Chapter 3

Computational Methods for RNA Structure Prediction and Analysis

David Dufour,[a,b] Emidio Capriotti,[c] and Marc A. Marti-Renom[a,b,*]

[a] *Genome Biology Group. Structural Genomics Team. Centre Nacional d'Anàlisi Genòmic (CNAG), Barcelona, Spain*
[b] *Structural Genomics Group. Centre de Regulació Genòmica (CRG), Barcelona, Spain*
[c] *Department of Mathematics and Computer Science, University of Balearic Islands, Palma de Mallorca, Spain*
mmarti@pcb.ub.cat

RNA cannot be considered anymore as a simple transfer molecule. On the contrary, a plethora of noncoding RNA molecules are being discovered, which is transforming our thinking about how the cell is regulated. Large and small RNAs carry now a large repertory of diverse biological functions within cells. Altogether, RNA is now considered as a major player in the molecular regulation of essential cellular processes. Similar to proteins, RNAs adopt three-dimensional (3D) structures that are necessary for performing their functional roles. Unfortunately, despite advances in understanding the folding and unfolding of RNA molecules, our knowledge of the atomic mechanism by which RNA molecules adopt their biologically active structures is still limited. Moreover, experimental

*Parc Científic de Barcelona—Torre I, Baldiri Reixac, 4, 08028 Barcelona, Spain

RNA Nanotechnology
Edited by Bin Wang

ISBN 978-981-4411-64-6 (Hardcover), 978-981-4411-65-3 (eBook)
www.panstanford.com

determination of RNA structures either by X-ray crystallography or nuclear magnetic resonance is challenging, given the instability of RNA molecules. Therefore, computational approaches for predicting the 3D structure of RNAs are becoming essential in the study of the molecular mechanisms of RNA function. Here we start by outlining the general principles of the RNA structure, and then we describe the databases and algorithms for analyzing and predicting RNA secondary and 3D structures.

3.1 Introduction

Ribonucleic acid (RNA) molecules are now known to act upon enzymatic functions [1], gene transcriptional regulation [1–3], protein biosynthesis regulation [4], development [5], and disease [6]. Characterizing the molecular details of such a diverse repertory of functions requires the knowledge of the three-dimensional (3D) structure of the RNA molecules as well as their interaction with other biomolecules in the cell. Since the seventies, when the first RNA structure was determined (i.e., the yeast phenylalanine transfer RNA [tRNA] [7]), the number of known RNA structures has steadily grown, and only recently the increase of new structures has been exponential (Fig. 3.1). However, the first computational algorithms to predict base pairing from an RNA sequence were developed in the late seventies/early eighties [8–10]. Another 10 years passed by before the first RNA 3D structure (i.e., a conserved core of group I introns) was predicted by the Westhof group [11].

Nowadays, only a limited number of automatic predictions of the 3D structures of large RNA molecules has so far been accomplished (for some examples see [12–14]). However, and given the limited number and diversity of known RNA structures, computational algorithms for RNA structure prediction has been one of the sources for characterizing the structural diversity in RNA molecules and its relationship to function [15]. Most of the existing algorithms rely in the principle that RNA folding is a hierarchical process and that knowledge of its secondary structure (i.e., the determination of all base pairing in an RNA sequence) may improve the prediction of its 3D conformation. Consequently, in recent years, several

computational programs have been developed, which try to predict the base pair interactions in RNA from its sequence (see, e.g., [16, 17, 18]). However, the growing amount of available structural data for RNA molecules and the initial attempts for classifying their motifs [19, 20] open the possibility for applying comparative approaches previously used for protein structure prediction [21, 22, 23]. It is known that in general, it will be more difficult to predict large RNA 3D structures based on comparative approaches than predicting protein structures [24]. Such a statement relies on (at least) two properties of RNA: (i) Its folding is essentially driven by its base pair and its regular motifs [25] (in contrast to the hydrophobic forces that drive protein folding [26]), and (ii) within the same functional family, RNA sequence conservation is usually limited to very short fragments of nucleotides, while still maintaining a substantial conservation of their secondary structure [25]. Both principles make necessary that the base pairs of an RNA molecule need be determined (or predicted) before attempting to predict its 3D structure and that reliable comparative approaches be limited to RNA sequences that align with more than 60% sequence identity to a known structure [15].

We begin this chapter by describing the RNA structure and the initial attempts for classifying of the RNA structural space. We continue by outlining recent developments and methods for secondary and tertiary structure prediction from sequences. We then conclude by discussing possible implications of the use of comparative approaches to predict the 3D conformation of RNA sequences based on existing known structures.

3.2 RNA Structure

Structurally, RNA is composed of a combination of riboses, phosphates, and aromatic bases. Riboses and phosphates are connected through the phosphodiester bond, forming a backbone from which the aromatic bases are attached in a regular fashion through the C1′ atom of the ribose moiety. As they are being synthesized, RNA molecules fold mainly by the driving force of hydrogen bonding and stacking interactions between bases. Long stretches of canonical

Watson–Crick (WC) base pairs result in an antiparallel double helix. However, each base has three sides (i.e., WC, Hoogsteen, and sugar sides) that can interact resulting in 28 different base-pairing possibilities between two nucleotides [19]. Nevertheless, canonical helices are maintained by WC base pairing. Additionally, nucleotide bases may also interact with the ribose or phosphate atoms as well as noncanonical base–base interactions, which may result in special RNA geometries. In contrast to deoxyribonucleic acid (DNA), this plethora of possible pairwise interactions between bases results in RNA adopting complex 3D structures. A base pair in RNA is maintained by a minimum of two hydrogen bonds between the paired nucleotide bases. The combination of base pairs in an RNA molecule defines the so-called secondary structure, which is composed of stems (double helices), loops, bulges, stem junctions, and pseudoknots. The final 3D RNA structure is maintained by tertiary interactions, including loop–loop interactions, stem–loop interactions, coaxial stacking, and triple and quadruple helices.

3.2.1 *RNA Base Pairs*

Over the last few years there has been a rapid growth in the number of RNA structures made available through the Protein Data Bank (PDB) (Fig. 3.1) [27, 28]. This increment is mostly due to the recent structural determination of ribosome machineries [29, 30, 31, 32]. Thus, the availability of such data has allowed the application of a more robust classification of base pair interactions in RNA molecules. Although there are differences in the interaction of two RNA bases, a stable classification depending on the edges involved in the interaction (i.e., WC, Hoogsteen, or sugar edges) has already been proposed [33, 34]. In such a classification, each base can form several nonbonded interactions that involve different types of atoms: (i) phosphate–phosphate interaction mediated by water molecules, (ii) phosphate–sugar interaction, (iii) sugar–sugar interaction, (iv) base–phosphate interaction, (v) base–sugar interaction, and (vi) base–base interaction. Moreover, these six different interaction types can be formed in either a *cis* or a *trans* state, resulting in 12 possible different conformations (Fig. 3.2A,B) [35]. Only about 60% of the base pairs in known RNA structures adopt the canonical WC–WC interaction in *cis* conformation.

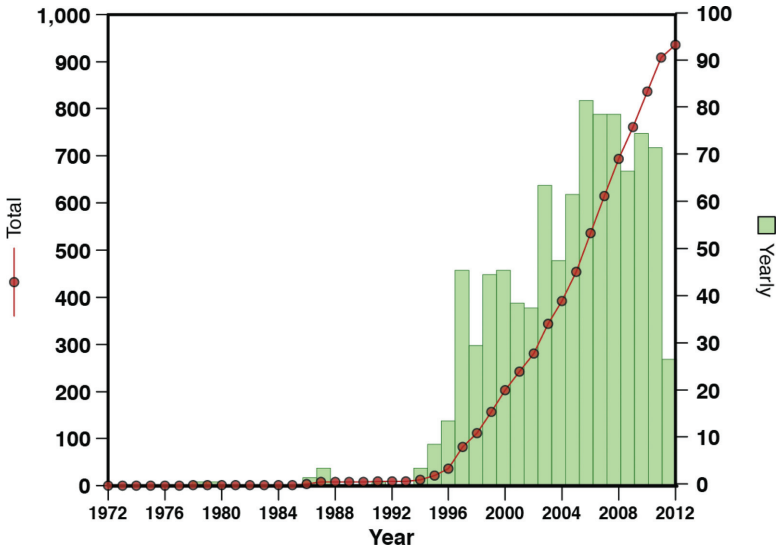

Figure 3.1 RNA structure deposition in the PDB database. Green bars (right y-axis) indicate yearly new PDB entries, and the red line (left y-axis) represents the total number of RNA structures in the PDB database. The data ends in July 2012.

3.2.2 *RNA Backbone*

Differently from proteins, RNA molecules are characterized by well-packed side chains stabilized by hydrogen bonds and a flexible backbone. RNA backbone conformation can be described in terms of the torsion angles α, β, γ, δ, ε, and ζ, while the ribose conformation is determined by the endocyclic torsion angles $?_0$ to $?_4$ (Fig. 3.2C). Richardson *et al*. have analyzed a set of RNA structures with crystallographic resolutions higher than 3 Å and no atom clashes, identifying 42 discrete RNA backbone conformers [36]. Different studies concluded that the RNA backbone is rotameric and can be classified into discrete conformers [37, 38]. These types of analyses have been possible because the quality and amount of determined RNA structures have considerably grown over the last few years [39, 40, 41]. However, most large RNA structures can only be determined at low resolutions. At this resolution, the phosphate and base plane

AQ: Please check what these ? symbol are.

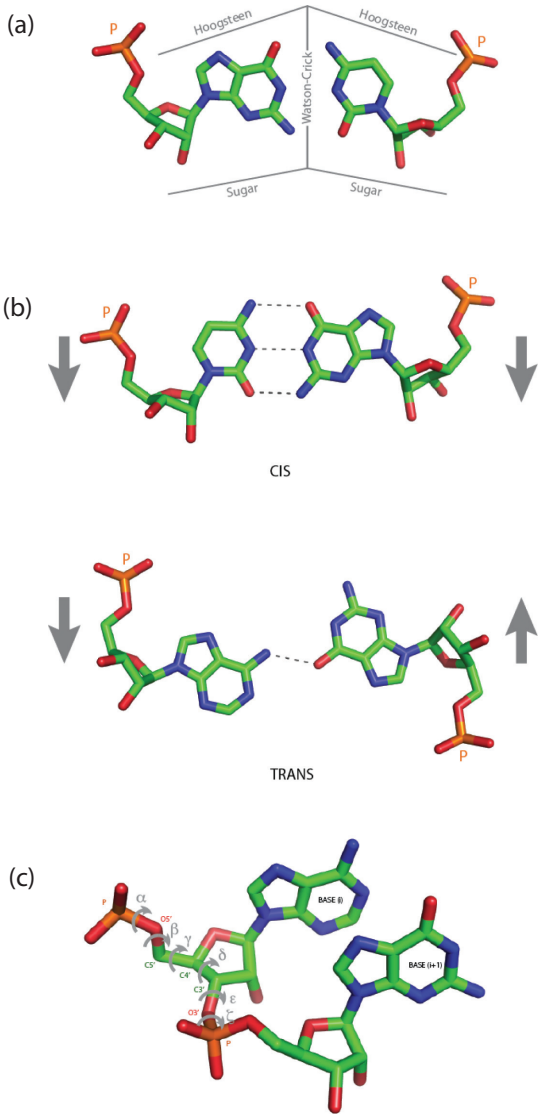

Figure 3.2 Base pair interactions. (A) WC, Hoogsteen, and sugar edges for a base pair interaction. (B) *cis* and *trans* states of a base pair interaction. (C) RNA backbone torsion angles.

can be accurately positioned but the sugar ring and the rest of the backbone atoms may contain errors. Indeed, on average, one error can occur every two bases based on the analysis and classification of RNA torsion angles from the RNABase database [39].

3.2.3 *RNA Motifs*

RNA motifs correspond to recurrent RNA structural elements subject to 3D spatial constraints [20, 42]. This broad definition of RNA motifs already indicates the difficulty in uniquely describing or classifying them. The RNA secondary structure partially explains some of the known RNA motifs such as bulges, hairpins, internal loops, and multihelical motifs. However, the prediction of pseudoknots is a more challenging task in secondary structure prediction programs. Pseudoknots contain two stem–loop motifs in which the first stem's loop forms part of the second stem. Structural data indicates that the final 3D RNA structure is mostly determined by base pair stacking (i.e., WC base pairs) and non-WC interactions. Thus, characterizing, analyzing, and ultimately predicting the stacking of those bases will help the goal of classifying complex RNA motifs.

3.3 RNA Structural Databases

Since the seventies, when the first RNA structures became available [7], there has been an attempt to store, organize, and classify the RNA structural space. Next, we briefly describe available databases that classify RNA structures.

The **NDB** database [27] stores all molecules containing nucleic acids and complements them with additional information such as classification of nucleic acids and their interaction with proteins, backbone conformation angles, and base pair classification. The **SCOR** database [41] organizes RNA motifs in a hierarchical classification system similar to the SCOP database for protein domains [43]. SCOR classifies RNA structures from three properties: first, the RNA structural classification describes RNA motifs according to the number of strands connecting double helices; second, the RNA

functional classification divides each entry by the biological function of their molecule, motif, and structural model; and third, the RNA tertiary interaction group classifies RNA structures by their inter- and intramolecular interactions differing from WC and non-WC base pairs. The SCOR database stores 8,270 structural motifs (October 2004), some of which are further classified into functional and RNA tertiary interaction classes (Table 3.1). The **Rfam** database [44, 45] classifies noncoding RNA molecules into families of members that conserve sequence and secondary structure. The conservation of RNA secondary structure implies a degree of conservation of its function [15]. The **MeRNA** database [46] was manually curated by analyzing each RNA structure and comparing them to previously described binding motifs and includes eight well-characterized metal ion–binding motifs. **PseudoBase** [47], a searchable database of pseudoknot secondary structures, contains over 250 records of pseudoknots determined by crystallography, nuclear magnetic reso-

Table 3.1 RNAbase classification. The number of RNA structure entries stored in RNAbase classified by their functional categories

Category	Entries
tRNAs	217
rRNAs	283
mRNAs	126
Transcription-related RNAs	86
Introns	26
Splicing-related RNAs	59
Signal recognition particle RNAs	22
Ribozymes	115
RNase P	21
Aptamers	30
Pseudoknots	31
Tetraloops	81
Bulges	69
DNA–RNA hybrids	115
PNA–RNA hybrids	1
Drug–RNA complexes	137
Viral and phage RNAs	221

Abbreviations: mRNA, messenger RNA, RNase P, ribonuclease P.

nance (NMR), mutational experiments, and sequence comparisons. **PseudoBase++** [48] is an extension of PseudoBase for searching, formatting, and visualization of pseudoknots. PseudoBase++ links each pseudoknot in PseudoBase to the GenBank record of the corresponding nucleotide sequence and allows scientists to automatically visualize RNA secondary structures with PseudoViewer. It also includes the capabilities of fine-grained reference searching and collecting of new pseudoknot information. The **RNAJunction** database [49] contains structure and sequence information for RNA structural elements such as helical junctions, kissing loops, internal loops, bulges, and loop–loop interactions. The database can be searched using the PDB code, structural classification, sequence, and interhelix angles. The **STRAND** database [50] provides a collection of known RNA secondary structures drawn from diverse public databases. The database is searchable based on one or various criteria defined by the user, like RNA type, organism of origin, external source, length, the number of molecules in the complex, and other features. Moreover, the RNA Secondary Structure Analyzer, a tool developed by the same group to analyze RNA secondary structures, provides comprehensive statistical information on the secondary structures in the database. **MODOMICS** [51] is a database devoted to the systems biology of RNA modification. It provides information on the chemical structure of modified nucleosides, pathways of their biosynthesis, sequences of RNAs containing these modifications, and RNA-modifying enzymes. It contains curated tRNA and ribosomal RNA (rRNA) sequences and all RNAs with 3D structures in the NDB database for which modified nucleosides are known. **BPS** is a database [52] that stores RNA base pair structures with quantitative information on the spatial arrangements of interacting bases, including higher-order base associations and isosteric pairs. The structures are taken from the NDB database, and the base pairs are identified and characterized with the 3DNA software package [53]. The interactions are classified in terms of residue identities, base pair positioning, and hydrogen-bonding patterns and related to the structural context in which they occur. BPS also includes an atlas with representative images of the various base pairs, higher-order base interactions and isosteric pairs, and links to statistical information about these groups of structures.

Table 3.2 SCOR classification. Number of RNA structure motifs stored in the SCOR database classified by structural, functional, and RNA tertiary interaction categories

Classification	Subclasses	RNA motifs
Structural classification	Internal loops	5,350
	Hairpin loops	2,920
Functional classification	Molecular function	480
	Motif function	179
	Structural models	137
RNA tertiary interaction	Coaxial helices	7
	Tetraloop-receptor	1
	A-minor motif	240
	Kissing-hairpin loops	32
	tRNA D-loop:T-loop	7
	Pseudoknots	17
	Ribose zipper	657

The **FRABASE** database [54] allows for the automatic search of 3D RNA fragments within a set of RNA structures by the input of either RNA sequence(s) and/or secondary structure(s). The database contains RNA sequences and secondary structures in the 'dot bracket' notation derived from the PDB, a collection of atom coordinates of unmodified and modified nucleotide residues occurring in RNA structures, calculated RNA torsion angles and sugar pucker parameters, and information about base pairs. The **CoSSMos** database [55] is an online database of 3D characteristics of internal, bulge, and hairpin loops. It contains each loop's structural information, including sugar pucker, glycosidic linkage, hydrogen-bonding patterns, and stacking interaction. Users can search via general PDB information, experimental parameters, sequences, and specific motifs and by specific structural parameters in the subquery page after the initial search.

3.4 RNA Secondary Structure Prediction

Predicting the secondary structure of an RNA sequence can prove very useful for gaining insight into its tertiary structure and its

function [56]. The RNA-folding process is hierarchical [57], which means that local interactions occur first and are energetically stronger than tertiary interactions [58]. Therefore, the RNA secondary structure provides a scaffold to its native 3D structure. This property already indicates that the RNA secondary structure can be predicted without the knowledge of tertiary interactions (to the exception of the so-called pseudoknots). The first methods for predicting the secondary structure of RNA molecules were developed assuming that the minimum free-energy conformation for the native state could be searched by dynamic programming [8, 9] and the Nussinov's algorithm [10]. The scoring functions for such approaches were based on free-energy parameters from physics, which were derived from empirical calorimetric experiments [59] or from known RNA structures deposited in the PDB [17]. Unfortunately, the minimum free-energy (MFE) approach does not guarantee that the selected or predicted final structure will be the native structure and typically corresponds to a near-native conformation [60]. Other implementations of the MFE principle include the use of a heuristic search for suboptimal secondary structures [8, 59, 61], the computation of all suboptimal alignments near the optimal folding space [62], and the selection of suboptimal solutions based on RNA shape analysis [63].

In the 1990, McCaskill first implemented a method based on the equilibrium partition function for secondary structure and associated probabilities of various substructures [64]. Such a method allowed the statistical characterization of the equilibrium ensemble of RNA secondary structures. It has been noted that higher base pair probabilities, computed by the partition function approach, correspond to higher predictive reliability when considering structures determined by comparative sequence analysis [61]. More recently, new computational approaches based on the statistical sampling of known RNA secondary structures [17] or genetic algorithms [65, 66, 67] have also been implemented for secondary structure prediction. However, most of the methods described so far are not suitable for predicting RNA pseudoknots, because they are based on the recursive approach. It has been demonstrated that the prediction of secondary structure motifs with pseudoknots is a NP-complete problem making it computational intractable [68]. To address this

problem, modified dynamic programming and stochastic context-free grammar algorithms [69] have been recently introduced. For example, the **PKNOTS** program implements thermodynamic folding in a rather large subclass of pseudoknots on $O(N^4)$ and $O(N^6)$ time space, which makes it only usable for short sequences [70]. The partition function approach implemented by Dirks [71, 72] has an $O(N^5)$ complexity. Despite this computational complexity, the accuracy for pseudoknots prediction has significantly increased by using an innovative dynamic partner sequence-stacking algorithm [73].

RNA secondary structure prediction from a single sequence somehow neglects the evolutionary forces acting upon RNA sequence variation. Therefore, the inclusion of multiple sequences for predicting the RNA secondary structure allows the incorporation of constraints based on the commonalities of the compared sequences [74]. Evolution tends to conserve the RNA secondary structure more than sequence [15]. Indeed, it is known that a mutation in an RNA molecule is usually compensated by a second mutation that restores base pair interaction [75, 76]. Several methods for secondary structure prediction use this principle by attempting to detect such covariance between different positions in the multiple sequence alignment. An initial implementation of such an approach used mutual information theory to extract the covariance between bases [77, 78]. However, these approaches resulted in limited accuracy [79] and have been replaced by more recent implementations such as the **RNAalifold** program [80], which scores possible solutions by combining free energy with a covariance term; the **Pfold** program [81], which uses an evolutionary SCFG approach; or the **ILM** program [82, 83], which combines thermodynamic and mutual information in a single score. The **Foldalign** program [84, 85] heuristically considers local sequence alignments and the maximum number of base pairs at the same time. The **Dynalign** program [86] is a pairwise alignment method that searches for common low-energy structures between two sequences. The algorithm complexity is reduced considering a maximum value of the sequence distance between two aligned residues and by limiting the size of any internal loop. Finally, the **Carnac** program [87, 88], which is not a strict implementation of a simultaneous align-and-fold approach,

AQ: Please expand this

relies on a thermodynamic model with energy minimization by combining information from locally conserved elements and mutual information between sequences. Finally, the **RNAforester** [89] and **MARNA** [90] programs first fold RNA sequences using single-sequence secondary structure prediction methods and then align the resulting structures using tree-based methods.

3.5 RNA Tertiary Structure Analysis and Prediction

The increase over the last decade of the number of available structures deposited in the PDB, including X-ray and NMR models (Fig. 3.1), has stimulated the structural biology community to develop computational tools for analyzing the RNA structural space. Next, we outline some of the existing methods for RNA structure analysis and prediction.

3.5.1 *RNA 3D Structure Analysis*

The **PRIMOS** program [91] describes an RNA structure with pseudo torsion angles η ($C4'_{i-1}$-P_i-$C4'_i$-P_{i+1}) and θ (P_i-$C4'_i$-P_{i+1}-$C4'_{i+1}$) obtained using the **AMIGOS** program [92]. Then the search comparison is done over such simplified version of the RNA structure, allowing the identification of common small motifs between two RNAs or an RNA structural motif and a database of RNA structures. The **NASSAM** program [93], which was designed for identifying common substructural motifs between two RNA structures, implements a simplified vector representation of each nucleic acid base with respect to its position in the structure. Then the vectors and their edges are transformed into a graph connecting the bases and compared using the Ullman subgraph isomorphism algorithm. The **ARTS** [94] and **DIAL** [95] programs for structural comparison of RNA structures were developed to overcome the limitation of sequence continuity. The ARTS program describes RNA structures by a set of contiguous *quadrats* (i.e., four phosphate atoms located in two successive base pairs). The program then identifies similar *quadrats* between two RNA structures and uses them as seeds for the final alignment. Finally, the algorithm finds the

maximal matching in a bipartite graph between the two structures by extending the structure alignment that maximizes the number of aligned bases and base pairs. The DIAL program uses a dynamic programming algorithm to align two RNA structures based on a scoring function that combines a base, a dihedral angle, and a base-paring similarity measure. DIAL can be run as a web server and provides the user with the option of producing global (Needleman–Wunsch), local (Smith–Waterman), or global/semiglobal (motif search) alignments. The **SARA** program [96] applies a unit-vector root mean square approach to pairwise structural alignment. It can also assign RNA structures to functional classes as defined in the SCOR database [97]. The **SARSA** program [98] uses vectors to obtain a structural alphabet of RNA backbone conformations. The input structures are represented using such an alphabet, and the two RNA structures are finally aligned using dynamic programming based on the alphabet sequence. The **FR3D** program [99], and its web-based interface **WebFR3D** [100], identifies recurrent motifs in a base-centered approach using geometric, symbolic, or sequence information. To score and rank candidate motifs, FR3D calculates a geometric discrepancy by rigidly rotating candidates to optimally align with the query motif and then comparing the relative orientations of the corresponding bases in the query and candidate motifs. The **FASTR3D** program [101] allows users to specify a range of nucleotides from a PDB file as a query to look for similar structures in a list of PDB files using the secondary structure information and backbone torsion angles of the query structure. Alternatively, it can take primary and/or secondary structures as an input. The **RNAMotifScan** program [102] detects similar RNA structure motifs based on the two-dimensional (2D) alignments. It was observed that many noncanonical base pairs in RNA structural motifs are isosteric, and these base pairs can interchange with each other without affecting the overall RNA structure, so RNAMotifScan takes into account isosteric base pairs and multipairings. The **FRASS** program [103] is capable of handling large RNA fragments and is designed for global similarity searching. The user can select an entire chain from a PDB file or upload a structure to the server. The searching method is based on Gauss integrals that are used to compare the shapes of backbones of RNA molecules.

The **WebR3DAlign** server [104] identifies all motifs conserved in the 3D structures of two possibly homologous RNA molecules. It produces a nucleotide-to-nucleotide alignment of two 3D structures to identify conserved motifs, while allowing for differences in the global structure of the molecules, like domain motions.

3.5.2 *RNA 3D Structure Prediction*

Predicting the 3D structure of an RNA molecule is straightforward and usually requires human intervention [105]. In contrast to the current status of protein structure prediction, a fully automated approach is not able to reliably predict a large RNA 3D structure from its sequence [24]. However, over the last few years, a plethora of methods has been developed that aid the manual or semiautomatic prediction of RNA structures. For example, the **ERNA-3D** program [12] automatically generates an RNA 3D structure, starting for its secondary structure. ERNA-3D, which has successfully been used to model the structure of transfer-messenger RNA molecules [13], is able to model RNA motifs by using high-resolution structural information from the SCOR database. The **MANIP** program [106] builds complete RNA structural models based on the assembly of RNA motifs or fragments from a selected library. The final refinement protocol combines canonical as well as noncanonical base-pairing constraints with restraints imposed by covalent geometry, stereochemistry, and van der Waals contacts. The **MC-Sym** program [107] builds RNA 3D structures using the coordinates and relations between bases from known RNA structures. Additional constraints can be applied to the model during the building procedure to ensure the conservation of particular structural features. Mc-Sym uses molecular dynamic simulations to minimize the energy of the predicted structure. The **RNA2D3D** program [108] builds RNA structural models by first spacing the atoms of a nucleotide along a fixed backbone and then predicting the final structure of the model by an helical winding procedure. The model is further refined by interactively moving groups of nucleotides to better-fit, known structural information or by minimizing it using molecular dynamics simulations. The **iFoldRNA** program [109] uses coarse-grained structural models to perform molecular dynamics simulations of

AQ: Should this be 'from'?

RNA structures. iFoldRNA has been used to predict the structure of RNA molecules smaller than 50 nucleotides to near-atomic resolution (i.e., 2 to 5 Å RMSD to its native structure). The **NAST** program [14] uses a knowledge-based, coarse-grained dynamics engine for modeling RNA structures. NAST allows the end user to provide secondary or tertiary experimentally derived restraints to filter the predicted 3D models. The **BARNACLE** program [110], which introduced a probabilistic model of the RNA structure, allows an efficient sampling of RNA conformations in continuous space. The **Rosetta/FARFAR** program [111, 112], inspired by the protein structure prediction method [113], has been applied to predict the 3D structure of 20 RNA sequences of ∼30 nucleotides. The authors report that their method is able to correctly predict the native conformation for ∼90% of WC and about one-third of non-WC base pairs. Their results also suggest that improvements in the energy function together with the use of predictions from phylogenetic approaches are necessary for an accurate structure prediction of more complex RNA molecules. The **RNAmoIP** framework [114, 115] uses integer programming to refine predicted or known secondary structures to accommodate the insertion of RNA 3D motifs. Then, the predictions are used as templates to generate complete 3D structures with the MC-Sym program. Integer programming techniques have gained a lot of interest recently as they provide state-of-the-art methods for predicting RNA secondary structures with pseudoknots. The **RNABuilder** program [116] is a software package that generates model RNA structures by treating the kinematics and forces at separate multiple levels of resolution. Kinematically, bonds in bases, certain stretches of residues, and some entire molecules are rigid, while other bonds remain flexible. Forces act on the rigid bases and selected individual atoms. The **Assemble** program [117], a graphical user interface (GUI) semiautomated modeling program that can be performed by homology and ab initio with or without electron density maps, allows interactive editing of the secondary structure and the use of a library of annotated tertiary structures. It combines automated and manual protocols within an iterative modeling process, where the user can insert 3D motifs and modify backbone angles of a coarse-grained input structure. **ModeRNA** [118] is a program for comparative modeling

AQ: Please expand this.

of RNA 3D structures. It requires a pairwise sequence alignment and a structural template to generate a 3D structural model of the target RNA sequence via either fully automated or script-based approaches. ModeRNA is capable of handling 115 different nucleotide modifications and bridging gaps using fragments derived from an extensive fragment library. The **Vfold** program [119] evaluates the stable structures and the folding free energies for RNA secondary structures and pseudoknotted structures. The program predicts the secondary structure of the RNA sequence allowing the build of a coarse-grained structure, which is later refined with an all-atom representation that relies in a library of PDB fragments. **RAGPOOLS** is a recent application of a graph theoretic approach to represent RNA molecules [120, 121], which allows for a simplified representation of RNA structural motifs that are then used for predicting and designing new complex structures of RNA molecules. Finally, the **RSIM** program [122] provides a fully automated application predicting RNA tertiary structures using fragment assembly from secondary structure constraints predicted by the ViennaRNA package. These tertiary structures are further refined with Monte Carlo simulations utilizing a novel sampling method, an expanded statistical potential, and a diverse fragment library. Finally, RSIM stores the refinement paths, which allows the representation of the predicted RNA conformational space as a graph with secondary structures as nodes and simulation paths as edges.

3.5.3 *RNA 3D Structure Assessment*

The large number of new tools for predicting the RNA 3D structure is likely to result in an increasing number of predicted structures that will need to be assessed. According to this, the scientific community is now using known 3D structures deposited in the PDB to develop knowledge-based potentials of mean force to assess the accuracy of predicted RNA structures. In the 2009, Jonikas *et al*. presented the first coarse-grained knowledge-based function to select native-like structures [14]. The statistical potential is embedded in the **NAST** algorithm and assesses the RNA structure using a representation based on C3′ atoms. The method has been validated against thousands of decoys and assessing how well it samples RNA

structures close to the solved crystal structures. Unfortunately, the NAST knowledge-based potential provides a nucleotide-based score that cannot distinguish between models with different atom positions except for C3′. The NAST algorithm was successfully used to generate and properly assess the accuracy of a modeled tRNA and P4–P6 structures. Later, the **FARFAR** method [112] implemented a statistical potential developed to score the structures predicted using the FARNA algorithm. The scoring function in the FARFAR algorithm is composed of several terms, including pairwise distance-dependent potentials, hydrogen bonds, hydrophobic and hydrophilic contributions, and a term describing the screened electrostatic interactions between phosphates. FARFAR was derived using a dataset of 32 motifs automatically extracted from a set of high-resolution crystallographic structures of ribozymes, riboswitches, and other noncoding RNAs. The **RASP** potential [123] is based on a distance-dependent scoring function including four different types of RNA molecule representations with increasing level of complexity. The simplest RASP potential included four atom types corresponding to the four C3′ atoms of each nucleotide. Other representations consisted of 28 atom types for the backbone atoms, 44 atom types for the backbone and sugar ring, and 23 for all atoms without taking into consideration the differences between nucleotides. All the RASP knowledge-based potentials were tested using a leave-one-out procedure over a set of 85 nonidentical RNA structures and their associated decoys sets composed of 500 structures with different RMSD and GDT-TS values. The RASP tool was favorably compared with previously existing methods such as NAST, FARFAR and AMBER [124]. The results showed that the RASP full atom was the most accurate method in the ranking native structures. More recently, Bernauer *et al*. have developed another distance-dependent statistical potential based on the Dirichlet process mixture model [125]. This procedure allows us to obtain an analytical form of the potential as a sum of Gaussian functions that makes the scoring function fully differentiable and suitable for energy minimization or molecular dynamics. The method has been trained using a set of 77 nonredundant and high-quality RNA structures, and its performances have been evaluated over a set of decoys generated by molecular dynamics simulations and a normal-

AQ: Please expand this.
AQ: If these are abbreviations, please expand them. Please check all abbreviations in the chapter.

mode perturbation method. The method has been compared with FARFAR, resulting in similar accuracies.

3.6 Perspectives

The increase in the number of known RNA structures in the PDB clearly shows the existence of regular and recurrent 3D RNA motifs. Thus, the next logical step for structural biologists is to detect, store, analyze, and classify such structural motifs to aid in ab initio and/or knowledge-based structural prediction of whole RNA sequences [19]. The current amount and diversity of known structures of RNA molecules has allowed the development of a plethora of approaches for RNA structure prediction, which have been briefly outlined in this chapter. However, it is difficult to predict whether such methods will be readily applicable to RNA and, more importantly, will result in reliable models. The first collective blind test experiment in RNA 3D structure prediction (called RNA-Puzzles) was recently organized. RNA-Puzzles is a CASP-like experiment that aims at evaluating the accuracy of both manual and automatic methods for RNA structure prediction [24]. The results from the RNA-Puzzles experiment provide deeper insights into the accuracy of available methods for different applications at the same time that stimulate the RNA structure prediction community for its ongoing efforts to improve its tools. In its first edition, seven different research groups that tried to predict the structure of several RNA molecules within three different types of scenarios participated in the RNA-Puzzles experiment. Overall, the methods implemented by the Bujnicki, Chen, and Das groups were scored among the top methods [24]. The website for automatic RNA model evaluation and additional information about the RNA-Puzzles experiment is http://paradise-ibmc.u-strasbg.fr/rnapuzzles/.

AQ: Please expand this.

Acknowledgments

MAM-R acknowledges support from the Spanish Ministry of Economy and Competitively (BFU2010-19310). EC acknowledges

support from the European Community through the Marie Curie International Outgoing Fellowship program (PIOF-GA-2009-237225). This chapter is an updated version of our previous review [126].

References

1. Staple, D.W., and Butcher, S.E. (2005). Pseudoknots: RNA structures with diverse functions. *PLoS Biol.*, **3**, e213.
2. Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
3. Dorsett, Y., and Tuschl, T. (2004). siRNAs: applications in functional genomics and potential as therapeutics. *Nat. Rev. Drug Discovery*, **3**, 318–329.
4. Doudna, J.A. (2000). Structural genomics of RNA. *Nat. Struct. Biol.*, **7**(Suppl.), 954–956.
5. Aravin, A.A., *et al*. (2003). The small RNA profile during Drosophila melanogaster development. *Dev. Cell*, **5**, 337–350.
6. Lu, J., *et al*. (2005). MicroRNA expression profiles classify human cancers. *Nature*, **435**, 834–838.
7. Kim, S.H., *et al*. (1974). Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science*, **185**, 435–440.
8. Zuker, M., and Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.*, **9**, 133–148.
9. Zuker, M., and Sankoff, D. (1984). RNA secondary structure and their prediction. *Bull. Math. Biol.*, **46**, 591–621.
10. Nussinov, R., and Jacobson, A.B. (1980). Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc. Natl. Acad. Sci. U S A.*, **77**, 6309–6313.
11. Michel, F., and Westhof, E. (1990). Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.*, **216**, 585–610.
12. Zwieb, C., and Muller, F. (1997). Three-dimensional comparative modeling of RNA. *Nucleic Acids Symp. Ser.*, 69–71.
13. Burks. J., Zwieb, C., Muller, F., Wower, I., and Wower, J. (2005). Comparative 3-D modeling of tmRNA. *BMC Mol. Biol.*, **6**, 14.

14. Jonikas, M.A., *et al*. (2009). Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA*, **15**, 189–199.
15. Capriotti, E., and Marti-Renom, M.A. (2010). Quantifying the relationship between sequence and three-dimensional structure conservation in RNA. *BMC Bioinf.*, **11**, 322.
16. Hofacker, I.L. (2003). Vienna RNA secondary structure server. *Nucleic Acids Res.*, **31**, 3429–3431.
17. Do, C.B., Woods, D.A., and Batzoglou, S. (2006). CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics*, **22**, e90–98.
18. Dowell, R.D., and Eddy, S.R. (2006). Efficient pairwise RNA structure prediction and alignment using sequence alignment constraints. *BMC Bioinf.*, **7**, 400.
19. Leontis, N.B., *et al*. (2006). The RNA Ontology Consortium: an open invitation to the RNA community. *RNA*, **12**, 533–541.
20. Leontis, N.B., Lescoute, A., and Westhof, E. (2006). The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.*, **16**, 279–287.
21. Baker, D., and Sali, A. (2001). Protein structure prediction and structural genomics. *Science*, **294**, 93–96.
22. Marti-Renom, M.A., *et al*. (2000). Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 291–325.
23. Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.
24. Cruz, J.A., *et al*. (2012). RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA*, **18**, 610–625.
25. Schuster, P., Stadler, P.F., and Renner, A. (1997). RNA structures and folding: from conventional to new issues in structure predictions. *Curr. Opin. Struct. Biol.*, **7**, 229–235.
26. Dill, K.A. (1990). Dominant forces in protein folding. *Biochemistry*, **29**, 7133–7155.
27. Berman, H.M., *et al*. (1992). The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.*, **63**, 751–759.
28. Berman, H.M., *et al*. (2002). The Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.*, **58**, 899–907.

29. Schuwirth, B.S., *et al*. (2005). Structures of the bacterial ribosome at 3.5 A resolution. *Science*, **310**, 827–834.
30. Gao, H., Ayub, M.J., Levin, M.J., and Frank, J. (2005). The structure of the 80S ribosome from Trypanosoma cruzi reveals unique rRNA components. *Proc. Natl. Acad. Sci. U S A*, **102**, 10206–10211.
31. Noller, H.F. (2005). RNA structure: reading the ribosome. *Science*, **309**, 1508–1514.
32. Felden, B. (2007). RNA structure: experimental analysis. *Curr. Opin. Microbiol*., **10**, 286–291.
33. Leontis, N.B., and Westhof, E. (2001). Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.
34. Leontis, N.B., Stombaugh, J., and Westhof, E. (2002). The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res*., **30**, 3497–3531.
35. Yang, H., *et al*. (2003). Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res*., **31**, 3450–3460.
36. Murray, L.J., Arendall, W.B., 3rd, Richardson, D.C., and Richardson, J.S. (2003). RNA backbone is rotameric. *Proc. Natl. Acad. Sci. U S A*, **100**, 13904–13909.
37. Schneider, B., Moravek, Z., and Berman, H.M. (2004). RNA conformational classes. *Nucleic Acids Res*., **32**, 1666–1677.
38. Hershkovitz, E., Sapiro, G., Tannenbaum, A., and Williams, L.D. (2006). Statistical analysis of RNA backbone. *IEEE/ACM Trans. Comput. Biol. Bioinform*., **3**, 33–46.
39. Murthy, V.L., and Rose, G.D. (2003). RNABase: an annotated database of RNA structures. *Nucleic Acids Res*., **31**, 502–504.
40. Griffiths-Jones, S., *et al*. (2005). Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res*., **33**, D121–D124.
41. Klosterman, P.S., Tamura, M., Holbrook, S.R., and Brenner, S.E. (2002). SCOR: a structural classification of RNA database. *Nucleic Acids Res*., **30**, 392–394.
42. Leontis, N.B., and Westhof, E. (2003). Analysis of RNA motifs. *Curr. Opin. Struct. Biol*., **13**, 300–308.
43. Andreeva, A., *et al*. (2004). SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res*., **32**, D226–D229.
44. Griffiths-Jones, S., *et al*. (2003). Rfam: an RNA family database. *Nucleic Acids Res*., **31**, 439–441.

45. Gardner, PP., *et al*. (2011). Rfam: Wikipedia, clans and the "decimal" release. *Nucleic Acids Res*., **39**, D141–D145.
46. Stefan, LR., *et al*. (2006) MeRNA: a database of metal ion binding sites in RNA structures. *Nucleic Acids Res*., **34**, D131–D134.
47. van Batenburg, F.H., *et al*. (2000). PseudoBase: a database with RNA pseudoknots. *Nucleic Acids Res*., **28**, 201–204.
48. Taufer, M., *et al*. (2009). PseudoBase++: an extension of PseudoBase for easy searching, formatting and visualization of pseudoknots. *Nucleic Acids Res.*, **37**, D127–D135.
49. Bindewald, E., *et al*. (2008). RNAJunction: a database of RNA junctions and kissing loops for three-dimensional structural analysis and nanodesign. *Nucleic Acids Res*., **36**, D392–D397.
50. Andronescu, M., Bereg, V., Hoos, H.H., and Condon, A. (2008). RNA STRAND: the RNA secondary structure and statistical analysis database. *BMC Bioinf*., **9**, 340.
51. Czerwoniec, A., *et al*. (2009). MODOMICS: a database of RNA modification pathways. 2008 update. *Nucleic Acids Res*., **37**, D118–D121.
52. Xin, Y., and Olson, W.K. (2009). BPS: a database of RNA base-pair structures. *Nucleic Acids Res*., **37**, D83–D88.
53. Lu, X.J., and Olson, W.K. (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res*., **31**, 5108–5121.
54. Popenda, M., *et al*. (2010). RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures. *BMC Bioinf*., **11**, 231.
55. Vanegas, P.L., *et al*. (2012). RNA CoSSMos: characterization of secondary structure motifs—a searchable database of secondary structure motifs in RNA three-dimensional structures. *Nucleic Acids Res.*, **40**, D439–D444.
56. Mathews, D.H., Moss, W.N., and Turner, D.H. (2010). Folding and finding RNA secondary structure. *Cold Spring Harb. Perspect. Biol.*, **2**, a003665.
57. Tinoco, I., Jr., and Bustamante, C. (1999). How RNA folds. *J. Mol. Biol.*, **293**, 271–281.
58. Mathews, D.H., and Turner, D.H. (2006). Prediction of RNA secondary structure by free energy minimization. *Curr. Opin. Struct. Biol.*, **16**, 270–278.

59. Mathews, D.H., Sabina, J., Zuker, M., and Turner, D.H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.
60. Reeder, J., *et al*. (2006). Beyond Mfold: recent advances in RNA bioinformatics. *J. Biotechnol.*, **124**, 41–55.
61. Mathews, D.H. (2004). Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA*, **10**, 1178–1190.
62. Wuchty, S., Fontana, W., Hofacker, I.L., and Schuster, P. (1999). Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, **49**, 145–165.
63. Giegerich, R., Voss, B., and Rehmsmeier, M. (2004). Abstract shapes of RNA. *Nucleic Acids Res.*, **32**, 4843–4851.
64. McCaskill, J.S. (1990). The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105–1119.
65. van Batenburg, F.H., Gultyaev, A.P., and Pleij, C.W. (1995). An APL-programmed genetic algorithm for the prediction of RNA secondary structure. *J. Theor. Biol.*, **174**, 269–280.
66. Gultyaev, A.P., van Batenburg, F.H., and Pleij, C.W. (1995). The computer simulation of RNA folding pathways using a genetic algorithm. *J. Mol. Biol.*, **250**, 37–51.
67. Shapiro, B.A., Wu, J.C., Bengali, D., and Potts, M.J. (2001). The massively parallel genetic algorithm for RNA folding: MIMD implementation and population variation. *Bioinformatics*, **17**, 137–148.
68. Lyngso, R.B., and Pedersen, C.N. (2000). RNA pseudoknot prediction in energy-based models. *J. Comput. Biol.*, **7**, 409–427.
69. Cai, L., Malmberg, R.L., and Wu, Y. (2003). Stochastic modeling of RNA pseudoknotted structures: a grammatical approach. *Bioinformatics*, **19**(Suppl 1), i66–73.
70. Rivas, E., and Eddy, S.R. (1999). A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J. Mol. Biol.*, **285**, 2053–2068.
71. Dirks, R.M., and Pierce, N.A. (2003). A partition function algorithm for nucleic acid secondary structure including pseudoknots. *J. Comput. Chem.*, **24**, 1664–1677.
72. Dirks, R.M., and Pierce, N.A. (2004). An algorithm for computing nucleic acid base-pairing probabilities including pseudoknots. *J. Comput. Chem.*, **25**, 1295–1304.

73. Huang, X., and Ali, H. (2007). High sensitivity RNA pseudoknot prediction. *Nucleic Acids Res.*, **35**, 656–663.
74. Gaspin, C., and Westhof, E. (1995). An interactive framework for RNA secondary structure prediction with a dynamical treatment of constraints. *J. Mol. Biol.*, **254**, 163–174.
75. Rousset, F., Pelandakis, M., and Solignac, M. (1991). Evolution of compensatory substitutions through G.U intermediate state in Drosophila rRNA. *Proc. Natl. Acad. Sci. U S A*, **88**, 10032–10036.
76. Kirby, D.A., Muse, S.V., and Stephan, W. (1995). Maintenance of pre-mRNA secondary structure by epistatic selection. *Proc. Natl. Acad. Sci. U S A*, **92**, 9047–9051.
77. Chiu, D.K., and Kolodziejczak, T. (1991). Inferring consensus structure from nucleic acid sequences. Comput. *Appl. Biosci*., **7**, 347–352.
78. Gutell, R.R., *et al*. (1992). Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. *Nucleic Acids Res.*, **20**, 5785–5795.
79. Lindgreen, S., Gardner, P.P., and Krogh, A. (2006). Measuring co-variation in RNA alignments: physical realism improves information measures. *Bioinformatics*, **22**, 2988–2995.
80. Hofacker, I.L., Fekete, M., and Stadler, P.F. (2002). Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.*, **319**, 1059–1066.
81. Knudsen, B., and Hein, J. (2003). Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Res.*, **31**, 3423–3428.
82. Ruan, J., Stormo, G.D., and Zhang, W. (2004). ILM: a web server for predicting RNA secondary structures with pseudoknots. *Nucleic Acids Res.*, **32**, W146–W149.
83. Ruan, J., Stormo, G.D., and Zhang, W. (2004). An iterated loop matching approach to the prediction of RNA secondary structures with pseudoknots. *Bioinformatics*, **20**, 58–66.
84. Havgaard, J.H., Lyngso, R.B., and Gorodkin, J. (2005). The FOLDALIGN web server for pairwise structural RNA alignment and mutual motif search. *Nucleic Acids Res.*, **33**, W650–W653.
85. Torarinsson, E., Havgaard, J.H., and Gorodkin, J. (2007). Multiple structural alignment and clustering of RNA sequences. *Bioinformatics*, **23**, 926–932.

AQ: Journal details added OK

86. Mathews, D.H., and Turner, D.H. (2002). Dynalign: an algorithm for finding the secondary structure common to two RNA sequences. *J. Mol. Biol.*, **317**, 191–203.
87. Perriquet, O., Touzet, H., and Dauchet, M. (2003). Finding the common structure shared by two homologous RNAs. *Bioinformatics*, **19**, 108–116.
88. Touzet, H., and Perriquet, O. (2004). CARNAC: folding families of related RNAs. *Nucleic Acids Res.*, **32**, W142–W145.
89. Hochsmann, M., Toller, T., Giegerich, R., and Kurtz, S. (2003). Local similarity in RNA secondary structures. *Proc. IEEE Comput. Soc. Bioinf. Conf.*, **2**, 159–168.
90. Siebert, S., and Backofen, R. (2005). MARNA: multiple alignment and consensus structure prediction of RNAs based on sequence structure comparisons. *Bioinformatics*, **21**, 3352–3359.
91. Duarte, C.M., Wadley, L.M., and Pyle, A.M. (2003). RNA structure comparison, motif search and discovery using a reduced representation of RNA conformational space. *Nucleic Acids Res.*, **31**, 4755–4761.
92. Duarte, C.M., and Pyle, A.M. (1998). Stepping through an RNA structure: A novel approach to conformational analysis. *J. Mol. Biol.*, **284**, 1465–1478.
93. Harrison, A., *et al*. (2003). Recognizing the fold of a protein structure. *Bioinformatics*, **19**, 1748–1759.
94. Dror, O., Nussinov, R., and Wolfson, H. (2005). ARTS: alignment of RNA tertiary structures. *Bioinformatics*, **21**(Suppl 2), ii47–ii53.
95. Ferre, F., Ponty, Y., Lorenz, W.A., and Clote, P. (2007). DIAL: a web server for the pairwise alignment of two RNA three-dimensional structures using nucleotide, dihedral angle and base-pairing similarities. *Nucleic Acids Res.*, **35**, W659–W668.
96. Capriotti, E., and Marti-Renom, M.A. (2008). RNA structure alignment by a unit-vector approach. *Bioinformatics*, **24**, i112–118.
97. Capriotti, E., and Marti-Renom, M.A. (2009). SARA: a server for function annotation of RNA structures. *Nucleic Acids Res.*, **37**, W260–W265.
98. Chang, Y.F., Huang, Y.L., and Lu, C.L. (2008). SARSA: a web tool for structural alignment of RNA using a structural alphabet. *Nucleic Acids Res.*, **36**, W19–W24.
99. Sarver, M., *et al*. (2008). FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *J. Math. Biol*., **56**, 215–252.

100. Petrov, A.I., Zirbel, C.L., and Leontis, N.B. (2011). WebFR3D—a server for finding, aligning and analyzing recurrent RNA 3D motifs. *Nucleic Acids Res.*, **39**, W50–W55.
101. Lai, C.E., *et al*. (2009). FASTR3D: a fast and accurate search tool for similar RNA 3D structures. *Nucleic Acids Res.*, **37**, W287–W295.
102. Zhong, C., Tang, H., and Zhang, S. (2010). RNAMotifScan: automatic identification of RNA structural motifs using secondary structural alignment. *Nucleic Acids Res.*, **38**, e176.
103. Kirillova, S., Tosatto, S.C., and Carugo, O. (2010). FRASS: the web-server for RNA structural comparison. *BMC Bioinf.*, **11**, 327.
104. Rahrig, R.R., Leontis, N.B., and Zirbel, C.L. (2010). R3D Align: global pairwise alignment of RNA 3D structures using local superpositions. *Bioinformatics*, **26**, 2689–2697.
105. Shapiro, B.A., Yingling, Y.G., Kasprzak, W., and Bindewald, E. (2007). Bridging the gap in RNA structure prediction. *Curr. Opin. Struct. Biol.*, **17**, 157–165.
106. Massire, C., and Westhof, E. (1998). MANIP: an interactive tool for modelling RNA. *J. Mol. Graph Model*, **16**, 197–205, 255–197.
107. Major, F. (2003). Building three-dimensional ribonucleic acid structures. *Comput. Sci. Eng.*, **5**, 44–53.
108. Yingling, Y.G., and Shapiro, B.A. (2006). The prediction of the wild-type telomerase RNA pseudoknot structure and the pivotal role of the bulge in its formation. *J. Mol. Graph Model*, **25**, 261–274.
109. Sharma, S., Ding, F., and Dokholyan, N.V. (2008). iFoldRNA: three-dimensional RNA structure prediction and folding. *Bioinformatics*, **24**, 1951–1952.
110. Frellsen, J., *et al*. (2009). A probabilistic model of RNA conformational space. *PLoS Comput. Biol.*, **5**, e1000406.
111. Das, R., and Baker, D. (2007). Automated de novo prediction of native-like RNA tertiary structures. *Proc. Natl. Acad. Sci. U S A*, **104**, 14664–14669.
112. Das, R., Karanicolas, J., and Baker, D. (2010). Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods*, **7**, 291–294.
113. Bradley, P., *et al*. (2005). Free modeling with Rosetta in CASP6. *Proteins*, **61**(Suppl 7), 128–134.
114. Poolsap, U., Kato, Y., and Akutsu, T. (2009). Prediction of RNA secondary structure with pseudoknots using integer programming. *BMC Bioinf.*, **10**(Suppl 1), S38.

115. Sato, K., Kato, Y., Hamada, M., Akutsu, T., and Asai, K. (2011). IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming. *Bioinformatics*, **27**, i85–93.
116. Flores, S.C., Wan, Y., Russell, R., and Altman, R.B. (2010). Predicting RNA structure by multiple template homology modeling. *Pac. Symp. Biocomput.*, 216–227.
117. Jossinet, F., Ludwig, T.E., and Westhof, E. (2010). Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics*, **26**, 2057–2059.
118. Rother, M., Rother, K., Puton, T., and Bujnicki, J.M. (2011). ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res.*, **39**, 4007–4022.
119. Cao, S., and Chen, S.J. (2011). Physics-based de novo prediction of RNA 3D structures. *Journal of Physical Chemistry B*, **115**, 4216–4226.
120. Izzo, J.A., Kim, N., Elmetwaly, S., and Schlick, T. (2011). RAG: an update to the RNA-As-Graphs resource. *BMC Bioinf.*, **12**, 219.
121. Kim, N., *et al*. (2007). RagPools: RNA-As-Graph-Pools–a web server for assisting the design of structured RNA pools for in vitro selection. *Bioinformatics*, **23**, 2959–2960.
122. Bida, J.P., and Maher, L.J., 3rd (2012). Improved prediction of RNA tertiary structure with insights into native state dynamics. *RNA*, **18**, 385–393.
123. Capriotti, E., Norambuena, T., Marti-Renom, M.A., and Melo, F. (2011). All-atom knowledge-based potential for RNA structure prediction and assessment. *Bioinformatics*, **27**, 1086–1093.
124. Case, D.A., *et al*. (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
125. Bernauer, J., Huang, X., Sim, A.Y., and Levitt, M. (2011). Fully differentiable coarse-grained and all-atom knowledge-based potentials for RNA structure evaluation. *RNA*, **17**, 1066–1075.
126. Capriotti, E,, and Marti-Renom, M.A. (2008). Computational RNA structure prediction. *Curr. Bioinf.*, **3**, 32–45.