RNA Structure Prediction

10

Manoj Kumar Gupta, Gayatri Gouda, Ravindra Donde, Piyali Goswami, N. Rajesh, Pallabi Pati, Sushil Kumar Rathore, Ramakrishna Vadde, and Lambodar Behera

Abstract

One of the significant forms of molecules present in living cells is ribonucleic acid (RNA). RNA structural elements moderate various biological process, including epigenetic function, modify mRNA stability, and alternate splicing. The study of the secondary structures of RNA is, therefore, crucial for interpreting the role as well as the regulatory mechanism of RNA transcripts. But experimental methods are tedious, time-consuming, pricey, requires special equipment, and, thus, cannot often be implemented. Methods for statistical simulation are an option and parallel to experimental approaches. Additionally, the findings from the RNA-Puzzles, joint research on the estimation of RNA structures, suggest that computational methods can be employed for effective RNA modeling. However, there is still space for improvement. Considering this, in the chapter, authors attempted to understand the various forms of RNA and how computational approaches can be employed to predict their structure more precisely. The RNA

M. K. Gupta · G. Gouda · R. Donde · L. Behera (✉)
Crop Improvement Division, ICAR-National Rice Research Institute, Cuttack, Odisha, India

P. Goswami
Department of Biotechnology, IIT Kharagpur, Kharagpur, West Bengal, India

N. Rajesh · R. Vadde
Department of Biotechnology and Bioinformatics, Yogi Vemana University, Kadapa, Andhra Pradesh, India

P. Pati
District Headquarter Hospital, Ganjam, Odisha, India

S. K. Rathore
Department of Zoology, Khallikote Autonomous College, Ganjam, Odisha, India

M. K. Gupta, L. Behera (eds.), *Bioinformatics in Rice Research*,
https://doi.org/10.1007/978-981-16-3993-7_10

is classified mainly according to its existence, role, and structure into three groups, messenger RNA, transfer RNA, and ribosomal RNA. To date, numerous algorithms, and tools, have been designed for predicting the secondary structure of RNA. However, since three-dimensional structures are highly required for getting insight into the function of the RNA, few approaches have also been developed for predicting tertiary structures of RNA atoms. However, the authors believe that, in the near future, by combining experimental and computational approaches, we will be able to predict the structure of RNA more accurately, which in turn will enable us to understand its structure and function more precisely.

Keywords

RNA · tRNA · mRNA · rRNA · Structure prediction

Abbreviations

CSA	Comparative sequence analysis
DP	Dynamic programming
dsRNAs	Double-strand RNAs
lncRNAs	Long noncoding RNA
miRNAs	microRNA
MMP	MC-fold/Mc-Sym Pipeline
mRNA	Messenger RNA
PPV	Positive predictive value
pre-rRNA	Precursors-rRNAs
rDNA	Ribosomal DNA
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
rRNA	Ribosomal RNA
siRNA	Short interfering RNA
snoRNAs	Small nucleolar RNAs
SSs	Secondary structures
t6A	Threonyl-carbamoyl adenosine
tRNA	Transfer RNA

10.1 Introduction

One of the significant forms of molecules present in living cells is ribonucleic acid (RNA) [1]. After the central dogma was hypothesized in 1950, the key function assigned to RNA was to serve as the intermediary between DNA and protein synthesis [2]. However, out of ~70% of the genome transcribed, only a limited

portion encodes for protein sequences [3], which means that most RNAs might have various biological functions. Earlier, several researchers have also suggested that RNAs are transporters for genetic material and are also associated with many biological processes that are incredibly significant [4]. For instance, RNA transcripts fold into structures (SSs) (Fig. 10.1), which have different catalytic, ligand, and scaffolding functions that shape a crucial biological regulatory activity. RNA structural elements moderate epigenetic function, modify mRNA stability and translation, scaffold large macromolecular complexes, transduce signals, and monitor alternate splicing. The study of the SSs of RNA is, therefore, crucial for interpreting the role as well as the regulatory mechanism of RNA transcripts [2].

RNA folds into a 3D system through hydrogen interaction and base-stacking, which in the sequence are not consecutive [7]. The 3D structure of the RNA molecule decides its function, like proteins. In order to construct a 3D model, high-resolution experimental methods such as crystallography [8, 9], cryo-EM [10], and nuclear magnet resonance spectroscopy may be taken advantage of [11]. But experimental methods are tedious, time-consuming, pricey, requires special equipment, and, thus, cannot often be implemented. Methods for statistical simulation are an option and parallel to experimental approaches. Additionally, the findings from the RNA-Puzzles [12], joint research on the estimation of RNA structures, suggest that computational methods can be employed for effective RNA modeling. However, there is still space for improvement.

Like proteins, RNAs can be divided into families [13], which originated from a common ancestor. RNA sequences from the same family will have higher similarity, and the study of sequence conservation may be used for identifying important conserved areas, such as areas binding ligands, active sites, or other important functions. The Watson crick basis pairing pattern for the RNA is often used to forecast SSs. According to the CompaRNA [14], RNA alignments methods such as PETfold [15] outweigh the predictive single sequence methods of the secondary RNA structure. RNA alignments may also be used to enhance the prediction of the tertiary structure [16]. For instance, recently, a group of researchers employed a novel approach for exploring tertiary structure predictions [13]. The methodology examines the usage of multiple alignment knowledge and simultaneous RNA homolog simulation to strengthen ab initio RNA structure modeling techniques. A new technique, called EvoClustRNA, is focused on a conventional strategy to predict RNA structures, utilizing evolutionary knowledge from distant sequence homologs [17]. On the basis of the empirical finding that RNA sequences of the same RNA family normally fold into identical 3D structures, they have checked whether computational modeling can be driven by searching for a global helical sequence for the target sequence, which is shared through de novo models of various sequence homologs. EvoClustRNA is the first effort to use this method for RNA 3D prediction. Thus, in this chapter, the authors attempt to understand the sources, form, and role of the RNA structure and how different computational approaches that researchers are adopting for determining various RNA structures.

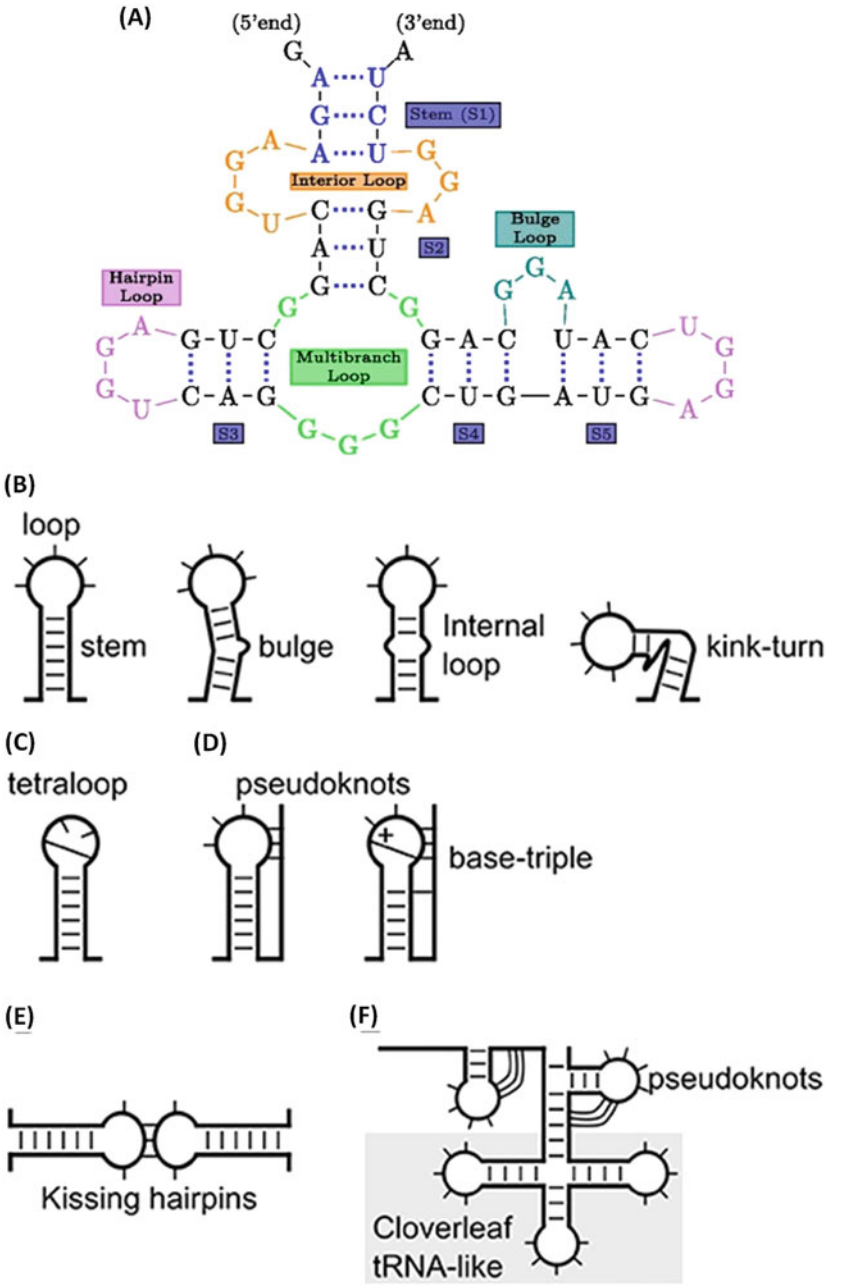

Fig. 10.1 (**a**) Basic structural motifs depicted within RNA secondary structures. (**b**) The simplest form of RNA structure is a stem-loop. A stem-loop is shown with a bulge, internal loop, or (**c**) tetraloop. (**d**) The loop can also base-pair with upstream or downstream sequences to form a pseudoknot. (**e**) Interaction between the loops of two stem-loops forms kissing hairpins. (**f**) A relatively complex structure is a cloverleaf or tRNA-like structure that often consists of multiple stem-loops and pseudoknots. (Adapted from [5, 6])

10.2 RNA Structure

The RNA is classified mainly according to its existence, role, and structure into three groups: messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA). The rRNA produces complex three-dimensional structures that interact with polypeptides to shape ribosomes responsible for protein synthesis in organelles. The ribosomes act as an mRNA encoding tool. The mRNA includes instructions that dictate protein amino acid sequences. The tRNA serves as an adapter to convert mRNA codons into those amino acids [18]. In addition, as discussed below, there are also other forms of RNAs, like long noncoding RNA (lncRNAs), small nucleolar RNAs (snoRNAs), microRNA(miRNAs), and short interfering RNA (siRNA) [19].

10.2.1 Messenger RNA

The "messenger" RNA is mRNA. The mRNA in the nucleus is synthesized using the DNA nuclear sequence as a reference. This process needs nucleotide triphosphates as substrate and is catalyzed by the RNA polymerase II enzyme. The DNA to mRNA processing is called transcription and takes place in the nucleus. The mRNA produced in the nucleus is transferred to ribosomes out of the nucleus and into a cytoplasm. Subsequently, the mRNA guides the protein synthesis that takes place in the cytoplasm. On the ribosomes, proteins are packaged using the mRNA sequence as a reference. Thus, mRNA bears a "message" to the cytoplasm from the nucleus for encoding protein. The processing of mRNA to proteins is called translation [20]. Earlier studies have reported that while the configuration and mode of action of the prokaryotic and eukaryotic mRNAs vary, similarities still exist. In mRNA, genetic information is encoded into a four-base nucleotide alphabet, which forms codons of three bases. Each codon codes for a certain amino acid except for stop codons that specify when the synthesis of protein stops. The mRNA is translated by the codon-reading ribosome. For all prokaryotes and eukaryotes, the beginning or initiator codon is an AUG sequence, and the sequences are read in 5′ to 3′ direction. Eukaryotic mRNA normally codes for one specific (monocistronic) protein, whereas the prokaryotic mRNA typically codes for a set of similar (polycistronic) proteins on the same mRNA. Polycistronic mRNA guides the synthesis of each coded polypeptide, which is more or less simultaneous. For example, the *trp* operon is a DNA, which is transcribed in mRNA and codes for six polypeptides, catalyzes the synthesis of tryptophan.

The mRNA has a shorter life in comparison to the DNA. An mRNA molecule may be stored, edited, and transported before translation following transcription [21]. For many factors, mRNA stabilization is an important control point in the regulation of gene expression. At first, an equilibrium between its synthesis and degradation is highly required for consistent normal function. Secondly, the consistency of individual mRNAs may be altered because of multiple environmental stimuli, such as carbon source, viral diseases, and developmental transformations, allowing rapid shifts in the expression of the gene. Thirdly, a method of competent

mRNA degradation to remove deleterious errors during mRNA synthesis. Finally, successful mRNA degradation is essential for the growth of both the prokaryotes and the eukaryotes [22].

10.2.2 Ribosomal RNA

The biosynthesis of rRNA and its integration into the ribosomes is a surprisingly complex process, which for over three decades, has been the focus of intensive study. Ribosome biogenesis starts in the nucleolus, in accordance with the "RNA-base machine," through the synthesis of the large primary mRNA transcripts via the RNA polymerase I (Pol I) [23]. In eukaryotes, the mature 80S cytoplasmic ribosome is composed of the 60S larger subunit and the 30S smaller subunit. The small subunit is composed of 18S rRNAs and more than 30 ribosomal proteins. The large subunit comprises 5.8S, 25S/28S, and 5S rRNAs and over 40 ribosomal proteins. Biogenesis of ribosome includes replication of ribosomal DNA (rDNA), production of precursors-rRNAs (pre-rRNA), modifications to the RNA, and assembly of ribosomal protein and assembly factors in rRNA. Ribosome biogenesis is an important, complicated, and energy-intensive mechanism strictly controlled by endogenous signals and environmental factors, such as ambient temperature. Within eukaryotic cells, irregular biogenesis of rRNA stimulates "RNA Nucleus Quality Regulation," inducing higher polyadenylation of some intermediate rRNA products as well as by-products, known as TRAMPs (Trf/Air/Mtr4 polyadenylation complex). The nuclear exosome complex sequentially degrades these intermediates. Ribosomal biogenesis failure results in significant developmental of deficiencies in higher plants and extreme hereditary disorders in mammals [24].

The catalytic function of rRNA was first shown by Harry Noller and his colleagues' 1992 experiments. These researchers found that even after about 95% of the ribosomal proteins have been discarded via traditional protein extraction methods, the large ribosomal unit would catalyze peptide bond formation ("Peptidyl Transferase Reaction"). In comparison, RNase treatment fully abolishes the development of peptide bindings, which clearly supports the theory that peptide binding development is an "RNA-catalyzed reaction." Further experiments have also validated as well as expanded these findings by showing that the "peptidyl transferase reaction" can be catalyzed by synthetic fragments of 23S rRNA in the complete absence of any ribosomal protein. These findings support that rRNA catalyzes the basic reaction in protein synthesis.

Apart from being the ribosomes' basic catalytic constituents, ribosomal proteins can also be used for promoting proper rRNA folding and for boosting the ribosomes' activity through proper tRNAs' positioning [25]. The direct presence of rRNA during the peptidyl transmission response has significant evolutionary consequences. RNAs are considered to be the first macromolecules that have self-replicating properties. Earlier studies have also supported this theory by stating that ribozymes like RNase P as well as self-splicing introns can catalyze RNA substrate reactions. The rRNA's function in the peptide attachment formation expands the

catalytic action of RNA to direct participation in the synthesis of protein. Few studies have also revealed that the "Tetrahymena rRNA ribozyme" can catalyze the amino acid binding of RNA, thus adding credence to the likelihood of RNAs, rather than protein, being the initial aminoacyl tRNA synthesis. Thus, the RNA molecules may also serve as a significant biomarker toward understanding the early evolution of cells in catalyzing the reactions needed for self-replication as well as for protein synthesis [25].

10.2.3 transfer RNA

tRNA is a small nucleotide chain. The tRNA acts as an "adapter" molecule with an L-shape configuration, which converts the three-nucleotide codon sequence of the mRNA into the required amino acid of that codon. The tRNAs define the genetic code as the bond between amino acids and nucleic acids. However, their functions extended beyond protein translation, providing a remarkable set of tasks in the synthesis of bacterial cell walls, viral replication, cell tension, and even regulation of animal behavior [26]. Bacteria have multiple antibiotic mechanisms, which in the clinic is a growing obstacle. Within bacterial outer membrane lipids, tRNA-dependent aminoacylation offers improved virulence and tolerance to the cationic antimicrobial peptide [27]. Earlier, Fields and his team have studied the well-documented pathways of lipid aminoacylation to illustrate the usage of aminoacyl-tRNA substrates as an amino acid donor in lipid changes for improved antibiotic tolerance by the aminoacyl-phosphatidylglycerol synthases [28]. Emerging data also suggest that the tRNA genes perform a new function in bacterial conjugation. For instance, Alamos and his team found that 36 out of its 95 tRNAs are encoded in an integrative-conjugative genetic variable within *acidithiobacillus ferrooxidans* [29]. Castillo and his team have also shown that the integrases encoded inside the conjugative factor recognize the area of the tRNA stem-loop for active and location-specific recombination [30].

Mature tRNAs are abundant in nucleotide-based post-transcriptional modifications. These improvements perform important roles in the management of translation and reading frames [31], tRNA reliability, and transport [32]. Modification may occur within the anticodon as well. Phylogenetic analysis has been performed by Rafels-Ybern and his team to show the production of adenine base modification to Inosin (I) at location 34. The A to I shift affects the tRNA's ability to decipher the codon's third nucleotide location. Whereas A34 forms an optimal relationship with U, I34 is equally well informed of U, C, or A. The I34 function is to broaden wobble decoding, as investigated by another group of researchers [33]. The switch to I34 requires one tRNA to read three codons of the same amino acid. While the alteration is widely used in eukaryotes, there were limited earlier examples in bacteria. Earlier, researchers have also found many possible I34 locations for tRNAs' modifications in Firmicutes as well as Cyanobacteria genomes [34].

Changes can often involve the shipment of tRNA within cellular compartments. In an interesting study, Kessler et al. explain how the transportation of tRNAs between various areas of the cell influences the maturation and alteration of tRNAs [35]. Seminary tests of the sleeping causative agent, *Trypanosoma brucei*, indicate that tRNA-Tyr is transferred to the cytoplasm where the intron in an immature tRNA is broken. The tRNA is then reimported to the nuclease in a process called "retrograde transport," in which the spliced tRNA is necessary for modification with queuosin [32]. Some tRNA variations are retained in all life types. In a few tRNAs, the modification threonyl-carbamoyl adenosine (t6A) is found to decipher the ANN codons and is necessary for the stabilization of the duplex codon–anticodon. The t6A modification at loci 37, next to the anticodon loop, has been found to be necessary for the operation of *Streptococcus mutans*' anticodon nuclease PrrC [31]. The nuclease facilitates the bacterial cell death under stress conditions or during phage infection when the $tRNA^{Lys}_{UUU}$ anticodon loop is precisely cleaved.

10.2.4 Small Nucleolar RNAs

snoRNAs are generally composed of 60–170 nuclear nucleotides (with few exceptions) [36, 37] and are mainly involved in directing post-transcriptional alteration of nonprotein-coding RNAs (rRNAs, snRNAs) [38]. snoRNAs are broadly categorized as either a "C/D box" or "H/ACA box" based on the given sequence as well as SSs component [37]. "C/D box" directs 2′-O-methylation and "H/ACA" nuclear pseudouridylation upon target molecules. Since the 5′ as well as 3′ ends of the molecule fold into a stem configuration, which in turn creates a "kink switch," the "C box" ("RUGAUGA", R = A or G) and the "D box" ("CUGA") sequence motifs of the "C/D box" are brought closely into contact. The majority of the C/D boxes have another less conserved C as well as D box motifs, namely the C′ and D′ boxes, within the "Central SnoRNA region." C/D box is mainly involved in the ribonucleoprotein (RNP) complexes that also include 15.5 K, NOP56, NOP58, and fibrillarine proteins [39, 40]. The latter catalyzes the 2′-O-methylation of ribose molecules within the target RNA [40]. "H/ACA box" is a well-designed SSs comprising two hairpins connected together through a single-stranded area designated as the "H box" ("ANANNA", N = A, C, G or U) as well as the "ACA box" ("AYA", Y = C or U) at the 3′ end [41]. "H/ACA" produces "H/ACA" snoRNA and a group of four proteins, namely, Nop10, Gar1, Dyskerin, and Nhp2, where Dyskerin functions as pseudouridine synthase [42]. Primary identification of "H/ACA box" often includes RNA–RNA interactions between single-stranded area within the inner loops of the two snoRNA hairpin systems, mostly with target RNA [43, 44].

10.2.5 microRNA

miRNAs are small noncoding RNAs that have a mean length of ~22 nucleotides. The majority of the miRNAs are transcribed into prime miRNAs from DNA sequences and converted into precursor miRNAs as well as mature miRNAs. In certain instances, miRNA interacts with the 3′ UTR of the objective mRNA for suppressing expression. However, there have also been records of the association of miRNAs with other regions, including the coding sequence, 5′ UTR, and gene promoters. Moreover, under some circumstances, miRNAs have been shown to cause gene expression. In recent research, miRNAs have been shuttled between various subcellular cells to regulate the rate of translation and also transcription [45]. miRNAs are critical for the natural growth of animals and active in various biological processes [46]. Aberrant expression of miRNAs is related to a variety of human diseases [47, 48]. miRNAs are often secreted within extracellular fluids. Extracellular miRNAs can serve as plausible biomarkers for a number of diseases and signaling molecules for cell–cell interaction [49].

10.2.6 Short Interfering RNA

siRNAs are derived from double-strand RNAs (dsRNAs), consisting of two antisenses as well as a sense RNA strand that forms 19–25 bp duplex with 3′ dinucleotide overhangs. The antisense strand is a perfect reverse complement to the expected mRNA target. Few important functions of siRNAs include mainly post-transcriptional gene silencing or translation inhibition, exogenous DNA defense, intervention in epigenetic processes, and preserving genome integrity by transcriptional silencing. It has been used for industrial purposes to easily research in vivo gene expression owing to its capacity to knock out genes. Many of the siRNA measurement applications are therefore planned to aim siRNA sequences optimally to knock out genes. Subsequently, siRNAs' prediction can also be used to establish protocols for screening and can be used to classify new pathways to confirm cellular targets correlated with diseases such as hepatitis, cancer, and HIV infection [50].

10.2.7 Long Noncoding RNA

lncRNAs are classified as >200 nucleotide RNA molecules. While this differentiation is rather subjective and dependent on functional aspects of RNA separation techniques, lncRNAs vary from miRNAs as well as other sRNAs. In significant amounts, lncRNAs are present within the genome. They may not usually have working open read frames (ORFs). However, the discovery of bifunctional RNAs with coding-independent and protein-coding functions is flexible by this distinction, which increases the probability that certain protein-coding genes might have non-coding functions, as well [51]. Many lncRNAs are poorly expressed and, thus, researchers experience difficulties during exploring lncRNAs and understanding

why lncRNAs were always considered to be "transcriptional noise." RNA-sequences in various tetrapods indicate that mostly (~81%), primate-specific lncRNAs are poorly retained in the DNA chain. However, it is worth remembering that many lncRNAs are highly conserved within the DNA sequence, and ~3% of lncRNAs might have originated earlier than 300 million years ago [52].

lncRNAs may be fast-evolving species of RNA that can play key roles in lineage specifying. A comparison of the matching tissues in *Rattus norvegicus*, *Mus musculus castaneus*, and *Mus musculus domesticus* indicates that shifts in the transcription levels of the adjacent protein-coding genes are linked with the appearance or disappearance of the lncRNAs [53]. There are several instances of lncRNAs with retained biological roles but low-level sequence survival, such as TUNA/megamind correlated with the growth of the brain in zebrafish, mouse, and humans [54, 55], and X-inelective unique transcript (*Xist*) involved in X-inactivation [56]. RNA molecules can require fewer sequence retention in order to maintain their function than proteins. Conversely, lncRNA promoters have a strong sequence conservation, which is even higher in comparison to protein-coding-gene promoters [57], indicating that lncRNA expression control is significant.

10.3 RNA Structure Prediction

RNA plays various cellular functions, and, thus, recognizing RNA structure is essential to understand its action mechanism [58]. Because the prediction of the three-dimensional RNA structure is difficult and expensive, scientists mainly depend on RNA's SSs. Hence, to date, numerous algorithms have been designed for predicting the SSs of RNA [19]. However, since three-dimensional structures are highly required for getting insight into the function of the RNA, few approaches have also been developed for predicting tertiary structures of RNA atoms [59] (Fig. 10.2 and Tables 10.1 and 10.2).

10.3.1 RNA SSs Prediction Methods

Present methods of prediction of SSs of RNA may be broadly categorized into comparative sequence analysis (CSA) and folding algorithms with thermodynamic, predictive, or probabilistic scoring schemes [81]. CSA distinguishes base pairs between homologous sequences. These approaches are incredibly accurate [82] if there are a sufficient number of compatible sequences and are aligned with professional expertise manually. However, to date, only a few thousand RNA families have been identified. Therefore, the most popular method used for the RNA SSs prediction is to fold an individual RNA sequence according to a suitable scoring feature. In this method, the RNA structure is separated into substructures, such as loops and trunks in the closest model [83]. Dynamic programming (DP) algorithms are then used to find minimal or probabilistic global structures from such substructures. Subsequently, experimental technique [84] (e.g., RNAshapes [85],

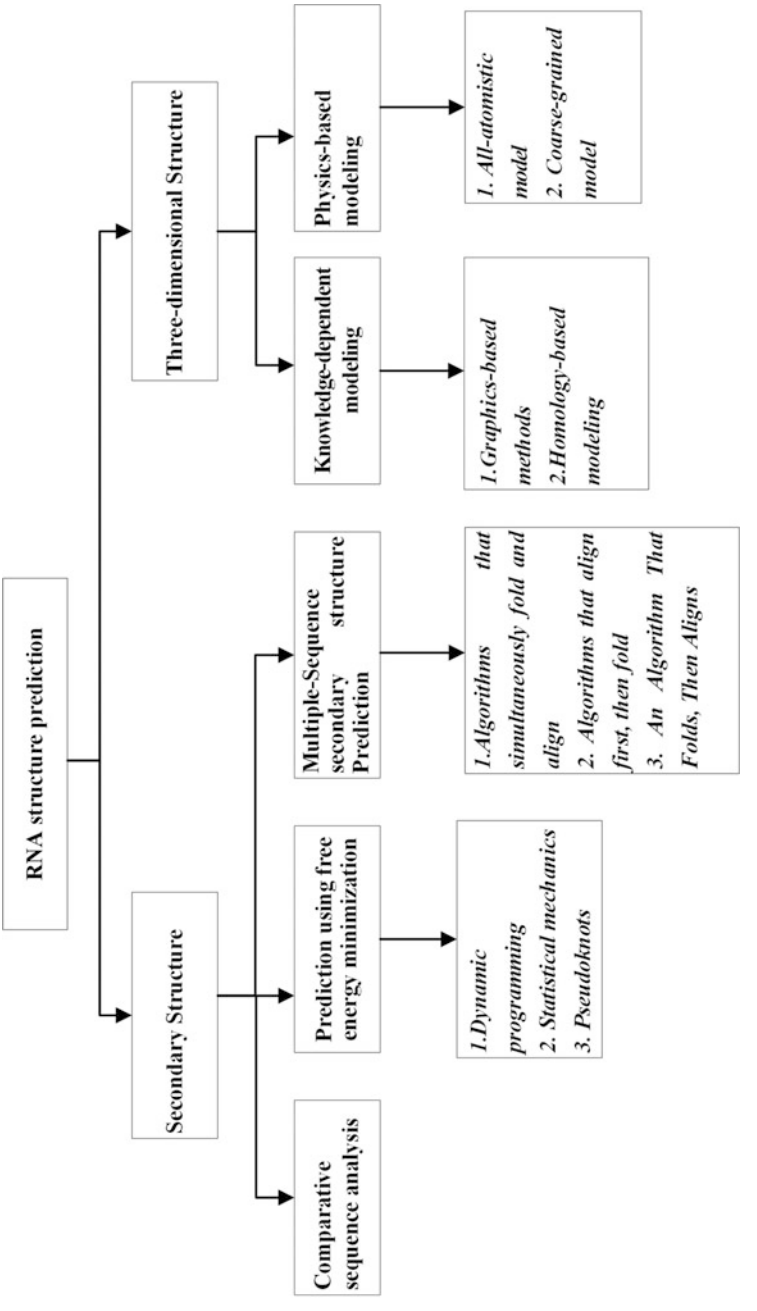

Fig. 10.2 Different approaches to predict secondary and three-dimensional structure of RNA

Table 10.1 Softwares and tools for predicting secondary structure of RNA

Name	Description	References
CentroidAlifold	Employs generalized centroid estimator	[60]
DAFS	Align and fold RNA sequences through dual decomposition.	[61]
MASTR	Uses Markov chain Monte Carlo in a simulated annealing framework	[62]
Multilign	Utilizes multiple Dynalign calculations for finding a low free energy structure that is common to numerous sequences. It does not need sequence identity.	[63]
Murlet	Uses iterative alignment dependent on Sankoff's algorithm having sharply decreased computational time as well as memory.	[64]
MXSCARNA	Employs progressive alignment	[65]
PARTS	Probabilistic model and requires pseudo free energies	[66]
Pfold	Utlizes a SCFG trained on rRNA alignments.	[67]
PETfold	Combines both the energy-based and evolution-based approaches	[15]
PhyloQFold	Consider the evolutionary history of a group of aligned RNA sequences	[68]
TurboFold	Utilizes probabilistic alignment as well as partition functions for mapping conserved pairs among sequences, and subsequently iterates the partition functions for improving the accuracy of structure prediction	[69]
Context Fold	Dependent on feature-rich trained scoring models.	https://www.cs.bgu.ac.il/~negevcb/contextfold/
E2Efold	A deep learning approach that employs a constrained optimization solver, without using dynamic programming.	[70]
SwiSpot	Detects alternative (secondary) configurations of riboswitches	[71]
Mfold	Prediction based on minimum free energy	[72]

RNAstructure [86], and RNAfold [87]) or machine learnings approaches (e.g., ContextFold [88] and CentroidFold [89]) are required to calculate the score parameters of every substructural device. But total accuracy (the percentage of correctly predicted basic pairs in all predicted base pairs) seems to have hit a "efficiency ceiling" [81] at around 80% [90, 91]. This is because all current approaches do not recognize some of all the base pairs arising from tertiary interactions [92]. These base-pairs are mostly pseudo-knotted (non-nested), lone (unstacked), and noncanonical base pairs (not G-U, A-U, and G-C) and triple interactions [92, 93]. While some methods can predict secondary RNA structures with pseudoknots (e.g., Knotty [94] and Probknot [95], pknotsRG [96]) and others

Table 10.2 Softwares and tools for predicting three-dimensional structure of RNA

Name	Description	References
BARNACLE	Employs probabilistic approach	[73]
FARNA	de novo prediction.	[74]
iFoldRNA	3D structure prediction as well as folding	[75]
MC-Fold MC-Sym Pipeline	Thermodynamics as well as nucleotide cyclic motifs for RNA structure prediction algorithm	[76]
ModeRNA	Based on a template RNA structure as well as a user-defined target-template sequence alignment	[77]
NAST	Coarse-grained modeling having knowledge-based potentials as well as structural filters	[78]
MMB	Turning limited experimental information into 3D models of RNA	[79]
RNA123	de novo and homology modeling of RNA 3D structures.	[80]
RNAComposer	Automated generation of large RNA 3D structures.	http://rnacomposer.cs.put.poznan.pl/

may predict noncanonical base pairs (e.g., CycleFold [97], MC-Fold-DP [98], and MC-Fold [76]).

10.3.1.1 Comparative Sequence Analysis

The most reliable way of the prediction of SSs for RNA is CSA as well as it is the method of first preference when deciding the SSs of a new RNA. It is built on the hypothesis that structure is more commonly conserved than sequences via evolution [58]. CSA was first used to address tRNA SSs [99, 100]. This research is used to be done with the assistance of modern structure prediction algorithms. Subsequently, tRNA crystal structures predicted was found to be right. Later, a CSA of the SSs of rRNAs have revealed that over 88% of the expected pairs find crystal structures subsequently fixed, and nearly all of the expected tertiary as well as noncanonical interactions were considered to be right [82]. Almost no means of SSs prediction can give something similar to this degree of precision, especially for longer RNAs, or have a similar insight into higher-order contacts that might also have functional or structural worth. CSA is typically the criterion under which structure prediction algorithms are tested since only a small number of sequences of such RNA families have been crystallized [58].

Identifying regions with orchestrated mutations that do not represent nucleotide identities but retain base pairs is a good indicator of an underlying structure that is conserved and of practical significance. The two-nucleotide sequence modifications that maintain base pairing are considered compensating base-pair changes. For example, the G-C base pair is more likely to mutate in one sequence into another canonical pair (AU, UA, CG, UG, GU) as the modification may include modifying two nucleotides rather than one, than mutating into one noncanonical pair or deleting one of the pairing partners with a single nucleotide. The bases of homologous RNAs

from distant species can have low identification, but SSs are fully conserved, as each transition under which the sequences diverge conserves the structure [58]. Further accessible descriptions of structures of variance sequences resolved through CSA are accessible in the Rfam database seed alignments [101]. Irrespective of all these signs, CSA is not necessarily feasible, especially when a sequence is not defined. Free energy minimization (FEM) is one of the most common approaches in such scenarios [58].

10.3.1.2 Secondary RNA Structure Prediction Using Free Energy Minimization

The most common approach for predicting SSs is the FEM, where only a single sequence is defined for a certain function [102, 103]. This approach mainly employs DP, statistical mechanics, and pseudoknots algorithms to achieve its aim.

Dynamic Programming

The most common methods of FEM for RNA prediction are focused on dynamic algorithms of programming [102–104]. In principle, these algorithms can take into account indirectly all potential SSs for a particular sequence with the explicit construction of these structures. To do this, the lowest folding free energies are calculated for all sequence fragments of the entire sequence and the outcomes retained. As the least folding-FE (FE) for longer fragments is measured, the mechanism speeds up to the free energies for shorter fragments. DP algorithms have been preferred because they are computationally powerful and usually produce the same results to ensure that the lower FE structure is provided with the stability laws.

Statistical Mechanics

The lowest FE configuration is the most possible configuration for RNA in equilibrium. When the expected lowest FE structure is compared to the well-known secondary sequence, the precision may be defined by sensitivity or positive predictive value (PPV). Sensitivity is the proportion of recognized base pairs in the SSs predicted. The PPV is the proportion of expected pairs in the established structure. Therefore, sensitivity states that the proportion of identified pairs can be estimated independently of erroneously estimated pairs. Good predictive value is the percentage of expected, accurate pairs influenced by inaccurate pair predictions. It is usually less than sensitivity since FE reduction expects more base pairs than the so-called base changes. It is usually less than sensitivity since FE reduction expects more base pairs than the so-called base compensating base changes.

In 2003, Ding and Lawrence [105] established a statistical sampling procedure for RNA SSs prediction. The SSs are sampled according to Boltzmann's likelihood by means of a partition-function approximation using a stochastic DP algorithm. The likelihood of any given base pair is the frequency of its existence in the ensemble of structures within sampled structures. Moreover, several new structural properties can be calculated, including the likelihood of the single-stranded of two neighboring nucleotides. This information is not given by the partition function estimation alone in a single estimation since the base pairs' pairing chances are not independent. In

other words, the possibility that two base pairs will appear in the same structure is not the consequence of its partition function. The predictive sampling approach improves the estimation of the SSs by identifying the SSs in the ensemble that better describes all of the structures [106]. This "centroid" configuration is selected as the least aggregate variance of all systems. The centroid of the ensemble is also not the lowest FE system. On average, centroids have slightly better sensitivity to base-pair predictions for different sequence databases of the established SSs but have a substantially higher predictive value. Thus, statistical sampling should also be used to boost the precision of SSs prediction.

Pseudoknots

Pseudoknots are troublesome since most DP algorithms cannot anticipate them, although 1.4% of simple pairs in various defined SSs are pseudo-knotted. The fundamental issue is that most DP algorithms presume, in order to speed up the estimation, that the overall folding FE change of a secondary system with two branches is the amount of the FE change of each branch calculated separately. If pseudoknots develop between the divisions, it does not work anymore. With increasing sequence length, DP algorithms that forecast pseudoknots scale poorly in time and are quite sluggish. For instance, the standard set of DP algorithms for FEM as well as partition function calculation scales $O(N^3)$ in time, whereby N is the number of nucleotides in the sequence. PKNOTS is a DP algorithm that can forecast the most known topology, yet $O(N^6)$ scales [107]. This means that doubling the sequence length takes 8 times more computing time by conventional methods but 64 times more when pseudoknots are taken into account. This restricts the use of these algorithms to sequences of up to 100–200 nucleotides. Numerous different DP algorithms scale better $O(N^5)$ or $O(N^4)$, but cannot forecast as many established pseudoknot topologies [108, 109]. Pknots RG, by Reeder and Giegerich [110], which scale $O(N^4)$ are one of the best to this group of algorithms.

Pseudoknots can be predicted using heuristics in acceptable computational time. However, the trade-off is that no certainty exists for estimating the lowest FE structure. In a software called ILM [111], one heuristic algorithm is introduced. It is based on a repeated (iterated) prediction of the structure with the so-called "loop matching algorithm." Each repetition forecasts a non-pseudoknotted structure, from which the highest score helix will be selected for the final structure. The paired nucleotides from the previously selected helixes are discarded in the next structure prediction iteration for each repeat. Since nucleotides are eliminated from successive measurements in pairs, the selected helices' collection may be pseudo-knotted in the final assembly. The algorithm scales $O(mN^3)$ in the worst case for m loop matching calculations. Another heuristic algorithm is introduced in the software HotKnots [112]. HotKnots commonly use many calls to a DP algorithm to assemble pseudoknots constructs, but at each point, several alternative helixes are expected simultaneously. This tests a number of SSs, which are ordered by increased FE changes at the end of the measurement.

10.3.1.3 Multiple-Sequence SSs Prediction

The multi-sequence SSs estimate tries to mimic a CSA by forecasting structures retained in two or more sequences. These techniques are not as reliable as manual CSA, but they can greatly increase precision over single sequence methods [63]. However, many of these techniques are more computationally costly than single sequence approaches.

Algorithms That Simultaneously Fold and Align

David Sankoff suggested the first method for folding homologous RNAs [113]. It concurrently considers the alignment as well as the folding of any amount of RNA sequences in a single measurement. The algorithm formally scales $O(N^{3s})$ in time, and $O(N^{2s})$ in storage, i.e., the memory usage, for s sequences of having length N. As other prediction methods of the single-sequence structure, this algorithm cannot predict pseudoknots. This algorithm is costly, particularly for more than two sequences. However, the limitation of alignment to eliminate impossible biological alignments and pairs provides major time-cost improvements. FOLDALIGN [114] was the first minimal version of the Sankoff algorithm. This algorithm used base-pair optimization instead of the FEM and nearest neighbor approach. It also removed branched structures from consideration, which reduced the algorithm to $O(N^4)$ in time, but removed a common and significant motif within RNA structure. Latest FOLDALIGN updates also provided support for the provision of branched systems, a FE model, and a heuristic trimming to speed up computation, which greatly increases algorithm accuracy [115].

Another approach that employs the Sankoff algorithm is LocARNA [116]. LocARNA maximizes the sum of pair probabilities for both sequences that are computed by different single-sequence partition function estimates and a similarity score for alignment, rather than minimizing energy. LocARNA runs easily since only significant base pairs are regarded, reducing the order of the algorithm to $O\ (N^2(N^2 + M^2))$, for two sequences of length N, and where M is the number of significant base pairs, also on order N. It does, however, lose any precision in contrast to FOLDALIGN [66]. While it is typically challenging to expand Sankoff's algorithm directly to more than two sequences in terms of computational costs, other means have been used to adapt it for the infinite number of sequences with greatly reduced complexity. FOLDALIGNM employed pair frequencies for all pair-wise FOLDALIGN sequence measurements [117].

Initially, LocARNA was also able to work on several sequences [116]. In order to generalize to several sequences, mLocARNA employs the output of the LocARNA multi-sequence alignment calculations in pairs. The chance of a pair of aligned columns is the square root of the pairing odds of the two alignments. RAF (99) is a distinct method that aligns and plies an infinite number of sequences concurrently. RAF functions in two sets at a time and aligns successive implementations of alignments rather than sets. mLocARNA and FOLDALIGNM seemed to work better at shorter sequences in benchmarks, and RAF & Multilign appeared to operate better for longer sequences, and both seemed to outperform single and double sequence

approaches for most forms of RNAs [118]. Many of these algorithms cost the current hardware fairly.

Algorithms That Align First, Then Fold

The second paradigm to approximate the arrangement of more than one sequence is to first align and fold the sequences. This model is seen in RNAalifold [119]. RNAalifold defines a minimum FE consensus framework that a community of compatible input sequences may build. Input alignment mostly emerges from a series alignment algorithm; however, individually curated alignments are endorsed and may increase precision. RNAalifold is quick and productive for matching sequences. Its accuracy is hindered if the input is incorrectly matched, which may happen when the sequence identity of the pairs is $<60\%$. In those instances, automatic sequence alignment algorithms battle [120]. The alignment of sequences of low identification, rendered by a professional investigator, or at least modified, could have provided good results. The CentroidFold algorithm [121] also exists in the model "align, then fold." It investigates the central frame in a way that is similar to the one of Sfold [122] rather than considering a minimum free-energy consensus structure. The central structure of the largest structural cluster represents the central structure, created by the stochastic sampling of homologous structures series. The chances of identifying the core consensus structure can be calculated by using an experimental discovery [123] or by using the nearest model from a database of unique structure sequences.

Recently, the TurboFold model requires an unlimited amount of sequences in the "align, then fold" model and then tests their pair-wise probabilistic alignment and their base-pair probabilities [124]. These alignments are employed for configurations among sequences. The single-sequence base pair probabilities for a provided sequence within the collection are referred to as the "intrinsic information." For any other sequence, the combined probabilistic aligning and the base couple probabilities are referred to as the "extrinsic information." The updated probabilities of the pair would then be used to recalculate extrinsic information. Many iterations strengthen and improve the predicted chances of the pair for series. The architectures are designed with the highest expected accuracy algorithm after the required number of iterations. In random assortments, TurboFold typically outperforms other predictive algorithms in multiple sequences that normally include identities that are less than 60% in pairs and that are commonly comparable in PPV sequences [124]. Although it may be more expensive to compute than any of the above alignment and then fold algorithms, up to 10 RNA sequences of normal lengths per minute are required. One of TurboFold's important advantages over most of the above-noted algorithms is that it does not enforce a common structure. Variable elements can also be properly predicted in homologous sequences, including the variable stem in tRNAs, allowing TurboFold a convincing alternative for structural prediction, where different sequences can be used in divergent identities and ambiguous alignment.

An Algorithm That Folds, Then Aligns

In multi-sequence structure prediction, the third paradigm is to "fold, then align." RNAshapes [125] follows this method. This algorithm lists separately the abstract "shape" space accessible for each sequence and determines the probability of each form, and then defines the thermodynamically optimal configuration with the typical form. Instead of full pairing details, abstract shapes encode RNA structure features. There are far less low FE sources than systems, so the solution is feasible. RNAshapes is fast; after single sequence structure measurements, it is roughly linear in time. It gives precision comparable to the above multi-sequence approaches. It does not provide series alignments but can be created with RNAforester [126] from the retained structure.

10.3.2 Three-Dimensional Structure Prediction Methods

Although the SSs offers the blueprint for an RNA molecule, information about the RNA 3D structure remains key to an overall understanding of its role. Initial 3D structure modeling was carried out successfully by RNA structure experts with the 3D structures of several typical RNA molecules, like tRNAs [100], the group I introns [127], and RNase P [128]. In recent years, a range of computational models for the prediction of RNA 3D structures has been developed [1]. These models may be broadly categorized into two groups, i.e., depending on the knowledge or physical property.

10.3.2.1 Knowledge-Dependent Modeling

RNA 3D structures may be predicted by assembling established motifs or the aligning sequence with already available experimentally defined structures in the database. Knowledge-based modeling primarily involves modeling on the basis of graphics and homology-based modeling (HBM) [129].

Graphics-Based Methods

The graphics modeling typically offers a visual interface, which enables users to create 3D RNA constructs by controlling or assembling segments of RNA [130–134]. Few of the major graphics-based algorithms are MANIP, ERNA-3D, and RNA2D3D. The MANIP helps users to design known 3D models on the computer screen using the corresponding SSs predicted through CSA [130]. Although the MANIP is not an automated process, it provides a quick and simple way to construct 3D RNA structures, particularly large RNAs, such as the RNase P RNA [130]. Moreover, multiple relationship tables, as well as base-pair tables that specifically contain RNAs' topological information, can be used to precisely model RNAs' interactions [130].

In order to create RNA 3D structures from sequences as well as SSs, the ERNA-3D offers users a graphic interface in order to freely position the A-form helixes and to explicitly draw the single inter-helical strands [135]. The 3D structures of mRNA, rRNAs, and tRNAs, including 16S rRNAs, 5S rRNA, and 23S rRNA, were

successfully produced using ERNA-3D [135]. The RNA2D3D will forecast rough 3D structures for large RNAs easily based on their SSs, e.g., viral kissing loops, ribozymes, and various RNA nanostructures [136]. Manual handling, though, must be performed to create a graphical interface to achieve a better structure, like compacting, energy refinement, stalking, and segment-positioning [137]. While the graphics-based approaches introduced above can be used for creating 3D structures for large RNAs with hundreds of nucleotides quickly and intuitively, since they are manual techniques, they require users to set up and optimize the RNA structure models according to particular concepts utilizing the tools provided in the software packages. Thus, in order to construct plausible systems, it is important for users to have extensive knowledge of RNA systems.

Homology-Based Modeling

Although a macromolecule's 3D structure experiences change much slower in comparison to its sequence, evolutionarily associated macromolecules normally preserve similar 3D structure though divergences at the sequence level. On this basis, 3D macromolecule structures are able to be built by aligning the target molecule sequence to molecular structure templates [134]. HBM, also referred to as comparative modeling or template-based modeling, was very effective in the prediction of 3D protein structure [138, 139]. Additionally, HBM has been expanded to include fragment assembly methods like 3dRNA [140] and RNAComposer [141]. 3dRNA is a quick and automatic 3D algorithm designed to construct RNA structure by assembling A-form helixes and various loops, whose structures are extracted in a database from known structures [140]. 3dRNA predicts reliable 3D structures based on its SSs for 300 RNA tested, including pseudoknots, duplexes, and hairpins. In addition, 3dRNA can also be used freely online as a database server, and the projected 3D structure can be accessed rapidly with the sequence and SSs as data [140].

ModeRNA enables both the simplified structure forecast from a series of templates/alignments as well as user-controlled structure manipulations, i.e., the fragment assembly [142]. ModeRNA understands as well as models post-transcriptional alteration of nucleosides compared with other modeling algorithms. It is pertinent to note that even though ModeRNA is not a method focused on graphics, it also demands that users should have alignment among the RNAs template and the target RNA and define the base pairs between the embedded fragment and the rest of the RNA [142]. RNAComposer is another web server that can use the RNA FRABASE database for predicting 3D structures for large RNAs [141]. The RNA FRABASE database can be considered as a dictionary linked to the RNA SSs with established fragments of a tertiary structure. The SSs that a user provides in the RNAComposer is, first of all, broken into elements such as stem, loops, and individual strands and subsequently scanned the related tertiary structural elements automatically from an RNA FRABASE database as well as assembled into full 3D structures.

The key benefit of HBM is that the size of the RNAs to be modeled is not necessarily limited. The consistency of the projected structures depends, however,

on the sequence alignment consistency, template structures, and secondary frameworks identified by the user. Although the amount of identified RNA structures stored, the PDB/NDB database is growing quickly, and it might yet be challenging to locate accurate template RNAs for a given target RNA. In addition, owing to their strong stability, the configurations of their RNA are normally modified with solutions such as ion conditions and temperature [143], and other ligands or macromolecules. Moreover, the creation of a good alignment of RNAs with complicated systems typically involves laborious manual planning dependent on proven expertise in the most significant RNA families. HBM is, therefore, not always accurate.

10.3.2.2 Physics-Based Modeling

Physics-related methods are based on biophysical concepts that concurrently scan for the conformation to fold with minimum free energy. Since complete atomic structure modeling for an RNA typically requires several degrees of freedom and thus tremendous computational sophistication, many CG predictive models with physical simplifications have also been developed at various resolution levels.

All-Atomistic Model

Until today, the "all-atomistic molecular dynamics" are highly required for understanding macromolecule simulation, which in turn provides an insight into the real movement of atoms, such as AMBER [140, 144] and CHARMM [145] with physics-based force fields. However, considering the several degrees of freedom, it remains challenging for folding RNA 3D structures even with advanced computing strategies. The models were then evolved considering the recognized or secondary fragments [146], such as the MC-fold/Mc-Sym Pipeline (MMP) [76] and FARNA/FARFAR [147]. Because SSs can provide enough structural constraints for automated construction of 3D structures, the MMP infers RNA SSs from sequence data and subsequently assembles a set of 3D structures based on their SSs [76]. Unlike the thermodynamics approaches such as Mfold [148]. The MC-Fold can forecast RNA SSs, including noncanonical and canonical base pairs, for the usage of a knowledge-related scoring function associated with the NCM (nucleotide cycle motif) databases. The NCMs that are circularly bound through covalent bonds, pairing or stacking interactions, were actually developed from a study of the X-ray crystallographic structures. The MC-Sym along with the 3D NCMs as well as the Las Vegas algorithm was employed for the fragment insertion simulation. MMP has been authenticated by constructing 3D structures of precursor microRNA as well as human immunodeficiency virus (HIV1) *cis*-acting-1 frame-shifting segment [76].

Das and Baker discussed FARNA's completely automatic, energy-based solution to predict RNA 3D structure [74]. FARNA integrates trinucleotide fragments obtained from the ribosome crystal structure into a completely atomistic structure that is compatible with the particular sequence by utilizing the Monte Carlo algorithm as well as the simpler knowledge-based energy feature that favors stacking, base pairing, and geometry. The CG core pairing capacity employed in FARNA is focused on the mathematical study of the ribosomal basis, and not only Watson–

Crick base pairs. However, the interactions along with Hoogsteen as well as sugar edges may be taken into account. FARFAR implements a high-resolution process of refining into FARNA in order to forecast and design the atomic precision of noncanonical small RNA structures [147]. In another study, RNAnbds was built for predicting RNA's 3D structures through fragmentary assembly, on the basis of statistics of bases as per their sequence/space neighbors in databases [149]. RNAnbds offers a good predictor for short fragments (<15 nucleotides), in specific RMSD loops <4, together with statistical potentials like base stacking and base pairing.

Coarse-Grained Model

Another important approach for minimizing the computational expense is to decrease the number of objects by handling a set of usable atoms with a single bead [150, 151]. The bead can either denote a few or a large number of atoms on the basis of the model's resolution. Following the initially "one-bead RNA model" designed by Malhotra and Harvey [152], several CG models were implemented for the purposes of predicting RNA 3D structures [153] or modeling interactions among RNAs as well as other molecules [154, 155], for example, NAST [156], YUP [157], and iFold [75]. The YUP is a very versatile molecular mechanic algorithm for CG and multi-scaling modeling [157]. The YUP is employed for modeling RNA, protein, and DNA structures on the basis of the related energy potentials and approaches such as Monte Carlo, energy minimization, and molecular dynamics. In YUP, one nucleotide is substituted by a pseudo-atom at the middle of phosphorous atoms in order to model high RNAs, which decreases device costs efficiently. While YUP needs users to supply the information about the SSs of RNAs and the force field, YUP is an adaptive RNA modeling kit for automated CG modeling [157].

Like YUP [157], NAST is another "one-bead RNA model" in which a nucleotide's C3' atom is picked to reflect the whole nucleotide [158]. The NAST will sample conformations that fulfill a certain range of secondary structure as well as tertiary interaction limit, with an RNA basic knowledge-based ability and a simple molecular dynamic algorithm. One benefit of NAST is its capacity to integrate experimental data as a filter for structurally equivalent conformation clusters, for example, with perfect small-angle X-ray dispersing data as well as experimental solvent accessibility data. Earlier, NAST was employed for predicting the yeast phenylalanine tRNA's 3D structures (76 nucleotides) and the Tetrahymena thermophile group I intron's the P4-P6 domain of 158 nucleotides within 8 Å as well as 16 Å RMSDs retrieved from experimental structures, respectively [158]. The iFold is another web-based algorithm that is being built by the Doknolyan community and can be used for predicting RNA's 3D structures [75]. The model employs three-nucleotide beads of CG representation and efficient molecular dynamic simulations with step-by-step potentials like the base pairing. iFold's strength has been seen in forecasting the 3D structures of 150 RNA with different sequences, with <4 Å deviations in experimental structures in the majority of predicted structures [159].

10.4 Conclusion and Future Perspective

In conclusion, to date, numerous computational approaches have been developed for predicting the secondary as well as the three-dimensional structure of RNA. However, the authors believe that there is still scope for the development of novel tools and techniques which can predict a more accurate structure [160]. Combining experimental and computational approaches for predicting the structure of RNA will enable us to understand its structure and function more precisely. For instance, smFRET as well as NMR spectroscopy are useful tools to evaluate the several states that the RNA can embrace. The PARIS approach for in vivo crosslinking also has a significant potential to include several instances of multiple-folded RNA. Improvements in cryoelectron microscopy, as well as tomography along with direct electron detectors, advanced contrast methods as well as single particle detection, can allow direct observation of single RNA molecules feasible in several configurations. The cryoelectron microscopy group has already designed novel sophisticated systems for categorizing related configurations of complex macromolecular assemblies. The least number as well as the length of RNA helices obtained from crystallography or cryoelectron microscopy often offer an important restriction for RNA folding [160]. Good measurements are also needed to assess differences among RNA structures. For instance, the analysisDist tool available within the program kit of the Vienna RNA provides several alternatives for measuring matrix distances with Ward's method, Shapiro's cost matrix for coarse structures, or Saitou's neighbor-joining method [87]. Thus, RNA structure prediction continues to progress with novel metrics as well as more experimentally specified examples of multi-conformation RNA assemblies. A single minimum free energy configuration for the RNA sequence would become a greater appreciation of the several potential different configurations encrypted in the RNA sequence and precise forecasts will direct the evaluation of detailed transcript and transcriptome-wide research studies.

Conflict of Interest None

References

1. Hajdin CE, Bellaousov S, Huggins W, Leonard CW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. PNAS. 2013;110(14):5498–503.
2. Vandivier LE, Anderson SJ, Foley SW, Gregory BD. The conservation and function of RNA secondary structure in plants. Annu Rev Plant Biol. 2016;67:463–88.
3. Kashi K, Henderson L, Bonetti A, Carninci P. Discovery and functional analysis of lncRNAs: methodologies to investigate an uncharacterized transcriptome. Biochim Biophys Acta. 2016;1859(1):3–15.
4. Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. Cell. 2014;157(1):77–94.

5. Achar A, Sætrom P. RNA motif discovery: a computational overview. Biol Direct [Internet]. 2015 [cited 2020 Dec 15];10. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4600295/.
6. Lim CS, Brown CM. Know your enemy: successful bioinformatic approaches to predict functional RNA structures in viral RNAs. Front Microbiol [Internet]. 2018 [cited 2020 Dec 15];8. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02582/full#h10.
7. Li H, Zhu D, Zhang C, Han H, Crandall KA. Characteristics and prediction of RNA structure, vol. 2014 [Internet]. BioMed Research International. Hindawi; 2014 [cited 2020 Oct 25]. p. e690340. Available from: https://www.hindawi.com/journals/bmri/2014/690340/.
8. Reyes FE, Garst AD, Batey RT. Chapter 6—Strategies in RNA crystallography. In: Methods in enzymology [Internet]. Biophysical, chemical, and functional probes of RNA structure, interactions and folding: part B; vol. 469. Academic Press; 2009 [cited 2020 Oct 25]. p. 119–39. Available from: http://www.sciencedirect.com/science/article/pii/S0076687909690066.
9. Westhof E. Twenty years of RNA crystallography. RNA. 2015;21(4):486–7.
10. Fernandez-Leiro R, Scheres SHW. Unravelling the structures of biological macromolecules by cryo-EM. Nature. 2016;537(7620):339–46.
11. Fürtig B, Richter C, Wöhnert J, Schwalbe H. NMR spectroscopy of RNA. Chembiochem. 2003;4(10):936–62.
12. Miao Z, Adamiak RW, Blanchet M-F, Boniecki M, Bujnicki JM, Chen S-J, et al. RNA-puzzles round II: assessment of RNA structure prediction programs applied to three large RNA structures. RNA. 2015;21(6):1066–84.
13. Magnus M, Kappel K, Das R, Bujnicki JM. RNA 3D structure prediction guided by independent folding of homologous sequences. BMC Bioinform [Internet]. 2019 [cited 2020 Oct 25];20. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6806525/.
14. Puton T, Kozlowski LP, Rother KM, Bujnicki JM. CompaRNA: a server for continuous benchmarking of automated methods for RNA secondary structure prediction. Nucleic Acids Res. 2013;41(7):4307–23.
15. Seemann SE, Gorodkin J, Backofen R. Unifying evolutionary and thermodynamic information for RNA folding of multiple alignments. Nucleic Acids Res. 2008;36(20):6355–62.
16. Weinreb C, Riesselman AJ, Ingraham JB, Gross T, Sander C, Marks DS. 3D RNA and functional interactions from evolutionary couplings. Cell. 2016;165(4):963–75.
17. Bonneau R, Strauss CEM, Baker D. Improving the performance of rosetta using multiple sequence alignment information and global measures of hydrophobic core formation. Proteins. 2001;43(1):1–11.
18. Sun XS. 2—Plant materials formation and growth. In: Wool RP, Sun XS, editors. Bio-based polymers and composites [Internet]. Burlington: Academic Press; 2005 [cited 2020 Oct 25]. p. 15–32. Available from: http://www.sciencedirect.com/science/article/pii/B9780127639529500034.
19. Sharma D, Singh S, Chand T, Kumar P. RNA: structure, prediction, and visualization tools. In: Intelligent communication, control and devices. New York: Springer; 2018. p. 335–45.
20. Feher J. 2.2—DNA and protein synthesis. In: Feher J, editor. Quantitative human physiology. 2nd ed [Internet]. Boston: Academic Press; 2017 [cited 2020 Oct 25]. p. 120–9. Available from: http://www.sciencedirect.com/science/article/pii/B9780128008836000112.
21. Goss DJ, Domashevskiy AV. Messenger RNA (mRNA): the link between DNA and protein. In: Bradshaw RA, Stahl PD, editors. Encyclopedia of cell biology [Internet]. Waltham: Academic Press; 2016 [cited 2020 Oct 25]. p. 341–5. Available from: http://www.sciencedirect.com/science/article/pii/B9780123944474100409.
22. Dunckley T, Parker R. RNA turnover. In: Brenner S, Miller JH, editors. Encyclopedia of genetics [Internet]. New York: Academic Press; 2001 [cited 2020 Oct 25]. p. 1748–51. Available from: http://www.sciencedirect.com/science/article/pii/B0122270800011381.

23. Nazar RN. Ribosomal RNA processing and ribosome biogenesis in eukaryotes. IUBMB Life. 2004;56(8):457–65.
24. Hang R, Wang Z, Deng X, Liu C, Yan B, Yang C, et al. Ribosomal RNA biogenesis and its response to chilling stress in Oryza sativa. Plant Physiol. 2018;177(1):381–97.
25. Cooper GM. The cell: a molecular approach. Washington, DC/Sunderland, MA: ASM Press/ Sinauer Associates; 2000.
26. Doherty J, Guo M. Transfer RNA. In: Bradshaw RA, Stahl PD, editors. Encyclopedia of cell biology [Internet]. Waltham: Academic Press; 2016 [cited 2020 Oct 25]. p. 309–40. Available from: http://www.sciencedirect.com/science/article/pii/B9780123944474100392.
27. O'Donoghue P, Ling J, Söll D. Transfer RNA function and evolution. RNA Biol. 2018;15 (4–5):423–6.
28. Fields RN, Roy H. Deciphering the tRNA-dependent lipid aminoacylation systems in bacteria: novel components and structural advances. RNA Biol. 2018;15(4–5):480–91.
29. Alamos P, Tello M, Bustamante P, Gutiérrez F, Shmaryahu A, Maldonado J, et al. Functionality of tRNAs encoded in a mobile genetic element from an acidophilic bacterium. RNA Biol. 2018;15(4–5):518–27.
30. Castillo A, Tello M, Ringwald K, Acuña LG, Quatrini R, Orellana O. A DNA segment encoding the anticodon stem/loop of tRNA determines the specific recombination of integrative-conjugative elements in Acidithiobacillus species. RNA Biol. 2018;15 (4–5):492–9.
31. Bacusmo JM, Orsini SS, Hu J, DeMott M, Thiaville PC, Elfarash A, et al. The t6A modification acts as a positive determinant for the anticodon nuclease PrrC, and is distinctively nonessential in Streptococcus mutans. RNA Biol. 2018;15(4–5):508–17.
32. Kessler AC, Kulkarni SS, Paulines MJ, Rubio MAT, Limbach PA, Paris Z, et al. Retrograde nuclear transport from the cytoplasm is required for tRNATyr maturation in T. brucei. RNA Biol. 2018;15(4–5):528–36.
33. Agris PF, Eruysal ER, Narendran A, Väre VYP, Vangaveti S, Ranganathan SV. Celebrating wobble decoding: half a century and still much is new. RNA Biol. 2018;15(4–5):537–53.
34. Rafels-Ybern À, Torres AG, Grau-Bove X, Ruiz-Trillo I, de Pouplana LR. Codon adaptation to tRNAs with inosine modification at position 34 is widespread among eukaryotes and present in two bacterial phyla. RNA Biol. 2018;15(4–5):500–7.
35. Kessler AC, d'Almeida GS, Alfonzo JD. The role of intracellular compartmentalization on tRNA processing and modification. RNA Biol. 2018;15(4–5):554–66.
36. Marz M, Gruber AR, Zu Siederdissen CH, Amman F, Badelt S, Bartschat S, et al. Animal snoRNAs and scaRNAs with exceptional structures. RNA Biol. 2011;8(6):938–46.
37. Jorjani H, Kehr S, Jedlinski DJ, Gumienny R, Hertel J, Stadler PF, et al. An updated human snoRNAome. Nucleic Acids Res. 2016;44(11):5068–82.
38. Decatur WA, Fournier MJ. rRNA modifications and ribosome function. Trends Biochem Sci. 2002;27(7):344–51.
39. Kiss T. New embo member's review. EMBO J. 2001;20(14):3617–22.
40. McKeegan KS, Debieux CM, Boulon S, Bertrand E, Watkins NJ. A dynamic scaffold of pre-snoRNP factors facilitates human box C/D snoRNP assembly. Mol Cell Biol. 2007;27 (19):6782–93.
41. Ganot P, Caizergues-Ferrer M, Kiss T. The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. Genes Dev. 1997;11(7):941–56.
42. Lafontaine DL, Bousquet-Antonelli C, Henry Y, Caizergues-Ferrer M, Tollervey D. The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. Genes Dev. 1998;12(4):527–37.
43. Ganot P, Bortolin ML, Kiss T. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. Cell. 1997;89(5):799–809.

44. Bortolin ML, Ganot P, Kiss T. Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. EMBO J. 1999;18(2):457–69.
45. O'Brien J, Hayder H, Zayed Y, Peng C. Overview of microRNA biogenesis, mechanisms of actions, and circulation. Front Endocrinol [Internet]. 2018 [cited 2020 Oct 25];9. Available from: https://www.frontiersin.org/articles/10.3389/fendo.2018.00402/full.
46. Fu G, Brkić J, Hayder H, Peng C. MicroRNAs in human placental development and pregnancy complications. Int J Mol Sci. 2013;14(3):5519–44.
47. Tüfekci KU, Öner MG, Meuwissen RLJ, Genç Ş. The role of microRNAs in human diseases. In: Yousef M, Allmer J, editors. miRNomics: microRNA biology and computational analysis [Internet]. Methods in molecular biology. Totowa, NJ: Humana Press; 2014 [cited 2020 Oct 25]. p. 33–50. Available from: https://doi.org/10.1007/978-1-62703-748-8_3.
48. Paul P, Chakraborty A, Sarkar D, Langthasa M, Rahman M, Bari M, et al. Interplay between miRNAs and human diseases. J Cell Physiol. 2018;233(3):2007–18.
49. Huang W. MicroRNAs: biomarkers, diagnostics, and therapeutics. In: Huang J, Borchert GM, Dou D, Huan J (Luke), Lan W, Tan M, et al., editors. Bioinformatics in microRNA research [Internet]. Methods in molecular biology. New York, NY: Springer; 2017 [cited 2020 Oct 25]. p. 57–67. Available from: https://doi.org/10.1007/978-1-4939-7046-9_4.
50. Hari R, Parthasarathy S. Prediction of coding and non-coding RNA. In: Ranganathan S, Gribskov M, Nakai K, Schönbach C, editors. Encyclopedia of bioinformatics and computational biology [Internet]. Oxford: Academic Press; 2019 [cited 2020 Oct 25]. p. 230–40. Available from: http://www.sciencedirect.com/science/article/pii/B978012809633820099X.
51. Fang Y, Fullwood MJ. Roles, functions, and mechanisms of long non-coding RNAs in cancer. Genomics Proteomics Bioinform. 2016;14(1):42–54.
52. Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, et al. The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature. 2014;505(7485):635–40.
53. Kutter C, Watt S, Stefflova K, Wilson MD, Goncalves A, Ponting CP, et al. Rapid turnover of long noncoding RNAs and the evolution of gene expression. PLoS Genet. 2012;8(7):e1002841.
54. Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell. 2011;147(7):1537–50.
55. Lin N, Chang K-Y, Li Z, Gates K, Rana ZA, Dang J, et al. An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. Mol Cell. 2014;53(6):1005–19.
56. Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, et al. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell. 1992;71(3):515–26.
57. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. Science. 2005;309(5740):1559–63.
58. Seetin MG, Mathews DH. RNA structure prediction: an overview of methods. In: Keiler KC, editor. Bacterial regulatory RNA: methods and protocols [Internet]. Methods in molecular biology. Totowa, NJ: Humana Press; 2012 [cited 2020 Oct 25]. p. 99–122. Available from: https://doi.org/10.1007/978-1-61779-949-5_8.
59. Waldispühl J, Reinharz V. Modeling and predicting RNA three-dimensional structures. In: Picardi E, editor. RNA bioinformatics [Internet]. Methods in molecular biology. New York, NY: Springer; 2015 [cited 2020 Oct 25]. p. 101–21. Available from: https://doi.org/10.1007/978-1-4939-2291-8_6.
60. Hamada M, Sato K, Asai K. Improving the accuracy of predicting secondary structure for aligned RNA sequences. Nucleic Acids Res. 2011;39(2):393–402.
61. Sato K, Kato Y, Akutsu T, Asai K, Sakakibara Y. DAFS: simultaneous aligning and folding of RNA sequences via dual decomposition. Bioinformatics. 2012;28(24):3218–24.

62. Lindgreen S, Gardner PP, Krogh A. Measuring covariation in RNA alignments: physical realism improves information measures. Bioinformatics. 2006;22(24):2988–95.
63. Xu Z, Mathews DH. Multilign: an algorithm to predict secondary structures conserved in multiple RNA sequences. Bioinformatics. 2011;27(5):626–32.
64. Kiryu H, Tabei Y, Kin T, Asai K. Murlet: a practical multiple alignment tool for structural RNA sequences. Bioinformatics. 2007;23(13):1588–98.
65. Tabei Y, Kiryu H, Kin T, Asai K. A fast structural multiple alignment method for long RNA sequences. BMC Bioinform. 2008;9:33.
66. Harmanci AO, Sharma G, Mathews DH. PARTS: probabilistic alignment for RNA joinT secondary structure prediction. Nucleic Acids Res. 2008;36(7):2406–17.
67. Knudsen B, Hein J. Pfold: RNA secondary structure prediction using stochastic context-free grammars. Nucleic Acids Res. 2003;31(13):3423–8.
68. Doose G, Metzler D. Bayesian sampling of evolutionarily conserved RNA secondary structures with pseudoknots. Bioinformatics. 2012;28(17):2242–8.
69. Seetin MG, Mathews DH. TurboKnot: rapid prediction of conserved RNA secondary structures including pseudoknots. Bioinformatics. 2012;28(6):792–8.
70. Chen X, Li Y, Umarov R, Gao X, Song L. RNA secondary structure prediction by learning unrolled algorithms. In 2019 [cited 2020 Dec 16]. Available from: https://openreview.net/forum?id=S1eALyrYDH.
71. Barsacchi M, Novoa EM, Kellis M, Bechini A. SwiSpot: modeling riboswitches by spotting out switching sequences. Bioinformatics. 2016;32(21):3252–9.
72. Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. 1981;9(1):133–48.
73. Frellsen J, Moltke I, Thiim M, Mardia KV, Ferkinghoff-Borg J, Hamelryck T. A probabilistic model of RNA conformational space. PLoS Comput Biol. 2009;5(6):e1000406.
74. Das R, Baker D. Automated de novo prediction of native-like RNA tertiary structures. PNAS. 2007;104(37):14664–9.
75. Sharma S, Ding F, Dokholyan NV. iFoldRNA: three-dimensional RNA structure prediction and folding. Bioinformatics. 2008;24(17):1951–2.
76. Parisien M, Major F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. Nature. 2008;452(7183):51–5.
77. Rother M, Milanowska K, Puton T, Jeleniewicz J, Rother K, Bujnicki JM. ModeRNA server: an online tool for modeling RNA 3D structures. Bioinformatics. 2011;27(17):2441–2.
78. Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, Herschlag D, et al. Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. RNA. 2009;15(2):189–99.
79. Flores SC, Altman RB. Turning limited experimental information into 3D models of RNA. RNA. 2010;16(9):1769–78.
80. Eriksson ESE, Joshi L, Billeter M, Eriksson LA. De novo tertiary structure prediction using RNA123—benchmarking and application to Macugen. J Mol Model. 2014;20(8):2389.
81. Rivas E. The four ingredients of single-sequence RNA secondary structure prediction. A unifying perspective. RNA Biol. 2013;10(7):1185–96.
82. Gutell RR, Lee JC, Cannone JJ. The accuracy of ribosomal RNA comparative structure models. Curr Opin Struct Biol. 2002;12(3):301–10.
83. Domer JE, Ichinose H. Cellular immune responses in guinea pigs immunized with cell walls of Histoplasma capsulatum prepared by several different procedures. Infect Immun. 1977;16(1):293–301.
84. Schroeder SJ, Turner DH. Optical melting measurements of nucleic acid thermodynamics. Methods Enzymol. 2009;468:371–87.
85. Janssen S, Giegerich R. The RNA shapes studio. Bioinformatics. 2015;31(3):423–5.
86. Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinform. 2010;11(1):129.

87. Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011;6(1):26.
88. Zakov S, Goldberg Y, Elhadad M, Ziv-ukelson M. Rich parameterization improves RNA structure prediction. J Comput Biol. 2011;18(11):1525–42.
89. Sato K, Hamada M, Asai K, Mituyama T. CentroidFold: a web server for RNA secondary structure prediction. Nucleic Acids Res. 2009;37(Web Server issue):W277–80.
90. Do CB, Woods DA, Batzoglou S. CONTRAfold: RNA secondary structure prediction without physics-based models. Bioinformatics. 2006;22(14):e90–8.
91. Xu X, Chen S-J. Physics-based RNA structure prediction. Biophys Rep. 2015;1(1):2–13.
92. Nowakowski J, Tinoco I. RNA structure and stability. Semin Virol. 1997;8(3):153–65.
93. Westhof E, Fritsch V. RNA folding: beyond Watson–Crick pairs. Structure. 2000;8(3): R55–65.
94. Jabbari H, Wark I, Montemagno C, Will S. Knotty: efficient and accurate prediction of complex RNA pseudoknot structures. Bioinformatics. 2018;34(22):3849–56.
95. Bellaousov S, Mathews DH. ProbKnot: fast prediction of RNA secondary structure including pseudoknots. RNA. 2010;16(10):1870–80.
96. Reeder J, Giegerich R. Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics. BMC Bioinform. 2004;5(1):104.
97. Sloma MF, Mathews DH. Base pair probability estimates improve the prediction accuracy of RNA non-canonical base pairs. PLoS Comput Biol. 2017;13(11):e1005827.
98. zu Siederdissen CH, Bernhart SH, Stadler PF, Hofacker IL. A folding algorithm for extended RNA secondary structures. Bioinformatics. 2011;27(13):i129–36.
99. Madison JT, Everett GA, Kung H. Nucleotide sequence of a yeast tyrosine transfer RNA. Science. 1966;153(3735):531–4.
100. Levitt M. Detailed molecular model for transfer ribonucleic acid. Nature. 1969;224 (5221):759–63.
101. Gardner PP, Daub J, Tate J, Moore BL, Osuch IH, Griffiths-Jones S, et al. Rfam: Wikipedia, clans and the "decimal" release. Nucleic Acids Res. 2011;39(Database issue):D141–5.
102. Baxevanis AD, Ouellette BFF, editors. Bioinformatics: a practical guide to the analysis of genes and proteins. Hoboken, NJ: Wiley-Interscience; 2004. 560 p
103. Mathews DH, Turner DH, Watson RM. RNA secondary structure prediction. Curr Protoc Nucleic Acid Chem. 2007;CHAPTER 11:Unit-11.2.
104. Eddy SR. How do RNA folding algorithms work? Nat Biotechnol. 2004;22(11):1457–8.
105. Ding Y, Lawrence CE. A statistical sampling algorithm for RNA secondary structure prediction. Nucleic Acids Res. 2003;31(24):7280–301.
106. Ding Y, Chan CY, Lawrence CE. RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. RNA. 2005;11(8):1157–66.
107. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. PNAS. 2004;101(19):7287–92.
108. Condon A, Davy B, Rastegari B, Zhao S, Tarrant F. Classifying RNA pseudoknotted structures. Theor Comput Sci. 2004;320(1):35–50.
109. Dirks RM, Pierce NA. An algorithm for computing nucleic acid base-pairing probabilities including pseudoknots. J Comput Chem. 2004;25(10):1295–304.
110. Stoddard CD, Montange RK, Hennelly SP, Rambo RP, Sanbonmatsu KY, Batey RT. Free state conformational sampling of the SAM-I riboswitch aptamer domain. Structure. 2010;18 (7):787–97.
111. Ruan J, Stormo GD, Zhang W. An iterated loop matching approach to the prediction of RNA secondary structures with pseudoknots. Bioinformatics. 2004;20(1):58–66.
112. Ren J, Rastegari B, Condon A, Hoos HH. HotKnots: heuristic prediction of RNA secondary structures including pseudoknots. RNA. 2005;11(10):1494–504.
113. Sankoff D. Simultaneous solution of the RNA folding, alignment and protosequence problems. SIAM J Appl Math. 1985;45(5):810–25.

114. Gorodkin J, Heyer LJ, Stormo GD. Finding the most significant common sequence and structure motifs in a set of RNA sequences. Nucleic Acids Res. 1997;25(18):3724–32.
115. Havgaard JH, Torarinsson E, Gorodkin J. Fast pairwise structural RNA alignments by pruning of the dynamical programming matrix. PLoS Comput Biol. 2007;3(10):e193.
116. Will S, Reiche K, Hofacker IL, Stadler PF, Backofen R. Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. PLoS Comput Biol. 2007;3(4):e65.
117. Torarinsson E, Havgaard JH, Gorodkin J. Multiple structural alignment and clustering of RNA sequences. Bioinformatics. 2007;23(8):926–32.
118. Do CB, Foo C-S, Batzoglou S. A max-margin model for efficient simultaneous alignment and folding of RNA sequences. Bioinformatics. 2008;24(13):i68–76.
119. Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. RNAalifold: improved consensus structure prediction for RNA alignments. BMC Bioinform. 2008;9(1):474.
120. Gardner PP, Wilm A, Washietl S. A benchmark of multiple sequence alignment programs upon structural RNAs. Nucleic Acids Res. 2005;33(8):2433–9.
121. Hamada M, Kiryu H, Sato K, Mituyama T, Asai K. Prediction of RNA secondary structure using generalized centroid estimators. Bioinformatics. 2009;25(4):465–73.
122. Ding Y, Lawrence CE. Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond. Nucleic Acids Res. 2001;29(5):1034–46.
123. McCaskill JS. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. Biopolymers. 1990;29(6–7):1105–19.
124. Harmanci AO, Sharma G, Mathews DH. TurboFold: iterative probabilistic estimation of secondary structures for multiple RNA sequences. BMC Bioinform. 2011;12(1):108.
125. Steffen P, Voss B, Rehmsmeier M, Reeder J, Giegerich R. RNAshapes: an integrated RNA analysis package based on abstract shapes. Bioinformatics. 2006;22(4):500–3.
126. Höchsmann M, Voss B, Giegerich R. Pure multiple RNA secondary structure alignments: a progressive profile approach. IEEE/ACM Trans Comput Biol Bioinform. 2004;1(1):53–62.
127. Michel F, Westhof E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J Mol Biol. 1990;216(3):585–610.
128. Harris ME, Nolan JM, Malhotra A, Brown JW, Harvey SC, Pace NR. Use of photoaffinity crosslinking and molecular modeling to analyze the global architecture of ribonuclease P RNA. EMBO J. 1994;13(17):3953–63.
129. Shi Y-Z, Wu Y-Y, Wang F-H, Tan Z-J. RNA structure prediction: progress and perspective. Chinese Phys B. 2014;23(7):078701.
130. Massire C, Westhof E. MANIP: an interactive tool for modelling RNA. J Mol Graph Model. 1998;16(4–6):197–205, 255–7
131. Zwieb C, Wower I, Wower J. Comparative sequence analysis of tmRNA. Nucleic Acids Res. 1999;27(10):2063–71.
132. Hammann C, Westhof E. Searching genomes for ribozymes and riboswitches. Genome Biol. 2007;8(4):210.
133. Bindewald E, Grunewald C, Boyle B, O'Connor M, Shapiro BA. Computational strategies for the automated design of RNA nanoscale structures from building blocks using NanoTiler. J Mol Graph Model. 2008;27(3):299–308.
134. Jossinet F, Ludwig TE, Westhof E. Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. Bioinformatics. 2010;26(16):2057–9.
135. Lu F, Ammiraju JSS, Sanyal A, Zhang S, Song R, Chen J, et al. Comparative sequence analysis of MONOCULM1-orthologous regions in 14 Oryza genomes. Proc Natl Acad Sci U S A. 2009;106(6):2071–6.
136. Martinez HM, Maizel JV, Shapiro BA. RNA2D3D: a program for generating, viewing, and comparing 3-dimensional models of RNA. J Biomol Struct Dyn. 2008;25(6):669–83.
137. Jossinet F, Westhof E. Sequence to structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. Bioinformatics. 2005;21(15):3320–1.

138. Wu S, Zhang Y. A comprehensive assessment of sequence-based and template-based methods for protein contact prediction. Bioinformatics. 2008;24(7):924–31.
139. Zhang Y, Skolnick J. Segment assembly, structure alignment and iterative simulation in protein structure prediction. BMC Biol. 2013;11(1):44.
140. Zhao Y, Huang Y, Gong Z, Wang Y, Man J, Xiao Y. Automated and fast building of three-dimensional RNA structures. Sci Rep. 2012;2(1):734.
141. Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, et al. Automated 3D structure composition for large RNAs. Nucleic Acids Res. 2012;40(14):e112.
142. Rother M, Rother K, Puton T, Bujnicki JM. RNA tertiary structure prediction with ModeRNA. Brief Bioinform. 2011;12(6):601–13.
143. Tan Z-J, Chen S-J. Chapter 22—Predicting electrostatic forces in RNA folding. In: Methods in enzymology [Internet]. Biophysical, chemical, and functional probes of RNA structure, interactions and folding: part B; vol. 469. Academic Press; 2009 [cited 2020 Oct 29]. p. 465–87. Available from: http://www.sciencedirect.com/science/article/pii/S0076687909690224.
144. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, et al. The Amber biomolecular simulation programs. J Comput Chem. 2005;26(16):1668–88.
145. Brooks BR, Brooks CL, MacKerell AD, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: the biomolecular simulation program. J Comput Chem. 2009;30(10):1545–614.
146. Bida JP, Maher LJ. Improved prediction of RNA tertiary structure with insights into native state dynamics. RNA (New York, NY). 2012;18(3):385–93.
147. Das R, Karanicolas J, Baker D. Atomic accuracy in predicting and designing noncanonical RNA structure. Nat Methods. 2010;7(4):291–4.
148. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003;31(13):3406–15.
149. Zhang J, Bian Y, Lin H, Wang W. RNA fragment modeling with a nucleobase discrete-state model. Phys Rev E. 2012;85(2):021909.
150. Paliy M, Melnik R, Shapiro BA. Coarse-graining RNA nanostructures for molecular dynamics simulations. Phys Biol. 2010;7(3):036001.
151. de Pablo JJ. Coarse-grained simulations of macromolecules: from DNA to nanocomposites. Annu Rev Phys Chem. 2011;62:555–74.
152. Harvey SC, Malhotra A, Tan RK-Z. Molecular modeling studies on the ribosome. Mol Eng. 1995;5(1):213–8.
153. Xia Z, Gardner DP, Gutell RR, Ren P. Coarse-grained model for simulation of RNA three-dimensional structures. J Phys Chem B. 2010;114(42):13497–506.
154. Hori N, Takada S. Coarse-grained structure-based model for RNA-protein complexes developed by fluctuation matching. J Chem Theory Comput. 2012;8(9):3384–94.
155. Denesyuk NA, Thirumalai D. Coarse-grained model for predicting RNA folding thermodynamics. J Phys Chem B. 2013;117(17):4901–11.
156. Jonikas MA, Radmer RJ, Altman RB. Knowledge-based instantiation of full atomic detail into coarse-grain RNA 3D structural models. Bioinformatics. 2009;25(24):3259–66.
157. Tan RKZ, Petrov AS, Harvey SC. YUP: a molecular simulation program for coarse-grained and multi-scaled models. J Chem Theory Comput. 2006;2(3):529–40.
158. Kerpedjiev P, Höner zu Siederdissen C, Hofacker IL. Predicting RNA 3D structure using a coarse-grain helix-centered model. RNA. 2015;21(6):1110–21.
159. Woodson SA. Metal ions and RNA folding: a highly charged topic with a dynamic future. Curr Opin Chem Biol. 2005;9(2):104–9.
160. Schroeder SJ. Challenges and approaches to predicting RNA with multiple functional structures. RNA. 2018;24(12):1615–24.