Macke et al. 2001; Mosig et al. 2009) or a covariance model (see Klein and Eddy 2003;

Nawrocki and Eddy 2007). Use either model to search more specifically for further members of the RNA group under inspection. Alternatively, reiterate from step 2.

5. Check the refined model for consistency with ConStruct or RNAalifold using thermodynamics and covariation analysis. If this gives new information—especially in terms of tertiary interactions and/or base triples—reiterate from step 4, otherwise, this final model could be refined further by verification from wet lab experiments.

If additional experimental data is available, for example, from chemical or enzymatic mapping (Ehresmann et al. 1987; Tullius and Greenbaum 2005), the initial structure prediction by RNAfold or mfold can accordingly be constrained and thus incorporated into the model (Deigan et al. 2009). If in addition information on the three-dimensional structure of one of the sequences from the set is available from X-ray or NMR analysis, the use of an editor like S2S (Jossinet and Westhof 2005) is advantageous.

Acknowledgments We thank Jana Sperschneider for useful discussions.

References

- Andronescu M, Condon A, Hoos H, Mathews D, Murphy K (2007) Efficient parameter estimation for RNA secondary structure prediction. *Bioinformatics* 23:i19–i28, <http://dx.doi.org/10.1093/bioinformatics/btm223>
- Batey R, Rambo R, Doudna J (1999) Tertiary motifs in RNA structure and folding. *Angew Chem Int Ed Engl* 38:2326–2343, [http://dx.doi.org/10.1002/\(SICI\)1521-3773\(19990816\)38:16<2326::AID-ANIE2326>3.0.CO;2-3](http://dx.doi.org/10.1002/(SICI)1521-3773(19990816)38:16<2326::AID-ANIE2326>3.0.CO;2-3)
- Bauer M, Klau G, Reinert K (2007) Accurate multiple sequence-structure alignment of RNA sequences using combinatorial optimization. *BMC Bioinformatics* 8:271, <http://dx.doi.org/10.1186/1471-2105-8-271>
- Bellman R, Kalaba R (1960) On k th best policies. *SIAM J Appl Math* 8:582–588, <http://dx.doi.org/10.1137/0108044>
- Bernhart S, Hofacker I, Will S, Gruber A, Stadler P (2008) RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* 9:474, <http://dx.doi.org/10.1186/1471-2105-9-474>
- Bindewald E, Shapiro B (2006) RNA secondary structure prediction from sequence alignments using a network of k -nearest neighbor classifiers. *RNA* 12:342–352, <http://dx.doi.org/10.1261/rna.2164906>
- Brion P, Westhof E (1997) Hierarchy and dynamics of RNA folding. *Annu Rev Biophys Biomol Struct* 26:113–137, <http://dx.doi.org/10.1146/annurev.biophys.26.1.113>
- Cao S, Chen S (2006) Predicting RNA pseudoknot folding thermodynamics. *Nucleic Acids Res* 34:2634–2652, <http://dx.doi.org/10.1093/nar/gkl346>
- Cao S, Chen S (2009) Predicting structures and stabilities for H-type pseudoknots with interhelix loops. *RNA* 15:696–706, <http://dx.doi.org/10.1261/rna.1429009>
- Chan C, Lawrence C, Ding Y (2005) Structure clustering features on the Sfold Web server. *Bioinformatics* 21:3926–3928, <http://dx.doi.org/10.1093/bioinformatics/bti632>
- Chiu D, Kolodziejczak T (1991) Inferring consensus structure from nucleic acid sequences. *Comput Appl Biosci* 7:347–352, <http://dx.doi.org/doi:10.1093/bioinformatics/7.3.347>

- Cho S, Pincus D, Thirumalai D (2009) Assembly mechanisms of RNA pseudoknots are determined by the stabilities of constituent secondary structures. *Proc Natl Acad Sci USA* 106:17349–17354, <http://dx.doi.org/10.1073/pnas.0906625106>
- Dalli D, Wilm A, Mainz I, Steger G (2006) STRAL: Progressive alignment of non-coding RNA using base pairing probability vectors in quadratic time. *Bioinformatics* 22:1593–1599, <http://bioinformatics.oxfordjournals.org/cgi/reprint/22/13/1593>, <http://dx.doi.org/10.1093/bioinformatics/btl142>
- Dawson W, Fujiwara K, Kawai G (2007) Prediction of RNA pseudoknots using heuristic modeling with mapping and sequential folding. *PLoS ONE* 2:e905, <http://dx.doi.org/10.1371/journal.pone.0000905>
- Deigan K, Li T, Mathews D, Weeks K (2009) Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci USA* 106:97–102, <http://dx.doi.org/10.1073/pnas.0806929106>
- Dirks R, Pierce N (2003) A partition function algorithm for nucleic acid secondary structure including pseudoknots. *J Comput Chem* 24:1664–1677, <http://dx.doi.org/10.1002/jcc.10296>
- Dirks R, Pierce N (2004) An algorithm for computing nucleic acid base-pairing probabilities including pseudoknots. *J Comput Chem* 25:1295–1304, <http://dx.doi.org/10.1002/jcc.20057>
- Doshi K, Cannone J, Cobaugh C, Gutell R (2004) Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinformatics* 5:105, <http://dx.doi.org/10.1186/1471-2105-5-105>
- Draper D (2008) RNA folding: thermodynamic and molecular descriptions of the roles of ions. *Biophys J* 95:5489–5495, <http://dx.doi.org/10.1529/biophysj.108.131813>
- Dsouza M, Larsen N, Overbeek R (1997) Searching for patterns in genomic data. *Trends Genet* 13:497–498, [http://dx.doi.org/10.1016/S0168-9525\(97\)01347-4](http://dx.doi.org/10.1016/S0168-9525(97)01347-4)
- Eddy S, Dowell R (2006) Efficient pairwise RNA structure prediction and alignment using sequence alignment constraints. *BMC Bioinformatics* 7:400, <http://www.biomedcentral.com/1471-2105/7/400>
- Ehresmann C, Baudin F, Mougél M, Romby P, Ebel JP, Ehresmann B (1987) Probing the structure of RNAs in solution. *Nucleic Acids Res* 15:9109–9128, <http://dx.doi.org/10.1093/nar/15.22.9109>
- Freyhult E, Bollback J, Gardner P (2007) Exploring genomic dark matter: a critical assessment of the performance of homology search methods on noncoding RNA. *Genome Res* 17:117–125, <http://dx.doi.org/10.1101/gr.5890907>
- Gardner P, Giegerich R (2004) A comprehensive comparison of comparative RNA structure prediction approaches. *BMC Bioinformatics* 5:140, <http://dx.doi.org/10.1186/1471-2105-5-140>
- Garst A, Batey R (2009) A switch in time: detailing the life of a riboswitch. *Biochim Biophys Acta* 1789:584–591, <http://dx.doi.org/10.1016/j.bbagr.2009.06.004>
- Gautheret D, Lambert A (2001) Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *J Mol Biol* 313:1003–1011, <http://dx.doi.org/10.1006/jmbi.2001.5102>
- Giegerich R, Voss B, Rehmsmeier M (2004) Abstract shapes of RNA. *Nucleic Acids Res* 32:4843–4851, <http://dx.doi.org/10.1093/nar/gkh779>
- Glueck T, Draper D (1994) Thermodynamics of folding a pseudoknotted mRNA fragment. *J Mol Biol* 241:246–262, <http://dx.doi.org/10.1006/jmbi.1994.1493>
- Glueck T, Gerstner R, Draper D (1997) Effects of Mg^{2+} , K^+ , and H^+ on an equilibrium between alternative conformations of an RNA pseudoknot. *J Mol Biol* 270:451–463, <http://dx.doi.org/10.1006/jmbi.1997.1119>
- Gräf S, Teune JH, Strothmann D, Kurtz S, Steger G (2006) A computational approach to search for non-coding RNAs in large genomic data. In: Nellen W, Hammann C (eds) *Small RNAs: analysis and regulatory functions*, vol 17, Nucleic acids and molecular biology series. Springer, Berlin, pp 57–74, <http://www.springer.com/life+sci/biochemistry+and+biophysics/book/978-3-540-28129-0>
- Griffiths-Jones S (2004) RALEE-RNA ALignment editor in Emacs. *Bioinformatics* 21:257–259, <http://dx.doi.org/10.1093/bioinformatics/bth489>

3 Methods for Predicting RNA Secondary Structure

- Gruber A, Bernhart S, Hofacker I, Washietl S (2008) Strategies for measuring evolutionary conservation of RNA secondary structures. *BMC Bioinformatics* 9:122, <http://dx.doi.org/10.1186/1471-2105-9-122>
- Gulyaev A (1991) The computer simulation of RNA folding involving pseudoknot formation. *Nucleic Acids Res* 19:2489–2494, <http://dx.doi.org/10.1093/nar/19.9.2489>
- Gulyaev A, van Batenburg F, Pleij C (1999) An approximation of loop free energy values of RNA H-pseudoknots. *RNA* 5:609–617, <http://dx.doi.org/10.1017/S135583829998189X>
- Hamada M, Tsuda K, Kudo T, Kin T, Asai K (2006) Mining frequent stem patterns from unaligned RNA sequences. *Bioinformatics* 22:2480–2487, <http://dx.doi.org/10.1093/bioinformatics/btl431>
- Hamada M, Kiryu H, Sato K, Mituyama T, Asai K (2009) Prediction of RNA secondary structure using generalized centroid estimators. *Bioinformatics* 25:465–473, <http://dx.doi.org/10.1093/bioinformatics/btn601>
- Harmanci A, Sharma G, Mathews D (2007) Efficient pairwise RNA structure prediction using probabilistic alignment constraints in Dynalign. *BMC Bioinformatics* 8:130, <http://www.biomedcentral.com/1471-2105/8/130>
- Harmanci A, Sharma G, Mathews D (2008) PARTS: probabilistic alignment for RNA joint secondary structure prediction. *Nucleic Acids Res* 36:2406–2417, <http://dx.doi.org/10.1093/nar/gkn043>
- Höchsmann M, Voss B, Giegerich R (2004) Pure multiple RNA secondary structure alignments: a progressive profile approach. *IEEE/ACM Trans Comput Biol Bioinform* 1:53–62, <http://dx.doi.org/10.1109/TCBB.2004.11>
- Hofacker I (2003) Vienna RNA secondary structure server. *Nucleic Acids Res* 31:3429–3431, <http://dx.doi.org/10.1093/nar/gkg599>
- Hofacker I, Fontana W, Stadler P, Bonhoeffer S, Tacker M, Schuster P (1994) Fast folding and comparison of RNA structures. *Monatsh Chem* 125:167–188, <http://www.springerlink.com/content/p88384567740kn15/fulltext.pdf>
- Hofacker I, Fekete M, Stadler P (2002) Secondary structure prediction for aligned RNA sequences. *J Mol Biol* 319:1059–1066, [http://dx.doi.org/10.1016/S0022-2836\(02\)00308-X](http://dx.doi.org/10.1016/S0022-2836(02)00308-X)
- Hofacker I, Bernhart S, Stadler P (2004) Alignment of RNA base pairing probability matrices. *Bioinformatics* 20:2222–2227, <http://bioinformatics.oxfordjournals.org/cgi/reprint/20/14/2222>
- Holmes I (2005) Accelerated probabilistic inference of RNA structure evolution. *BMC Bioinformatics* 6:73, <http://www.biomedcentral.com/1471-2105/6/73>
- Huang C, Lu C, Chiu H (2005) A heuristic approach for detecting RNA H-type pseudoknots. *Bioinformatics* 21:3501–3508, <http://dx.doi.org/10.1093/bioinformatics/bti568>
- Jossinet F, Westhof E (2005) Sequence to structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics* 21:3320–3321, <http://dx.doi.org/10.1093/bioinformatics/bti504>
- Katoh K, Toh H (2008) Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. *BMC Bioinformatics* 9:212, <http://dx.doi.org/10.1186/1471-2105-9-212>
- Kiryu H, Tabei Y, Kin T, Asai K (2007) Murel: a practical multiple alignment tool for structural RNA sequences. *Bioinformatics* 23:1588–1598, <http://bioinformatics.oxfordjournals.org/cgi/reprint/23/13/1588.pdf>
- Klein R, Eddy S (2003) RSEARCH: finding homologs of single structured RNA sequences. *BMC Bioinformatics* 4:44, <http://www.biomedcentral.com/1471-2105/4/44>
- Klump H (1977) Thermodynamic values of the helix-coil transition of DNA in the presence of quaternary ammonium salt. *Biochim Biophys Acta: Nucleic Acids and Protein Synthesis* 475:605–610, [http://dx.doi.org/10.1016/0005-2787\(77\)90321-5](http://dx.doi.org/10.1016/0005-2787(77)90321-5)
- Knudsen B, Hein J (2003) Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Res* 31:3423–3428, <http://nar.oxfordjournals.org/cgi/reprint/31/13/3423.pdf>
- Leontis N, Stombaugh J, Westhof E (2002) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res* 30:3497–3531, <http://dx.doi.org/10.1093/nar/gkf481>

- Lindgreen S, Gardner P, Krogh A (2007) MASTR: multiple alignment and structure prediction of non-coding RNAs using simulated annealing. *Bioinformatics* 23:3304–3311, <http://dx.doi.org/10.1093/bioinformatics/btm525>
- Liu B, Shankar N, Turner D (2010) Fluorescence competition assay measurements of free energy changes for RNA pseudoknots. *Biochemistry* 49:623–634, <http://dx.doi.org/10.1021/bi901541j>
- Lu Z, Gloor J, Mathews D (2009) Improved RNA secondary structure prediction by maximizing expected pair accuracy. *RNA* 15:1805–1813, <http://dx.doi.org/10.1261/ma.1643609>
- Lyngsø R, Pedersen C (2000) RNA pseudoknot prediction in energy-based models. *J Comput Biol* 7:409–427, <http://dx.doi.org/10.1089/106652700750050862>
- Macke T, Ecker D, Gutell R, Gautheret D, Case D, Sampath R (2001) RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res* 29:4724–4735, <http://dx.doi.org/10.1093/nar/29.22.4724>
- Markham N, Zuker M (2005) DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res* 33:W577–W581, <http://dx.doi.org/10.1093/nar/gki591>
- Markham N, Zuker M (2008) UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol* 453:3–31, http://dx.doi.org/10.1007/978-1-60327-429-6_1
- Mathews D, Sabina J, Zuker M, Turner D (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 288:911–940, <http://dx.doi.org/10.1006/jmbi.1999.2700>
- Mathews D, Disney M, Childs J, Schroeder S, Zuker M, Turner D (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci USA* 101:7287–7292, <http://dx.doi.org/10.1073/pnas.0401799101>
- Matsui H, Sato K, Sakakibara Y (2005) Pair stochastic tree adjoining grammars for aligning and predicting pseudoknot RNA structures. *Bioinformatics* 21:2611–2617, <http://dx.doi.org/10.1093/bioinformatics/bti385>
- McConaughy B, Laird C, McCarthy B (1969) Nucleic acid reassociation in formamide. *Biochemistry* 8:3289–3295, <http://dx.doi.org/10.1021/bi00836a024>
- Meyer I, Miklós I (2007) SimulFold: simultaneously inferring RNA structures including pseudoknots, alignments, and trees using a Bayesian MCMC framework. *PLoS Comput Biol* 3:e149, <http://dx.doi.org/10.1371/journal.pcbi.0030149>
- Michov B (1986) Specifying the equilibrium constants in Tris-borate buffers. *Electrophoresis* 7:150–151, <http://dx.doi.org/10.1002/elps.1150070310>
- Moretti S, Wilm A, Higgins D, Xenarios I, Notredame C (2008) R-Coffee: a web server for accurately aligning noncoding RNA sequences. *Nucleic Acids Res* 36:W10–W13, <http://dx.doi.org/10.1093/nar/gkn278>
- Mosig A, Zhu L, Stadler P (2009) Customized strategies for discovering distant ncRNA homologs. *Brief Funct Genomic Proteomic* 8:451–460, <http://dx.doi.org/10.1093/bfgp/elp035>
- Nagel J, Pleij C (2002) Self-induced structural switches in RNA. *Biochimie* 84:913–923, [http://dx.doi.org/10.1016/S0300-9084\(02\)01448-7](http://dx.doi.org/10.1016/S0300-9084(02)01448-7)
- Nawrocki E, Eddy S (2007) Query-dependent banding (QDB) for faster RNA similarity searches. *PLoS Comput Biol* 3:e56, <http://dx.doi.org/10.1371/journal.pcbi.0030056>
- Nissen P, Ippolito J, Ban N, Moore P, Steitz T (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc Natl Acad Sci USA* 98:4899–4903, <http://dx.doi.org/10.1073/pnas.081082398>
- Nixon P, Giedroc D (1998) Equilibrium unfolding (folding) pathway of a model H-type pseudoknotted RNA: the role of magnesium ions in stability. *Biochemistry* 37:16116–16129, <http://dx.doi.org/10.1021/bi981726z>
- Nixon P, Giedroc D (2000) Energetics of a strongly pH dependent RNA tertiary structure in a frameshifting pseudoknot. *J Mol Biol* 296:659–671, <http://dx.doi.org/10.1006/jmbi.1999.3464>
- Nussinov R, Pieczenik G, Griggs J, Kleitman D (1978) Algorithms for loop matchings. *SIAM J Appl Math* 35:68–82, <http://dx.doi.org/10.1137/0135006>

- Perriquet O, Touzet H, Dauchet M (2003) Finding the common structure shared by two homologous RNAs. *Bioinformatics* 19:108–116, <http://bioinformatics.oxfordjournals.org/cgi/reprint/19/1/108.pdf>
- Pleij C, Rietveld K, Bosch L (1985) A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res* 13:1717–1731, <http://dx.doi.org/10.1093/nar/13.5.1717>
- Qiu H, Kaluarachchi K, Du Z, Hoffman D, Giedroc D (1996) Thermodynamics of folding of the RNA pseudoknot of the T4 gene 32 autoregulatory messenger RNA. *Biochemistry* 35:4176–4186, <http://dx.doi.org/10.1021/bi9527348>
- Ramesh A, Winkler W (2010) Magnesium-sensing riboswitches in bacteria. *RNA Biol* 7:77–83, <http://dx.doi.org/10.4161/rna.7.1.10490>
- Record M, Lohman T (1978) A semiempirical extension of polyelectrolyte theory to the treatment of oligoelectrolytes: Application to oligonucleotide helix-coil transitions. *Biopolymers* 17:159–166, <http://dx.doi.org/10.1002/bip.1978.360170112>
- Reeder J, Giegerich R (2004) Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics. *BMC Bioinformatics* 5:104, <http://dx.doi.org/10.1186/1471-2105-5-104>
- Reeder J, Giegerich R (2005) Consensus shapes: an alternative to the Sankoff algorithm for RNA consensus structure prediction. *Bioinformatics* 21:3516–3523, <http://dx.doi.org/10.1093/bioinformatics/bti577>
- Reeder J, Steffen P, Giegerich R (2007) pknotsRG: RNA pseudoknot folding including near-optimal structures and sliding windows. *Nucleic Acids Res* 35:W320–W324, <http://dx.doi.org/10.1093/nar/gkm258>
- Ren J, Rastegari B, Condon A, Hoos H (2005) HotKnots: heuristic prediction of RNA secondary structures including pseudoknots. *RNA* 11:1494–1504, <http://dx.doi.org/10.1261/rna.7284905>
- Reuter J, Mathews D (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *Bioinformatics* 11:129, <http://dx.doi.org/10.1186/1471-2105-11-129>
- Riesner D, Steger G (1990) Viroids and viroid-like RNA. In: Saenger W (ed) *Nucleic acids, subvolume d, physical data II, theoretical investigations*, Landolt-Börnstein, group vii biophysics, vol 1. Springer, Berlin, pp 194–243
- Rietveld K, Van Poelgeest R, Pleij C, Van Boom J, Bosch L (1982) The tRNA-like structure at the 3' terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA. *Nucleic Acids Res* 10:1929–1946, <http://dx.doi.org/10.1093/nar/10.6.1929>
- Rivas E, Eddy S (1999) A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J Mol Biol* 285:2053–2068, <http://dx.doi.org/10.1006/jmbi.1998.2436>
- Ruan J, Stormo G, Zhang W (2004) An iterated loop matching approach to the prediction of RNA secondary structures with pseudoknots. *Bioinformatics* 20:58–66, <http://bioinformatics.oxfordjournals.org/cgi/reprint/20/1/58.pdf>
- Sadhu C, Gedamu L (1987) In vitro synthesis of double stranded RNA and measurement of thermal stability: effect of base composition, formamide and ionic strength. *Biochem Int* 14:1015–1022
- Sankoff D (1985) Simultaneous solution of the RNA folding, alignment and protosequence problems. *SIAM J Appl Math* 45:810–825, <http://dx.doi.org/10.1137/0145048>
- Seibel P, Müller T, Dandekar T, Schultz J, Wolf M (2006) 4SALE—a tool for synchronous RNA sequence and secondary structure alignment and editing. *BMC Bioinformatics* 7:498, <http://dx.doi.org/10.1186/1471-2105-7-498>
- Serra M, Turner D (1995) Predicting thermodynamic properties of RNA. *Methods Enzymol* 259:242–261
- Shelton V, Sosnick T, Pan T (1999) Applicability of urea in the thermodynamic analysis of secondary and tertiary RNA folding. *Biochemistry* 38:16831–16839, <http://dx.doi.org/10.1021/bi991699s>
- Soto A, Misra V, Draper D (2007) Tertiary structure of an RNA pseudoknot is stabilized by “diffuse” Mg²⁺ ions. *Biochemistry* 46:2973–2983, <http://dx.doi.org/10.1021/bi0616753>
- Sperschneider J, Datta A (2008) KnotSeeker: heuristic pseudoknot detection in long RNA sequences. *RNA* 14:630–640, <http://dx.doi.org/10.1261/rna.968808>

- Sperschneider J, Datta A (2010) DotKnot: pseudoknot prediction using the probability dot plot under a refined energy model. *Nucleic Acids Res* 38:e103, <http://dx.doi.org/10.1093/nar/gkq021>
- Steffen P, Voss B, Rehmsmeier M, Reeder J, Giegerich R (2006) RNASHAPes: an integrated RNA analysis package based on abstract shapes. *Bioinformatics* 22:500–503, <http://bioinformatics.oxfordjournals.org/cgi/reprint/22/4/500.pdf>
- Steger G (2004) Secondary structure prediction. In: Bindereif A, Hartmann R, Schön A, Westhof E (eds) *Handbook of RNA biochemistry*. Wiley-VCH, Weinheim, pp 513–535, <http://www.wiley-vch.de/publish/en/books/bySubjectCH00/bySubSubjectCHB1/3-527-30826-1/?sID=18eede5181376097ccb16fc47772d157>
- Steger G, Müller H, Riesner D (1980) Helix-coil transitions in double-stranded viral RNA: fine resolution melting and ionic strength dependence. *Biochim Biophys Acta* 606:274–284
- Stombaugh J, Zirbel C, Westhof E, Leontis N (2009) Frequency and isostericity of RNA base pairs. *Nucleic Acids Res* 37:2294–2312, <http://dx.doi.org/10.1093/nar/gkp011>
- Tabaska J, Cary R, Gabow H, Stormo G (1998) An RNA folding method capable of identifying pseudoknots and base triples. *Bioinformatics* 14:691–699, <http://bioinformatics.oxfordjournals.org/cgi/reprint/14/8/691>
- Tabei Y, Tsuda K, Kin T, Asai K (2006) SCARNA: fast and accurate structural alignment of RNA sequences by matching fixed-length stem fragments. *Bioinformatics* 22:1723–1729, <http://bioinformatics.oxfordjournals.org/cgi/reprint/22/14/1723>
- Tabei Y, Kiryu H, Kin T, Asai K (2008) A fast structural multiple alignment method for long RNA sequences. *BMC Bioinformatics* 9:33, <http://dx.doi.org/10.1186/1471-2105-9-33>
- Taufner M, Licon A, Araiza R, Mireles D, van Batenburg F, Gulyaev A, Leung M (2008) PseudoBase++: an extension of PseudoBase for easy searching, formatting and visualization of pseudoknots. *Nucleic Acids Res* 37:D127–D135, <http://dx.doi.org/10.1093/nar/gkn806>
- Taylor W (2007) Protein knots and fold complexity: some new twists. *Comput Biol Chem* 31:151–162, <http://dx.doi.org/10.1016/j.compbiolchem.2007.03.002>
- Theimer C, Giedroc D (1999) Equilibrium unfolding pathway of an H-type RNA pseudoknot which promotes programmed -1 ribosomal frameshifting. *J Mol Biol* 289:1283–1299, <http://dx.doi.org/10.1006/jmbi.1999.2850>
- Theimer C, Giedroc D (2000) Contribution of the intercalated adenosine at the helical junction to the stability of the gag-pro frameshifting pseudoknot from mouse mammary tumor virus. *RNA* 6:409–421, <http://dx.doi.org/10.1017/S1355838200992057>
- Theimer C, Wang Y, Hoffman D, Krisch H, Giedroc D (1998) Non-nearest neighbor effects on the thermodynamics of unfolding of a model mRNA pseudoknot. *J Mol Biol* 279:545–564, <http://dx.doi.org/10.1006/jmbi.1998.1812>
- Tinoco I, Bustamante C (1999) How RNA folds. *J Mol Biol* 293:271–281, <http://dx.doi.org/10.1006/jmbi.1999.3001>
- Torarinsson E, Lindgreen S (2008) WAR: webserver for aligning structural RNAs. *Nucleic Acids Res* 36:W79–W84, <http://dx.doi.org/10.1093/nar/gkn275>
- Torarinsson E, Havgaard J, Gorodkin J (2007) Multiple structural alignment and clustering of RNA sequences. *Bioinformatics* 23:926–932, <http://bioinformatics.oxfordjournals.org/cgi/reprint/23/8/926.pdf>
- Touzet H, Perriquet O (2004) CARNAC: folding families of related RNAs. *Nucleic Acids Res* 32:W142–W145, <http://dx.doi.org/10.1093/nar/gkh415>
- Tullius T, Greenbaum J (2005) Mapping nucleic acid structure by hydroxyl radical cleavage. *Curr Opin Chem Biol* 9:127–134, <http://dx.doi.org/10.1016/j.cbpa.2005.02.009>
- van Batenburg F, Gulyaev A, Pleij C (2001) PseudoBase: structural information on RNA pseudoknots. *Nucleic Acids Res* 29:194–195, <http://dx.doi.org/10.1093/nar/29.1.194>
- Varani G (1995) Exceptionally stable nucleic acid hairpins. *Annu Rev Biophys Biomol Struct* 24:379–404, <http://dx.doi.org/10.1146/annurev.bb.24.060195.002115>
- Walter A, Turner D, Kim J, Lyttle M, Muller P, Mathews D, Zuker M (1994) Coaxial stacking of helices enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc Natl Acad Sci USA* 91:9218–9222

- Waterman M (1995) Introduction to computational biology. Maps, sequences and genomes. Chapman & Hall, London
- Waterman M, Byers T (1985) A dynamic programming algorithm to find all solutions in a neighborhood of the optimum. *Math Biosci* 77:179–188, [http://dx.doi.org/10.1016/0025-5564\(85\)90096-3](http://dx.doi.org/10.1016/0025-5564(85)90096-3)
- Will S, Reiche K, Hofacker I, Stadler P, Backofen R (2007) Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comput Biol* 3:e65, <http://dx.doi.org/10.1371/journal.pcbi.0030065>
- Wilm A, Mainz I, Steger G (2006) An enhanced RNA alignment benchmark for sequence alignment programs. *Algorithms Mol Biol* 1:19, <http://dx.doi.org/10.1186/1748-7188-1-19>, <http://www.biomedcentral.com/content/pdf/1748-7188-1-19.pdf>
- Wilm A, Higgins D, Notredame C (2008a) R-Coffee: a method for multiple alignment of non-coding RNA. *Nucleic Acids Res* 36:e52, <http://dx.doi.org/10.1093/nar/gkn174>
- Wilm A, Linnenbrink K, Steger G (2008b) ConStruct: improved construction of RNA consensus structures. *BMC Bioinformatics* 9:219, <http://dx.doi.org/10.1186/1471-2105-9-219>
- Wimberly B, Varani G, Tinoco I (1993) The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry* 32:1078–1087, <http://dx.doi.org/10.1021/bi00055a013>
- Wu J, Gardner D, Ozer S, Gutell R, Ren P (2009) Correlation of RNA secondary structure statistics with thermodynamic stability and applications to folding. *J Mol Biol* 391:769–783, <http://dx.doi.org/10.1016/j.jmb.2009.06.036>
- Wuchty S, Fontana W, Hofacker I, Schuster P (1999) Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers* 49:145–165, <http://www3.interscience.wiley.com/journal/40003742/abstract?CRETRY=1&SRETRY=0>
- Wyatt J, Puglisi J, Tinoco I (1990) RNA pseudoknots. Stability and loop size requirements. *J Mol Biol* 214:455–470, [http://dx.doi.org/10.1016/0022-2836\(90\)90193-P](http://dx.doi.org/10.1016/0022-2836(90)90193-P)
- Xia T, McDowell J, Turner D (1997) Thermodynamics of nonsymmetric tandem mismatches adjacent to G.C base pairs in RNA. *Biochemistry* 36:12486–12497, <http://dx.doi.org/10.1021/bi971069v>
- Xia T, SantaLucia J, Burkard M, Kierzek R, Schroeder S, Jiao X, Cox C, Turner D (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* 37:14719–14735, <http://dx.doi.org/10.1021/bi9809425>
- Xu X, Ji Y, Stormo G (2007) RNA sampler: a new sampling based algorithm for common RNA secondary structure prediction and structural alignment. *Bioinformatics* 23:1883–1891, <http://dx.doi.org/10.1093/bioinformatics/btm272>
- Yao Z, Weinberg Z, Ruzzo W (2005) CMfinder—a covariance model based RNA motif finding algorithm. *Bioinformatics* 22:445–452, <http://dx.doi.org/10.1093/bioinformatics/btk008>
- Zuker M (1989) On finding all suboptimal foldings of an RNA molecule. *Science* 244:48–52, <http://dx.doi.org/10.1126/science.2468181>
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406–3415, <http://dx.doi.org/10.1093/nar/gkg595>
- Zuker M, Stiegler P (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* 9:133–148, <http://nar.oxfordjournals.org/cgi/reprint/9/1/133>