Science**PG**
Science Publishing Group

Review Article

A Comprehensive Review of Predicting Method of RNA Tertiary Structure

Yurong Yang, Zhendong Liu[*]

School of Computer Science and Technology, Shandong Jianzhu University, Jinan, China

Email address:
Liuzd2000@126.com (Zhendong Liu)
[*]Corresponding author

Abstract: In recent years, great progress has been made in the research of RNA function, and more and more RNA functions have been discovered. The function of RNA is highly dependent on its 3D structure, the RNA tertiary structure includes the RNA 3D structure and RNA tertiary interaction, so the RNA tertiary structure prediction has also attracted extensive attention. There are many RNA tertiary structure prediction algorithms. According to the traditional classification methods, the existing RNA tertiary structure prediction algorithms can be divided into two categories: the RNA tertiary structure prediction algorithm based on knowledge mining and the RNA tertiary structure prediction algorithm based on physics. On this basis, this paper further refines the RNA tertiary structure prediction algorithm based on physics in traditional classification, and proposes a new refinement classification method based on conformational sampling method, namely RNA tertiary structure prediction algorithm based on physical fragment assembly conformational sampling method and RNA tertiary structure prediction algorithm based on Stepwise ansatz conformational sampling method. We make a comparative analysis of RNA tertiary structure prediction algorithms, and put forward some suggestions for improving the energy function in the next step, in order to find an RNA tertiary structure prediction algorithm that can achieve atomic accuracy.

Keywords: RNA Structure Prediction, Tertiary Structure, Conformational Sampling, Atomic Accuracy, Stepwise Ansatz

1. Introduction

Ribonucleic acid (RNA), as a genetic information carrier, exists in biological cells as well as some viruses and viroids. RNA has various functions in organisms, and its main function is to convert the genetic information stored in DNA into proteins, and to guide the synthesis of proteins. More and more attention has been paid to the function of RNA. In recent studies, non-coding RNA has been found to control protein synthesis, regulate transcription and translation. In addition, non-coding RNA has some more complex biological functions, such as dose compensation, chromatin regulation, genomic imprinting and nuclear tissue [1]. The newly discovered RNA molecular riboswitch can autonomously sense changes in metabolite concentration and regulate gene expression at different levels [2].

The function of RNA highly depends on its tertiary structure. Understanding the RNA tertiary structure is very important, which can not only help us to further understand the relationship between structure and function, but also provides theoretical basis for the design of targeted ribosomal drugs [3]. The experimental methods used to obtain the RNA tertiary structure mainly include x-ray crystallography and frozen electron microscopy. As the available conformations increases exponentially with the RNA length, using these experimental methods to determine the RNA tertiary structure is accurate and reliable when facing the current massive RNA sequences, but time consuming and expensive [4]. Therefore, it is very necessary to predict the RNA tertiary structure by using bioinformatics methods and techniques, combined with the known biological molecular structure and its functional characteristics.

2. RNA Tertiary Structure

The RNA tertiary structure includes the spatial coordinates of all atoms in RNA (3D), and the spatial relationship between atoms embodied by atomic coordinates (tertiary interaction). There are three kinds of RNA folding structures in biology: primary structure (single-stranded base sequence), secondary structure (a collection of base pairs) and tertiary structure (spatial position of atoms) [5].
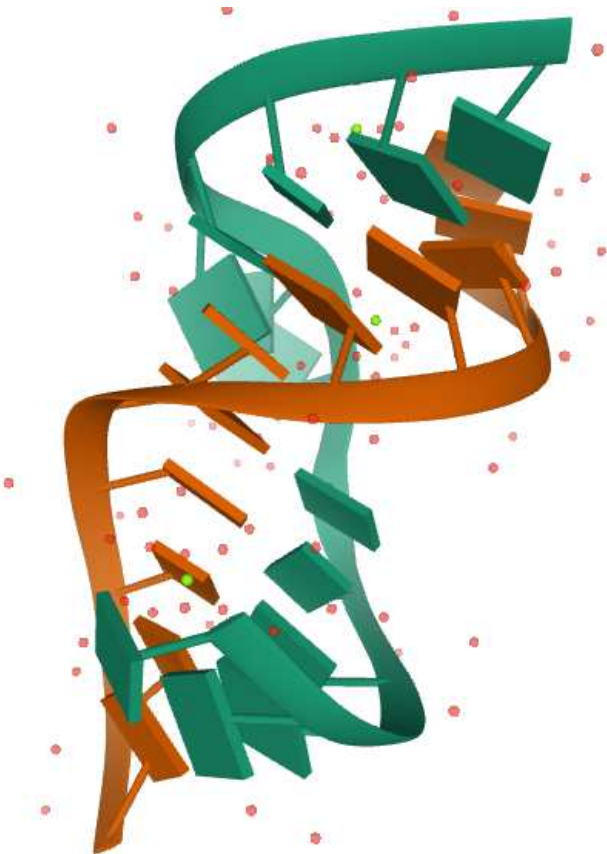

Figure 1. RNA three-dimensional (3D) structure.

The essence of the tertiary structure in the folded structure is the spatial coordinates of all the atoms in RNA molecule (3D), which is shown in Figure 1. Although the RNA secondary structure has been able to provide us with a blueprint for RNA, RNA can only exert its normal function after forming a specific tertiary structure. Therefore, knowing the RNA tertiary structure can help us understand and analyze its function and further understand the involved physiological activities.

The tertiary interactions mainly include bonded interactions, base interactions, hydrogen bond interactions, electrostatic interactions, van der Waals interactions, and other non-bonded interactions [6]. Among them, hydrogen bond interaction is the most important and characteristic interaction in RNA. Hydrogen bonds are widespread in RNA. As the bases are planar structures, as shown in Figure 2, the hydrogen donor/acceptor at the base edge can be approximately divided into three paired edges, i.e., Watson-Crick edge (W), Hoogstein edge (H) and Sugar edge (S) [7].
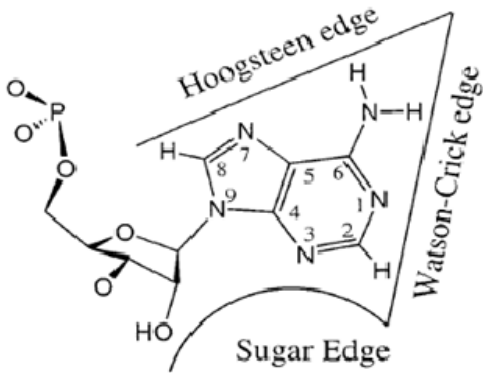

Figure 2. Definition of three sides of nucleotide base,

These edges can all be used as interaction edges. In addition, the paired edges also have cis-trans directionality. From this, we can know that in theory, four bases can form 12 kinds of hydrogen bond pairing modes [8], as shown in Table 1.

Table 1. 12 hydrogen bond pairing modes.

No.	Glycosidic Bond Orientation	Interacting Edges	Symbol	Default local strand Orientation
1	Cis	W/W		Anti-parallel
2	Trans	W/W		Parallel
3	Cis	W/H		Parallel
4	Trans	W/H		Anti-parallel
5	Cis	W/S		Anti-parallel
6	Trans	W/S		Parallel
7	Cis	H/H		Anti-parallel
8	Trans	H/H		Parallel
9	Cis	H/S		Parallel
10	Trans	H/S		Anti-parallel
11	Cis	S/S		Anti-parallel
12	Trans	S/S		Parallel

The Cis W/W interaction was the basic element in the RNA helical region, while the other 11 hydrogen bond interactions constituted the RNA structural module and the RNA tertiary structural element. A-U Cis W/W, G-C Cis W/W, and G-U Cis W/W were known as canonical base pairs. However, the study founds that canonical base pairs accounted for only about 80% of the observed RNA molecules. Although noncanonical base pairs account for 20%, they are important to ameliorate the accuracy of RNA tertiary structure prediction. Therefore, noncanonical base pairs are the key and difficult point of RNA tertiary structure prediction.

3. Algorithms of RNA Tertiary Structure Prediction

The RNA secondary structure prediction with pseudoknots has been proved to be an NP-hard problem [9], and it can be concluded that the RNA tertiary structure prediction is also an NP-hard problem. RNA tertiary structure prediction algorithms can be divided into two types by traditional classification methods, Knowledge mining-based prediction algorithm (Knowledge Mining-based) and Physical-based prediction algorithm (Physics-based). In this paper, the Physics-based prediction algorithms are further divided according to the different conformational sampling methods.

3.1. RNA Tertiary Structure Prediction Algorithm Based on Knowledge-mining

3.1.1. Fragment Assembly-based Algorithms

This kind of algorithm is to splice the known 3D RNA fragments into a tertiary structure satisfying the conditions by computer algorithm, also known as Graphics-based. MANIP [10] is a classical algorithm of this kind of algorithm, which allows users to form a complete RNA structure by assembling known 3D motif based on the secondary structure obtained by sequence alignment. However, this algorithm requires expert users to deeply master and understand the relevant knowledge of RNA structure, which is difficult for general users.

3.1.2. Homology-based Algorithms

The algorithm uses the known tertiary structure of the template sequence to determine the tertiary structure structure of the target sequence [11]. Typical algorithms include RNABuilder [12] and ModeRNA [13] which can contain post-translation information. The results of these algorithms depend on the template structure and alignment sequence, but it is difficult to find a suitable template RNA.

3.2. RNA Tertiary Structure Prediction Algorithm Based on Physics

According to the principle of biophysics, the algorithm finds out the conformation with the lowest free energy by searching the conformation space of RNA 3D structure. The algorithm is dynamic, and usually adopts Monte Carlo method or molecular dynamics simulation method [14]. The representative algorithms are FARNA, FARFAR, SWA and SWM. The key components of RNA tertiary structure prediction algorithm include molecular representation, degree of freedom, energy function and conformation sampling method [15]. Among them, the energy function and conformation sampling method are the key to ameliorate the accuracy of RNA tertiary structure modeling. In this paper, we put forward a new refined classification method based on conformational sampling, i.e., RNA tertiary structure prediction algorithm based on physical fragment assembly conformational sampling method and RNA tertiary structure prediction algorithm based on Stepwise ansatz conformational sampling method.

3.2.1. Base on Physical Fragment Assembly

The algorithm adopts a conformation sampling method based on a fragment assembly method, and on the basis, a physical energy function is introduced to guide the assembly process and assemble a 3D structure with lower energy, so that the problem that the traditional fragment assembly method is excessively dependent on a database is overcome to a certain extent, and the assembly precision of RNA fragments is improved. The representative algorithms are FARNA algorithm and FARFAR algorithm.

(1) Fragment Assembly of RNA (FARNA)

Rhiju Das and David Baker describe a physically based energy function and fully automated algorithm, which is inspired by Rosetta structure prediction method, which seeks the lowest energy tertiary structure of a given RNA sequence without using evolutionary information to minimize the dependence of fragment assembly on the database [16].

FARNA is an de novo method, which is different from the previous method. During conformational sampling, the nucleotide fragments in the initial structure were randomly replaced by the Monte carlo method while fragment assembly was performed. Moreover, when selecting RNA template, almost all the base pairing modes are included in the rRNA molecules determined by experiments, so that the RNA template obtained can be more comprehensive, so FARNA could effectively predict various RNA noncanonical base pairs.

(2) Fragment Assembly of RNA with Full-atom Refinement (FARFAR)

Since the RNA tertiary structure obtained by FARNA method is not accurate enough, an algorithm using statistical potential is designed based on FARNA by adding more accurate all-atom energy function. The FARFAR algorithm is highly accurate and achieves de novo structure prediction and design of complex motifs with unprecedented resolution, combining our previous FARNA protocol with Rosetta energy function [17].

After a rigorous test of FARFAR on a benchmark set of 32 motifs, it is found that some RNA did not achieve high resolution, Rhiju Das observed the bottleneck of this sampling method that conformations close to the natural conformation could not be sampled and lower energy could not be achieved.

3.2.2. Base on Stepwise Ansatz

It is impossible to realize high-accuracy structure prediction of RNA which is due to the imcomplete sampling of biopolymers with many degrees of freedom. So Rhiju Das put forward a effective hypothesis, called the "Stepwise ansatz", which is used to recursively construct a well-packed atomic detail model in small steps, enumerating millions of conformations for monomers, and covering all build-up paths [18].

(1) Stepwise Assembly (SWA)

SWA is the implementation of "Stepwise ansatz" in Rosetta framework. The RNA loop-modeling problem is a

typical case of high-accuracy structure prediction, which is challenging. In order to verify the SWA algorithm, we applied the SWA on a benchmark consisting of 15 single-stranded loops, and the results show that SWA is valid in all tests and SWA is obviously better than FARFAR in modeling accuracy. Futhermore, blind trail is the most stringent test for the RNA structure prediction algorithm. We therefore attempt a blind high-accuracy RNA structure modeling and test the model by a chemical mapping experiment. Finally, we draw a conclusion that SWA is an ab initio build-up and enumerative algorithm, and the overall performance of this algorithm is superior to the existing knowledge mining-based methods for RNA tertiary structure prediction.

For SWA method, the ability of conformational sampling

is no longer its bottleneck, and the inaccuracies of Rosetta all-atom energy function affects its accuracy.

(2) Stepwise Monte Carlo (SWM)

As we all know that the accurate prediction of noncanonical base pairs is the key to ameliorate the accuracy of RNA modeling. SWM could predict the noncanonical base pairs of complex RNA structures [19]. The algorithm randomly performs the add or delete moves which are guided by the Rosetta all-atom free energy function, selecting a random position on which to prepend a new nucleotide, rather than enumerating all of the additions at all possible positions as was implemented in the SWA algorithm. SWM procedure is shown in Figure 3.
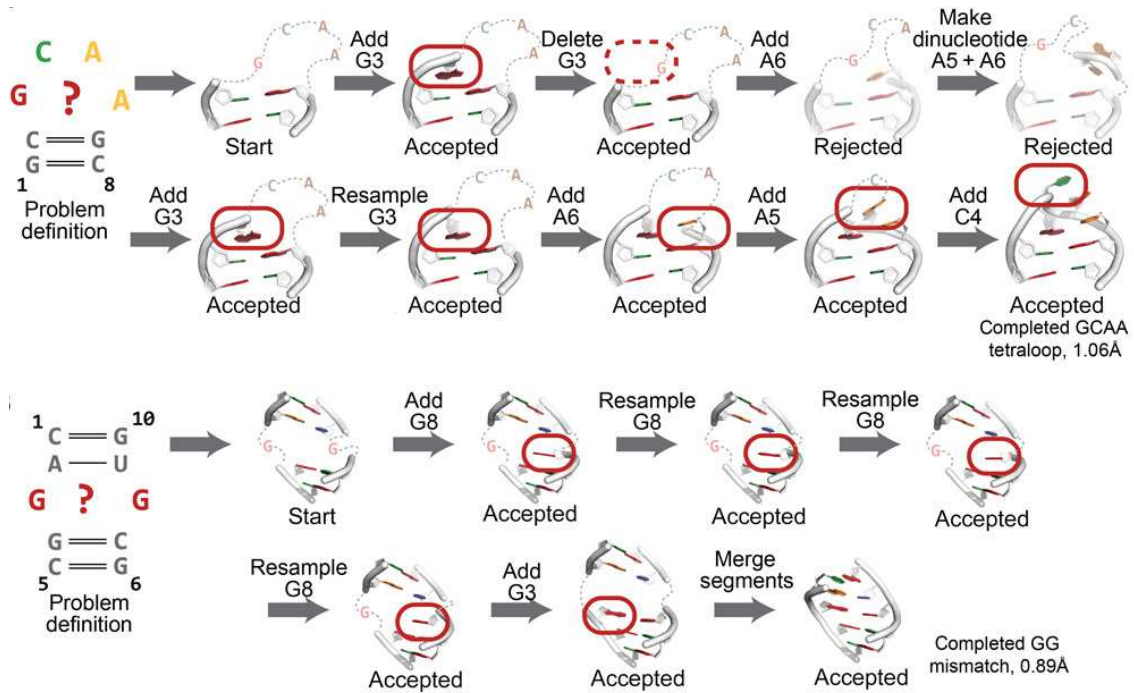

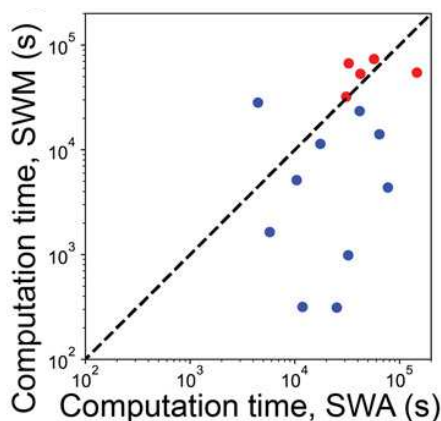**Figure 3.** *SWM procedure.*

Figure 4. *SWM method is compared with the SWA method based.*

A series of tests have been carried out on this method. First, the algorithm efficiently traverses the minimum of the energy landscape, allowing the ab initio recovery of a set of

15 single-stranded RNA loops, which proves that the lastest updates of Rosetta energy function improved modeling accuracy of single-stranded RNA loops. In addition, the SWM algorithm is compared with the SWA algorithm based on the benchmark, and the results are shown in Figure 4. Compared with SWA, SWM algorithm needs less CPU time to converge on the premise of ensuring the modeling accuracy.

Second, on a larger benchmark consisting of 82 complex and multi-strand RNA motifs, experimental results show that SWM can effectively recover complex noncanonical pairs.

Third, we applied SWM algorithm to three tetraloop/ receptors with unsolved structures, and prospectively validated these models through chemical mapping experiment.

Last, SWM solved a recent RNA-Puzzle and successfully achieved blind prediction of all noncanonical pairs of the Puzzle. As shown in Figure 4, the left figure in Figure 5 shows the modeling result of RNA Puzzles 18 using SWM

algorithm, and the right figure shows the subsequently released crystal structure diagram, which are basically the same. That is to say, SWM successfully achieved blind prediction of all noncanonical pairs of Puzzle 18. RNA-Puzzles was a group blinded experiment similar to CASP for the evaluation of RNA tertiary (3D) structure prediction to identify capacity and bottlenecks in RNA prediction problems.

These results indicate that step-by-step nucleotide structure formation is the principle of high-resolution RNA structure prediction algorithm. And SWM algorithm could greatly improve the calculation speed of ab initio structure prediction.

4. Conclusion and Perspective

RNA tertiary structure prediction is in the primary development stage at present. Modeling the complex structures of RNA and other molecules is more challenging due to the limited number of structures found in available experiments and the lack of data on complex interactions between RNA and other molecules. There is no perfect algorithm to solve the problem of RNA tertiary structure prediction. In this paper, physical-based RNA tertiary structure prediction algorithms is subdivided according to the conformation sampling method.

The bottleneck of RNA tertiary structure prediction algorithm based on physical fragment assembly is the lack of conformational sampling ability. For example, the assumption of FARFAR method limits its conformational sampling ability. RNA tertiary structure prediction algorithm based on "Stepwise ansatz" hypothesis is to add one residues at a time rather than directly listing all possible conformations of RNA, nor through the low-resolution Coarse-grading or through small perturbations to fully build conformations, which is a great progress in the field of RNA tertiary structure prediction. The RNA tertiary structure prediction algorithm based on Stepwise ansatz conformational sampling method achieves efficient conformational sampling and overcomes the problem of poor conformational sampling ability compared with other prediction algorithms.

The main factor limiting the high-precision modeling of the current tertiary structure prediction algorithm is the inaccuracies of Rosetta all-atom energy function. With the advancement of physicochemical technology, as shown in Table 2, the energy function is no longer limited to the Rosetta energy function for protein, and more energy parameters specific to RNA can be found. The discovery of these parameters is conducive to further improving the accuracy of the energy function and the modeling accuracy [20]. In particular, the test on the hepatitis C virus internal ribosome entry site indicates that a modified torational potential may address the problem that low modeling precision which is lead by energy functions, such as the modification of rna_torsion, rna_sugar_close and other RNA torsion terms.
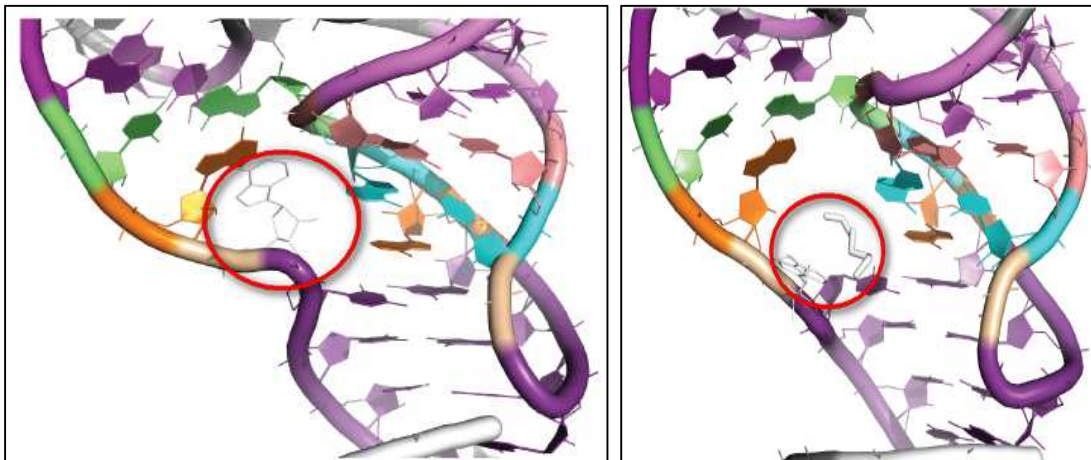

Figure 5. SWM blind (left) and Crystallographic (right).

Table 2. Energy terms for biomolecules.

Biomolecule	Term	Biological Process	Unit
Proteins	fa_atr	Attractive energy between two atoms on different residues separated by distance, d	kcal/mol
	fa_rep	Repulsive energy between two atoms on different residues separated by distance, d	kcal/mol
	fa_intra_rep	Repulsive energy between two atoms on the same residue, separated by distance, d	kcal/mol
	fa_elec	Energy of interaction between two non-bonded charged atoms separated by distance, d	kcal/mol
	hbond_sc	Energy of side chain to side chain hydrogen bonds	kcal/mol
	fa_dun	Probability that a chosen rotamer is native-like given backbone Ø angles	kT
RNA	fa_stack	π-π stacking energy for RNA bases	kT
	stack_elec	Electrostatic energy for stacked RNA bases	kT
	fa_elec_rna_phos	Electrostatic energy (fa_elec) between RNA phosphate atoms	kT
	rna_torsion	Knowledge-based torsional potential for RNA	kT
	rna_sugar_close	Penalty for opening an RNA sugar	kT

All in all, The function of RNA is highly dependent on its 3D structure, so the RNA tertiary structure prediction is very important. Before the RNA structure prediction has reached the level of atomic accuracy, this problem is still concerned by people. With the improvement of the algorithm, especially the emergence of SWM and other machine learning methods makes the modeling accuracy and speed have been improved, RNA tertiary structure prediction algorithm will be able to better provide the basis for RNA function research.

Acknowledgements

References

[1] Xiang J. Non-coding RNA Tertiary Structure Prediction and Analysis [D]. Wuhan: Huazhong University of Science & Technology, 2018: 1-6.

[2] Wang Sh, Li X. Structure and Regulation Mechanism of Riboseswitch [J]. Biotechnology Bulletin, 2010 (5): 16–22.

[3] Shi Y, Wu Y, et al. RNA Structure Prediction: Progress and Perspective [J]. Chinese Physics B, 2014, 23 (07): 92-101.

[4] Fürtig Boris, Richter Christian, et al. NMR Spectroscopy of RNA [J]. ChemBioChem, 2003, 4 (10): 936-62.

[5] Kong Q. Research on RNA fold structure prediction algorithm based on basin hopping graph with pseudoknots [D]. Jinan: Shandong Jianzhu University, 2019: 6-14.

[6] Jin L. Coarse-grained modeling for double-stranded RNA 3D structure and thermal stability [D]. Wuhan: Wuhan University, 2019: 2-17.

[7] Wang Y, Assessment of tertiary structure and statistical analysis of torsion angles of non-coding RNA [D]. Wuhan: Huazhong University of Science & Technology, 2013: 2-4.

[8] Leontis Neocles B, Stombaugh Jesse, Westhof Eric. The non-Watson-Crick base pairs and their associated isostericity matrices [J]. Nucleic acids research, 2002, 30 (16): 3497-531.

[9] Liu Zh. Studies on algorithms for RNA structure prediction including pseudoknots [D]. Jinan: Shandong University, 2014: 1-7.

[10] Massire. C, E. Westhof. MANIP: An interactive tool for modelling RNA [J]. Journal of Molecular Graphics and Modelling, 1998, 16 (4): 197-205.

[11] Bruce A Shapiro, Yaroslava G Yingling, et al. Bridging the gap in RNA structure prediction [J]. 2007, 17 (2): 157-165.

[12] Samuel Coulbourn Flores, Russ B. Altman. Turning limited experimental information into 3D models of RNA [J]. Rna-a Publication of the Rna Society, 2010, 16 (9): 1769-1778.

[13] Rother M, Rother K, et al. ModeRNA: a tool for comparative modeling of RNA 3D structure [J]. Nucleic Acid Res, 2011, 39 (10): 4007-22.

[14] Wang J. Prediction of nucleic acid molecular structure [D]. Huazhong University of Science and Technology, 2017: 10-26.

[15] Li J. RNA Tertiary Structure Prediction [D]. Nanjing: Nanjing University, 2017: 7-14.

[16] Rhiju Das, David Baker. Automated de novo Prediction of Native-Like RNA Tertiary Structures [J]. Proceedings of the National Academy of Sciences of the United States of America, 2007, 104 (37): 14664-14669.

[17] Rhiju Das, John Karanicolas, David Baker. Atomic accuracy in predicting and designing noncanonical RNA structure [J]. Nature Methods, 2010, 7 (4): 291-294.

[18] Parin Sripakdeevong, Wipapat Kladwang, Rhiju Das. An enumerative stepwise ansatz enables atomic-accuracy RNA loop modeling [J]. Proceedings of the National Academy of Sciences of the United States of America, 2011, 108 (51): 20573-20578.

[19] Andrew M. Watkins, Caleb Geniesse, et al. Blind prediction of noncanonical RNA structure at atomic accuracy [J]. Science Advances, 2018, 4 (5): eaar 5316.

[20] Alford Rebecca F, Leaver-Fay Andrew, et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design [J]. Journal of chemical theory and computation, 2017, 13 (6): 3031-3048.