

ORIGINAL RESEARCH

**OPEN ACCESS**

Full open access to this and thousands of other papers at <http://www.la-press.com>.

## A Comparative Taxonomy of Parallel Algorithms for RNA Secondary Structure Prediction

Ra'ed M. Al-Khatib, Rosni Abdullah and Nur'Aini Abdul Rashid

The Parallel and Distributed Computing Center (PDCC), School of Computer Sciences, University Sains Malaysia, 11800 Penang, Malaysia. Email: [rmaak.cod09@student.usm.my](mailto:rmaak.cod09@student.usm.my)

---

**Abstract:** RNA molecules have been discovered playing crucial roles in numerous biological and medical procedures and processes. RNA structures determination have become a major problem in the biology context. Recently, computer scientists have empowered the biologists with RNA secondary structures that ease an understanding of the RNA functions and roles. Detecting RNA secondary structure is an NP-hard problem, especially in pseudoknotted RNA structures. The detection process is also time-consuming; as a result, an alternative approach such as using parallel architectures is a desirable option. The main goal in this paper is to do an intensive investigation of parallel methods used in the literature to solve the demanding issues, related to the RNA secondary structure prediction methods. Then, we introduce a new taxonomy for the parallel RNA folding methods. Based on this proposed taxonomy, a systematic and scientific comparison is performed among these existing methods.

**Keywords:** RNA secondary structure, dynamic programming (DP), pseudoknot, free energy minimization, FPGA, GPU

---

*Evolutionary Bioinformatics* 2010:6 27–45

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



## Introduction

Bioinformatics is a new discipline resulting from the combination of two science fields: Computer Science, and Biology. This new scientific application has grown rapidly and nowadays it is becoming a cornerstone for each molecular biological study.<sup>1</sup> It utilizes computer implementations and algorithms for collecting, accumulating, storing, analyzing, and integrating biological data and genetic macromolecules like deoxyribonucleic acid (DNA), ribonucleic acids (RNA), or proteins. The DNA has directions on how to build other components of cells, such as proteins and RNA molecules. The RNA is a type of nucleic acid that provides a mechanism to copy the genetic information of the DNA. There are two different types of RNA which are the Messenger RNA (mRNA) and the Transfer RNA (tRNA). They play important roles in the living cells and protein synthesis. Recently, researchers have found that they can use the RNA interference (RNAi) process<sup>2,3</sup> for producing modern drugs. Mainly, this could be used in the therapeutics process of discovering antiviral drugs for difficult diseases like Cancer, AIDs, and Herpes.<sup>4,5</sup>

It is important to study the folding of RNA molecules to understand their roles and functionalities.<sup>6</sup> Medical researchers and biologists can find different vital roles of RNA molecules by scrutinizing their RNA secondary structures. This will pave the way in front of biomedical researchers to utilize the RNA molecules in useful application like when there are used for productive treatments.<sup>4</sup> The experimental methods that are commonly and broadly used for determining three-dimensional (3D) structures of RNA are listed as follows: X-ray crystallography and Nuclear Magnetic Resonance (NMR). In a biological context, these experimental methods are the most prominent accurate methods to determine the RNA tertiary structure, which is a 3D structure. But, both of these physical methods are time-consuming, expensive and computationally difficult to accomplish. To give the reasons, which make these purification techniques for determining the RNA 3D structures are tedious and difficult. In the rest of this section, we elaborated more about how both of these experimental methods are carried out by chemical or biological researchers.

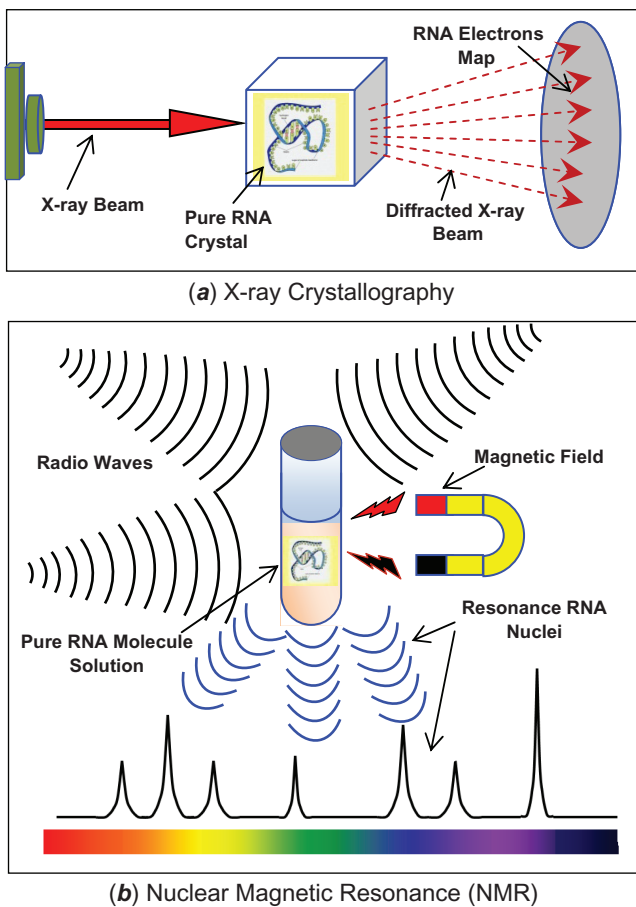
- *X-ray crystallography*. Primarily, it is the faster and popular purification experimental methods,

to determine the RNA tertiary structures. The researchers fixed a pure crystal of single RNA molecule. They needed to obtain sufficient amounts of pure RNA crystal, a milligram of RNA quantity is necessary,<sup>7</sup> which is complex and non-trivial to acquire. Next, the biologist will bombard this RNA crystal with X-ray, which is depicted in Figure 1 (a) as an adapted from.<sup>8</sup> The X-ray beams will collide with the fixed and stable RNA crystal and will go through it. The X-ray will diffract whilst colliding with the electrons that are situated around the nuclei of RNA molecule. Thus, the result is the map for the electrons of the RNA molecule, which gives the nearest accurate form of the folding RNA molecule.

- *Nuclear Magnetic Resonance (NMR)*. In this NMR experimental method,<sup>7</sup> the charged RNA molecule is fixed via magnetizing the RNA molecule, as shown in Figure 1 (b). The magnetic field will stabilize the RNA sample. The phosphate groups within the backbone of the RNA have a negative charge which causes the solution of the RNA molecule to be charged.<sup>9</sup> Then, the fixed molecule will be bombarded by using radio waves. This bombarding process will cause the resonance of the RNA nuclei. The 3D structure of the RNA molecule will be constructed from the resonating of these bombarded nuclei. However, this bombarding process needs to be performed from thousands of different angles. Therefore, this makes the NMR physical method incredibly time-consuming, costly and tedious.

Due to these constraints and difficulties from the experimental physical sides, computational methods from computer scientists and bioinformatic researchers have become more demanding to do the RNA structure prediction process. Predicting the RNA 3D structure from the primary sequence is difficult to accomplish. Hence, the bioinformatic researchers first detect and predict the RNA secondary structure from a given RNA primary sequence. Then, the result of this RNA secondary prediction process will assist the biologists to determine the RNA tertiary structure.

Thus, solving the RNA secondary structure problem is becoming a main issue among bioinformatic researchers.<sup>10</sup> Recent approaches and current work concentrate on applying the parallel techniques to



**Figure 1.** Experimental Methods of 3D RNA Structures Determination: (a) X-ray Crystallography sequence. (b) Nuclear Magnetic Resonance (NMR). (The idea adapted from<sup>7,8</sup>).

the previous RNA computational algorithms. In this paper, the works of the researchers explore the state-of-the-art of parallel techniques were proposed for solving RNA structural problems.

The earliest prediction methods for solving the RNA secondary structure problems were presented by Waterman and Smith<sup>11</sup> and Nussinov et al.<sup>12</sup> These computational algorithms proposed two different methods for the RNA/DNA folding structure, which require  $O(n^3)$  time complexity. Later, several stable algorithms for the RNA secondary structures were proposed like Zuker's algorithm.<sup>13</sup> It was based on a thermodynamic energy minimization model. The execution time is still also  $O(n^3)$ , where  $n$  is the length of the RNA sequences, which are implemented in the *ViennaRNA package*.<sup>14</sup>

Basically, most computational algorithms, approaches and methods for solving the RNA secondary structures problems were restricted to the length of the bases, which

were only a few hundreds of characters. Computational methods become much more desirable to solve traditional RNA cases and original RNA algorithms.<sup>15</sup> Then subsequently, promising passageway simplifies these constraints and solves the bottleneck for the original RNA computational prediction algorithms, by implementing them on new parallelization approaches. Some of the more parallel popular approaches to reach this objective are to design them on the Field Programmable Gate-Array (FPGA), the Graphics Processing Unit (GPU), the multi-core or the cluster master-slave parallel architecture systems.<sup>16</sup>

### Schematic Study

We begin this comparative taxonomy study for the RNA parallel prediction methods by compromising and organizing this paper into three main research issues; (i) literature and background of the RNA predictions, (ii) RNA research models, (iii) the discussion and evaluation of existing methods with a comparative taxonomy. The main objective of this research is to present a comprehensive summary of the state-of-the-art researches on the RNA secondary structure predictions methods. Thus, we combined and joined among these all parts as follows:

- *The First part* is on the literature and background of the RNA prediction work. It explains the cause of the formation of folds and helices in each RNA molecule, in the chemical context. Next, the authors elucidated the RNA primary and the RNA secondary structure as a base step to clarify the RNA problem domain and the research statement on the computer side.
- *The next part* was for the RNA research models and methods. The researchers presented the existing RNA methods including the sequential and the parallel structural predictions. Then, the researchers made a comparison among either of the group. Lastly, in the discussion and comparison section, the authors highlighted the results of this study. In the discussion, the researchers concluded with the most suitable efficient paradigm that could be used for parallelizing the RNA prediction algorithms. Also, the reasons that guided to choose this decision were given.
- Finally, there is the discussion and the evaluation of existing methods with a comparative taxonomy.



The researchers proposed and described a general taxonomy for RNA parallel methods. By using this taxonomy, they applied a comparative process for the previous existing parallel RNA prediction methods. This process enabled the researchers in this field to distinguish among the parallel RNA prediction approaches. Future researchers will be able to choose the suitable parallel paradigm that will fit their interests. This selection will depend on the group of the RNA molecules that the bioinformatic researchers working on or needs.

*Roadmap of the study:* The following section highlights the elementary RNA chemical structure and the RNA (*Primary* and *Secondary*) structures. Next, the authors presented the fundamental concept for the RNA secondary structure problem statement. Then, they listed and investigated the sequential RNA prediction methods and categorized the RNA parallel techniques based on the parallelization hardware architecture which are listed as follows: (i) An individual RNA parallel algorithm<sup>15</sup> on a multi-core Central Processing Unit (multi-core CPU). (ii) A one solitary RNA parallel method on the GPU.<sup>16</sup> (iii) One work of the RNA parallel algorithm on the Beowulf cluster.<sup>17</sup> (iv) Three different parallel RNA algorithms were implemented on the FPGA.<sup>18–20</sup> To complete the categorization, the researchers explained a suggested parallel taxonomy for RNA folding algorithms. Then, they systematically applied this proposed taxonomy on the mentioned RNA parallel methods, comparatively. Next, they discussed and compared the results for the main parallel RNA methods. Finally, the authors gave some concluding remarks. Also, they introduced some recommendations for bioinformatics researchers. These hints could be used as a constructive future research in the RNA structural prediction domain.

## Background

The RNA molecules had been confirmed to be very resourceful materials by.<sup>3,4</sup> The RNA molecules play different crucial functions and roles in living organisms and in many biological processes. Recently, it became clear that the RNA molecules play various roles, not just an intermediate in protein synthesis. But also, the latest researches are looking to utilize the RNAi process to discover new medications as a treatment for dangerous virus diseases.<sup>4,21</sup>

## RNA chemical structure

To understand the roles and functions of the RNA molecules, the biomedical engineers and researchers need to determine and scrutinize the RNA tertiary structures. As the first step to reach this goal, they ought to know the RNA molecules chemical structure and the motives that forced them to make these folds of 3D structures. Accordingly, the RNA was the common name for ribonucleic acids; they are made of long chains (single-stranded) of nucleotide units. There are three different components of the RNA nucleotides: the nitrogenous base, the sugar, and the phosphate group. While, the RNA backbone is made up of ribose five atom carbon-sugar counted from 1' through 5' and it is attached by two phosphate groups in 3' and 5', respectively. From the sugar group "*ribose*", the RNA molecule acquired its nickname "Ribonucleic Acids". The nitrogenous bases in the RNA group were made up of four different bases Adenine (A), Cytosine (C), Guanine (G), and Uracil (U). These bases attached to the five-carbon sugar in 1' position and they give RNA molecules characteristic possessions and properties.

Veritably, the phosphate groups in the backbone of the RNA have a negative charge, which makes the RNA a charged molecule.<sup>9</sup> Due to this, the charged RNA molecule inside the living cells is not stable. Thus to gain more stability, some parts of the single-stranded RNA fold back on itself forming double helices. This RNA folding process makes the determination methods intricate and not easy to determine the RNA 3D structures. The details about the RNA molecules and its chemical structure was explained in.<sup>22</sup>

## The RNA primary and secondary structure

Both the phosphate groups are attached to the 3' and 5' positions from ribose sugar in the backbone of RNA molecules. Due to this fact and based on the convention and general agreement among biologists, the RNA primary structure is a string series of bases reported from the 5' end to the 3' end, as shown in Figure 3(a).

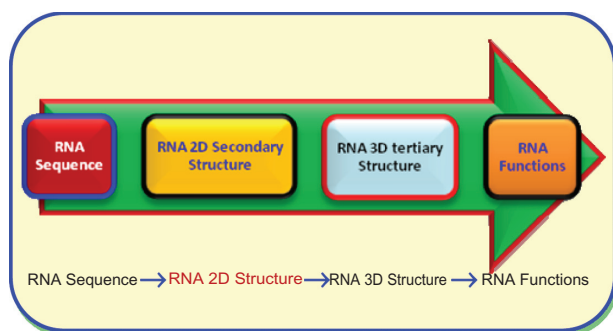
The RNA secondary structure derives from the pairing up of these four nucleotides according to the rules of *Watson-Crick* and *Wobble* "*WW*": *Watson-Crick*

base pairs ( $G \equiv C$ ) and ( $A = U$ ) and a *Wobble* base pair ( $G-U$ ). By applying these rules, the single-stranded RNA secondary structure forms two large groups: [Stem-loops and Pseudoknots (PK)], as shown in Figure 3 (b and d).

## Problem Statement

Recently, RNA molecules have been confirmed to be very resourceful materials in the medical process and biological systems.<sup>23</sup> Biologists can determine the RNA tertiary structure from its secondary structure. The biologists need this 3D structure of the RNA molecule to derive its function and essential role. Thus, the bioinformatic researchers introduced computational prediction methods, which can predict the RNA secondary structures from a given RNA primary sequence. In Figure 2, the authors depicted and explained the vital crucial position of the RNA secondary structure in the chain of the RNA research. Also, this figure showed the flow chain of the RNA research study, consecutively. Moreover, it confirmed that the computational secondary structure prediction processes is a required preliminarily step for determining the RNA 3D structure.

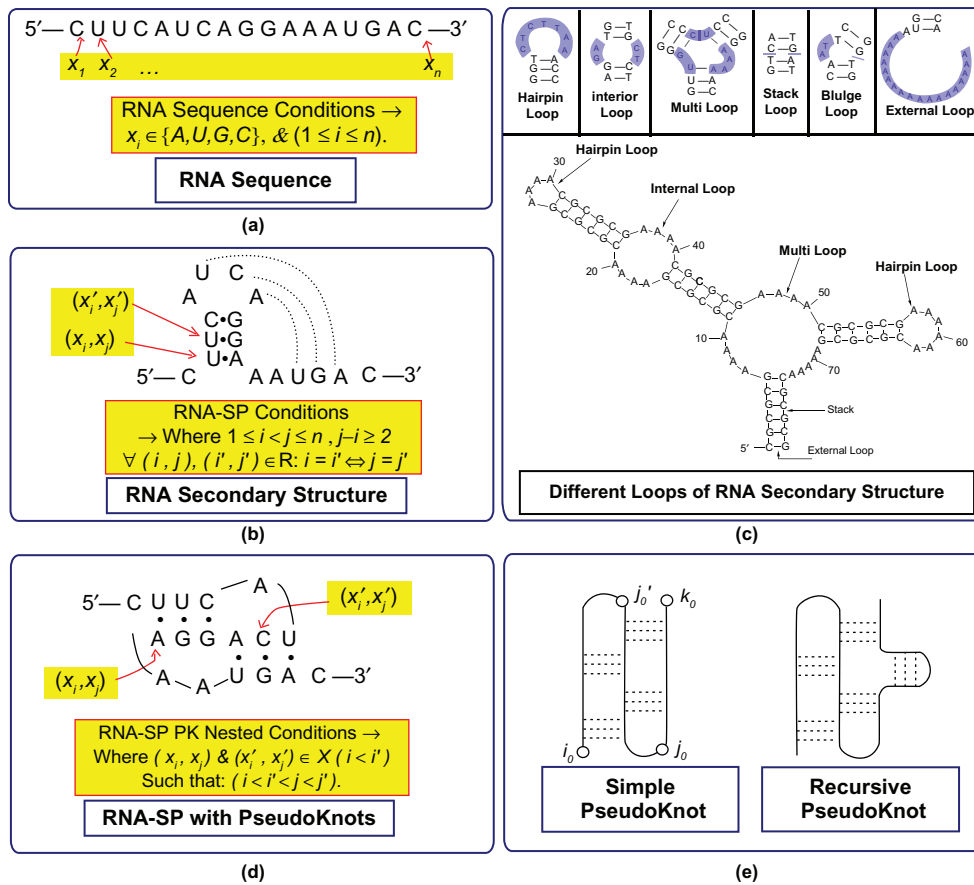
The RNA folding recognition methods attempt to predict an accurate and more stable RNA folding structure based on the Minimum Free Energy (MFE) models. As shown in Figure 3 (d), the nature of some types of the RNA structure forms the pseudoknots, in some parts. The RNA molecules with the pseudoknots structure make the calculation process of the RNA secondary structure prediction algorithms complex. These complexities of the RNA prediction algorithms conform and confirm the execution time and memory storage space, computationally.



**Figure 2.** The main systematic chain steps of RNA research study.

The components of the RNA involved in understanding the RNA functions, which are extracted from,<sup>24</sup> and can be presented as follows:

- RNA primary sequence structure is a string of  $n$  characters,  $x_i = x_1x_2\dots x_n$  where  $x_i \in \{A \text{ or } a, C \text{ or } c, G \text{ or } g, U \text{ or } u\}$  the four bases in uppercases or lowercases letters, as well  $1 \leq i \leq n$ , as seen in the Figure 3 (a).
- A single-stranded RNA secondary structure is a list of base-pairs which can be viewed as a set of,  $X$ , forms on acceptable base pairs  $(x_i, x_j)$ . These pair of letters is called a complementary base pair, according to the “*WW*” rules, in Figure 3 (b):
  - These “*WW*” rules are:  $(x_i, x_j) = (a, u)$  or  $(A, U)$ ,  $(x_i, x_j) = (g, c)$  or  $(G, C)$  for the Watson-Crick rule. Also,  $(x_i, x_j) = (g, u)$  or  $(G, U)$  along with the Wobble rule later.
  - In addition, these base pairs  $(x_i, x_j)$  for RNA secondary structure where that:
    - It should be at *first*,  $1 \leq i < j \leq n$ .
    - *Second*,  $j - i > t$  where  $t$  is a small constant, i.e.  $j - i \geq 2$ .
- For all base pairs  $(x_i, x_j)$  and  $(x'_i, x'_j)$  in  $X$ ,  $i = i'$ , if and only if  $j = j'$ , (such that  $\forall (i, j), (i', j') \in X: i = i' \Leftrightarrow j = j'$ ) as depicted in Figure 3 (b).
- Thus, namely two bases that form the canonical pair must be located at different locations. While, the RNA sequence does not fold back on itself too sharply. Also, each base can be paired and combined at most only with another base. Restrictedly, the implementers allowed just “*WW*” the canonical RNA base pairs rules<sup>26</sup>:  $\{(A,U), (C,G), (G,U)\}$ . The RNA secondary structures without pseudoknots folding in different kinds of loops,<sup>15</sup> which are: Hairpin Loops, Internal Loops, Multiloops, Stacks, Bulges and External Loops, as depicted in Figure 3 (c).
- *The RNA secondary structure with pseudoknots*: this is defined as,  $X$ , if and only if the base pairs exist in nested condition  $(x_i, x_j), (x'_i, x'_j) \in X (i < i' < j < j')$ , as seen in the Figure 3 (d). Whereas the RNA secondary structures with pseudoknots coming in two main groups of pseudoknotted RNA<sup>10</sup>: *simple* or *recursive*, those depicted in Figure 3 (e). Thus, a given RNA sequence  $X$ , can fold with a maximum number of base pairs and in an exponential number of possible structures.



**Figure 3.** RNA Structures:- (a) RNA sequence. (b) RNA secondary structure. (c) RNA Stem-Loops Structure. (d) RNA PseudoKnots. (e) PseudoKnots Types [Simple and Recursive], some parts adapted from.<sup>10,15,25</sup>

### Thermodynamic algorithm for RNA prediction

From the energy stability point of view, the phosphate groups in the backbone of RNA molecule have a negative charge. Thus, the RNA molecules inside the living cells are not stable.<sup>9,27</sup> They will fold back on themselves to reach more stability.

Then, the main goal of the RNA secondary structure prediction computational method is to arrive at more stable equilibrium of the RNA folding form, based on the free-energy model. Hence, to calculate the RNA free energy stabilities, it is necessary to predict RNA secondary structure by calculating the MFE, which is named as an optimal RNA structure.<sup>28</sup>

The summation of energies for all loops is the energy of RNA secondary structure (Eq. 1), which extracted with condensation from.<sup>18,28,29</sup> The empirical calculations explained and confirmed that over 99% from the execution time for the RNA prediction algorithm is in computing the MFE.<sup>20</sup>

$$\text{Total MFE for RNA} = \sum \text{of RNA Loops Energies} \quad (1)$$

(At fixed temperature + ionic concentration)

In reality, to the best knowledge of the authors, thermodynamic prediction approach for calculating the energy of the optimization RNA, had been expressed and introduced for the most time by Lyngso *et al* in.<sup>28</sup> They were used a four different arrays to hold and include the MFE lookup tables inside the shared cache memory during the execution time of the prediction algorithm.

These MFE calculative motifs complicate the general pseudoknotted RNA secondary structure prediction algorithms. The pseudoknots type turns the RNA prediction algorithms to be *NP-Complete* Problem.<sup>10,30</sup> In addition, the algorithms for solving the pseudoknotted RNA secondary structure problem, need to allow energy functions to operate and run in the worst case polynomial time. In fact, two researchers Akutsu in<sup>10</sup> and Lyngso and Pedersen in<sup>30</sup> proved that finding the pseudoknotted RNA structure with the MFE

is the *NP-hard* problem, particularly by applying the standard nearest-neighbour energy function.

Consequently, researchers of pseudoknotted RNAs faced with three problems: *First*, the RNA secondary structure prediction with the pseudoknots is computationally intricate and difficult to carry out.<sup>30</sup> Considering the execution time and the memory space complexities, made the problem an *NP-complex* problem.<sup>30</sup> Besides that, most professional algorithms exist only for partial classes of pseudoknots.<sup>10,25,31,32</sup> Namely there were restricted in a subclass and not for all classes of pseudoknotted RNA.<sup>33</sup> *Second*, almost majority of RNAs computational methods have analyzed nested “Stem-Loops” RNA secondary structure prediction,<sup>34</sup> either by neglecting the RNA pseudoknots for simplicity, or not being aware of the existence of the pseudoknotted RNA types. *Lastly*, most of the existing methods for RNA structure prediction are not acceptable at levels of accuracy, reliability and robustness.<sup>33</sup>

### Existing Sequential Methods

Basically, one of the important tasks in front of bioinformatics and computer application researchers is the RNA secondary structure prediction dilemma. From a biological point of view, there are various complexities and constraints that are faced the biomedical researchers in experimental methods to determine the RNA tertiary structures.<sup>7</sup> Recently, many computational efforts have been presented on the computer side, to detect the RNA secondary structure from the primary sequence. The predictive approaches have been introduced to solve the related issues in the RNA structural detection field as follows: (i) Energy thermodynamic models to predict RNA secondary structures; i.e. Mfold,<sup>13</sup> RNAfold<sup>14</sup> and RNAalifold.<sup>35</sup> (ii) Comparative analysis methods to predict the RNA secondary structures from multiple homologous sequence alignment.<sup>36</sup> (iii) Stochastic Context-Free Grammar methods (SCFS) are a comparative sequence analysis for prediction consensus of the RNA secondary structures from multiple sequence alignment.<sup>37</sup>

The original existing methods that have been proposed in solving the RNA secondary structures are divided into two major groups. The first group predicts the *non-pseudoknotted* RNA secondary structures “*Stem-Loop of RNA*”, while the other group

is solving the RNA secondary structure prediction problem including pseudoknots type “*RNA with Pseudoknots*”.

Then, the researchers in this paper classified these sequential RNA prediction methods in a schematic diagram, as shown in Figure 4. Next, the researchers converged these RNA prediction methods, as shown in Table 1, which was adapted and extracted from.<sup>22</sup> This table included the most well-known existing RNA sequential prediction methods and approaches that have been produced lately to predict RNA secondary structures.

### Parallel Methods and Schemes

The experimental methods are completely accurate for determining the RNA 3D folding structures. But due to their time consuming and expensive nature, many computational approaches have been proposed to predict the RNA secondary structure, which includes: (i) RNA prediction methods based on a thermodynamic energy minimization model. (ii) RNA structural comparative approaches from

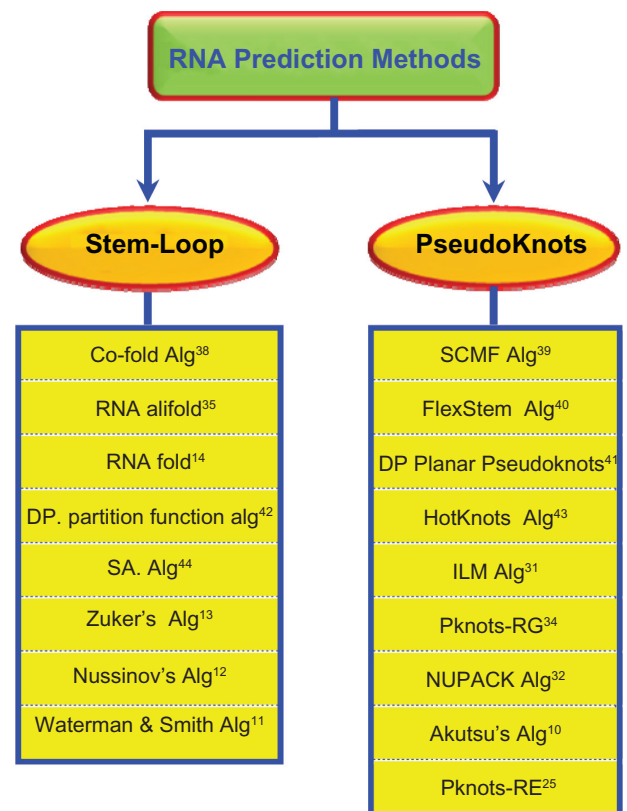


Figure 4. Schematic diagram of RNA structural prediction methods.

**Table 1.** Existing sequential methods for RNA secondary structure prediction.

| No. | RNA prediction method | Reference                       | Method complexities |                   | Major contribution  | RNA types   |
|-----|-----------------------|---------------------------------|---------------------|-------------------|---|-------------|
|     |                       |                                 | Execution time      | Space requirement |   |             |
| 1.  | SCMF Alg.             | Jens and Andrew <sup>39</sup>   | $O(m n^2)$          | $O(n^2)$          | A near optimal algorithm to predict RNA secondary structure with pseudoknots.<br>Where $n$ : is the RNA sequence length. And $m$ : is the number of iteration steps with the $n$ bases.   | Pseudoknots |
| 2.  | FlexStem Alg.         | Chen et al <sup>40</sup>        | $O(n^4)$            | $O(n^2)$          | A prediction algorithm named for RNA secondary structures, it adapted a comprehensive energy models for complex pseudoknots type.<br>Where $n$ : is the RNA sequence length.              | Pseudoknots |
| 3.  | Co-fold Alg.          | Ziv-Ukelson et al <sup>38</sup> | $O(n^4 \zeta(n))$   | –                 | An optimal alignment alg. to predict RNA secondary structures based on Sankoff's Alg. <sup>44</sup><br>Where $n$ : is the RNA sequence length.<br>And $\zeta(n)$ can converge to $O(n)$ . | Stem-Loops  |
| 4.  | DP Planar Pseudoknots | Hengwu et al <sup>41</sup>      | $O(n^4)$            | $O(n^3)$          | A DP algorithm to predict RNA secondary structures with arbitrary planar and simple non-planar pseudoknots type by using MFE model.<br>Where $n$ : is the RNA sequence length.            | Pseudoknots |
| 5.  | HotKnots Alg.         | Ren et al <sup>43</sup>         | $O(n^4)$            | $O(n^2)$          | A heuristic algorithm to predict pseudoknotted RNA based on MFE.<br>Where $n$ : is the RNA sequence length.   | Pseudoknots |
| 6.  | ILM Alg.              | Ruan et al <sup>31</sup>        | $O(n^4)$            | $O(n^2)$          | A heuristic algorithm for predicting pseudoknotted RNA based on MFE or comparative or both.<br>Where $n$ : is the RNA sequence length.  | Pseudoknots |
| 7.  | Pknots-RG             | Reeder et al <sup>34</sup>      | $O(n^4)$            | $O(n^2)$          | A DP algorithm to predict optimal RNA secondary structures by using MFE model.<br>Where $n$ : is the RNA sequence length.   | Pseudoknots |
| 8.  | NUPACK Alg.           | Dirks and Pierce <sup>32</sup>  | $O(n^5)$            | $O(n^4)$          | A DP algorithm to predict base-pairing probabilities of RNA with pseudoknots based on a partition function and MFE.<br>Where $n$ : is the RNA sequence length.                            | Pseudoknots |

(Continued)





Table 1. (Continued)

| No. | RNA prediction method       | Reference                        | Method complexities                                |   | Major contribution   | RNA types   |
|-----|-----------------------------|----------------------------------|--|---|--|-------------|
|     |                             |                                  | Execution time                                     | Space requirement                                 |  |             |
| 9.  | RNAalifold                  | Hofacker et al <sup>35</sup>     | $O(m \times n^4 + n^3)$                            | $O(n^2)$  | The algorithm computes the consensus RNA secondary structures from multiple alignments with modifying energy models. Where $n$ : is the RNA sequence length. And $m$ : is the number of sequences alignments.                                      | Stem-Loops  |
| 10. | Akutsu's Alg.               | Akutsu <sup>10</sup>             | $O(n^4)$   | $O(n^3)$ enhanced by <sup>30</sup> to be $O(n^2)$ | A simple DP algorithm to predict RNA secondary structure with pseudoknots. Where $n$ : is the RNA sequence length.   | Pseudoknots |
| 11. | Pknots-RE                   | Rivas and Eddy <sup>25</sup>     | $O(n^6)$   | $O(n^4)$  | An adaption of DP algorithm for predicting a tractable subclass of pseudoknotted RNA based on complex MFE model. Where $n$ : is the RNA sequence length.   | Pseudoknots |
| 12. | RNAfold                     | Hofacker et al <sup>14</sup>     | $O(n^3)$   | $O(n^2)$  | An implementing of Zuker's RNA prediction alg. <sup>13</sup> based on MFE model with employing thermodynamic parameters of. <sup>29</sup> Where $n$ : is the RNA sequence length.  | Stem-Loops  |
| 13. | DP. partition function alg. | McCaskill <sup>42</sup>          | $O(n^3)$   | $O(n^2)$  | A DP algorithm used MFE model to predict the partition function of unpseudoknotted RNA. Where $n$ : is the RNA sequence length.  | Stem-Loops  |
| 14. | SA. Alg.                    | Sankoff <sup>44</sup>            | $O(n^6)$   | $O(n^4)$  | A DP algorithm for RNA secondary structural alignment. Where $n$ : is the RNA sequence length.   | Stem-Loops  |
| 15. | Zuker's Alg.                | Zuker and Stiegler <sup>13</sup> | $O(n^4)$ optimized by <sup>45</sup> to be $O(n^3)$ | $O(n^2)$  | An improved DP algorithm to predict RNA secondary structures from single sequence by computing MFE. It has been re-implemented by Mfold, <sup>13</sup> RNAfold <sup>14</sup> and RNAalifold. <sup>35</sup> Where $n$ : is the RNA sequence length. | Stem-Loops  |
| 16. | Nussinov's Alg.             | Nussinov et al <sup>12</sup>     | $O(n^3)$   | $O(n^2)$  | A simplest DP algorithm computes RNA secondary structure based on MFE. Where $n$ : is the RNA sequence length.   | Stem-Loops  |
| 17. | Waterman and Smith Alg.     | Waterman and Smith <sup>11</sup> | $O(n^3)$   | $O(n^2)$  | A simple DP algorithm for predicting RNA secondary structure without pseudoknots. Where $n$ : is the RNA sequence length.  | Stem-Loops  |



multiple homologous sequences. (iii) A comparative prediction consensus of the RNA structures by using SCFG methods. (iv) The Genetic Algorithm (GA) for predicting RNA structures.<sup>18</sup>

Lately, many different parallel methods were introduced, in order to face the computational complexities of the RNA secondary structure prediction problem such as: (i) The scalable program using the parallel multi-core.<sup>15</sup> (ii) An implementing computer algorithms to run on the graphics hardware like GPU.<sup>16</sup> (iii) A parallel implementation on the Beowulf cluster by using Message Passing Interface (MPI) library.<sup>17</sup> (iv) The fine-grained hardware implemented on FPGA.<sup>18–20</sup> In the remaining parts of this section, the authors discussed the existing parallel methods for predicting the RNA structure from a given primary sequence.

### Multi-core RNA parallel algorithm

Recently, the mainly accepted and accurate approaches for predicting the RNA secondary structures sequentially are *Mfold*<sup>29,46</sup> and *RNAfold*.<sup>14</sup> These two sequential approaches require the  $O(n^4)$  execution time steps and the  $O(n^2)$  spatial storage complexities. Having these complexities, the prediction of the large RNA sequences would not be feasible, particularly in the sequential implementation. Lately, one RNA research presented a parallel and scalable design “*GTfold*” in.<sup>15</sup> *GTfold* was implemented on the multi-core CPU for solving the RNA secondary structure problem. This parallel RNA prediction method integrated the *Mfold* and *RNAfold* algorithms together, in the parallel blueprint. This proposed re-implemented parallel method “*GTfold*”<sup>15</sup> obtained more accuracy in predicting RNA secondary structure. Also, it could compute the larger RNA sequences.

In fact, the value and the significance of the *GTfold* is in predicting the accuracy of the large RNA sequences. This accurate result when compared to many other existing RNA approaches such as the *Mfold*,<sup>29,46</sup> and the *RNAfold*.<sup>14</sup> Also, Mathuriya et al in<sup>15</sup> made an optimal improvements by reducing the time complexity from  $O(n^4)$  to  $O(n^3)$ . This improvement was calculated in computing the internal loop in the RNA fold, by improving the Internal Loop Speedup Algorithm “*ILSA*”. Basically, this *ILSA* enhancement enabled the execution steps of the *GTfold* algorithm, to run in a shorter time.

Therefore, the *ILSA* enhancement gave the opportunity for the *GTfold* to predict *Homo sapiens* 23S ribosomal with 5,184 nucleotides of the RNA sequence only in minutes compared to the nine hours before.<sup>15</sup> Furthermore, the *GTfold* calculated the HIV-1 viral RNA genomes in 84 seconds (two months before<sup>47</sup>); HIV-1 includes 9,781 bases of RNA nucleotides. The *GTfold* was a parallel implementation of the RNA secondary structure prediction in the 16-core dual CPU symmetric multiprocessor system on an IBM P5-570 server machine. The *GTfold* algorithm achieved a good execution time one to two times (a factor 1.6× speed-up time). This algorithm has shown enhancements in both efficiency and performance when compared to the existing sequential algorithms,<sup>14</sup> in the large sequences.

### Parallel algorithm on GPU

The Graphics Processing Units (*GPUs*) started like a specific processor for accelerating and manipulating 3D computer graphical operations and games. Fortunately, due to the GPU’s highly parallel structure, extraordinary powerful and common function computing engines, this technique open a promising way in the parallel bio-computing sector. Recently, the General Purpose of GPUs (GPGPU) also has been growing very fast. The GPUs are going to become the cornerstone for the high computational complexity algorithms,<sup>16</sup> like the pseudokotted RNA structural prediction methods, which was proved as the *NP-hard* problem.<sup>10,30</sup>

Basically, the researchers utilized the latest modern GPUs to speed-up the algorithms in solving RNA secondary structure problems. Rizk and Lavenier in<sup>16</sup> explored a new implementation for the previous function, which was used in solving RNA prediction problems. The researchers re-implemented the *hybrid-ss-min*<sup>48</sup> function on GPU. This function was used in the Unafold package to compute the MFE of the RNA folding problem. Also, the original RNA algorithm “Unafold” has a time complexity of  $O(n^3)$ . Consequently, the new parallel design<sup>16</sup> on modern GPUs hardware fulfilled more accelerated speed-up time of up to ×17. These results was comparing with the same function which run on a system with a single CPU sequentially.

The significance of the research results in,<sup>16</sup> is to obtain faster execution time in the RNA secondary



prediction algorithm. Utilizing the GPU design, reduced the execution time and the computation complexities without any extra cost. This improves the results when compared with the other competitive RNA prediction methods like the Multi-core, the clusters or the multiprocessors systems.

### Parallel framework on cluster

The computer hardware Beowulf cluster showed some strong parallel features, based on the master-slave paradigm.<sup>49</sup> Many researchers exploited this parallel architecture to compute the traditional RNA secondary structure detection methods. The parallel implementation on the Beowulf cluster was utilized for re-implementing the original RNA prediction methods.<sup>17</sup> This work pointed out that there were good results with a higher accuracy and a faster execution time in the RNA structural algorithms, comparing to the original RNA detection methods.<sup>25,34</sup> The pseudoknotted RNA complicated structure algorithms could also benefit from this parallel design. Despite those RNA pseudoknots structures were always recognized in most of the RNA molecules and they were known among RNA researchers.<sup>24</sup> But, due to that the pseudoknotted RNA molecules are computationally demanded nature. Thus, the RNA pseudoknots types mostly ignored from many RNA prediction methods. Namely, some researchers compute their RNA prediction methods without pseudoknots to get more simplicity.<sup>34</sup> These prediction methods to solve RNA secondary structures would be inaccurate when they neglected pseudoknotted types.

A prominent work has opened a new epoch in pseudoknotted RNA secondary structure prediction research. The *CompPknots*<sup>17</sup> integrated two main pseudoknotted RNA secondary structure prediction methods.<sup>25,34</sup> The researchers implemented a parallel master–slave framework between two existing RNA methods. These two existing pseudoknotted RNA prediction methods are:

- *Pknots-RE*<sup>25</sup>: an optimal method in solving pseudoknotted RNA secondary structures. It was based on applying the standard RNA thermodynamic stability of pseudoknots type. The *Pknots-RE* method requires  $O(n^6)$  in time and  $O(n^4)$  memory space, where  $n$  is the length of the RNA sequences.
- *Pknots-RG*<sup>34</sup>: It was the latest improvement on time and space complexities for the same algorithm, which was used before in the *Pknots-RE*. The authors of *Pknots-RG* utilized the MFE model, to achieve better performance on the complexity, when comparing with the *Pknots-RE*.<sup>25</sup> This new version is the *Pknots-RG*, which enhanced the execution time complexity from  $O(n^6)$  to  $O(n^4)$ . Also, *Pknots-RG* improved the storage complexity to  $O(n^2)$  from  $O(n^4)$  comparing with the *Pknots-RE* method.

The main contribution of the *CompPknots*,<sup>17</sup> was to apply a parallel calculation in bioinformatics for the pseudoknotted RNA structural methods. The authors used the MPI library with the combination of these two previous pseudoknotted RNA algorithms<sup>25,34</sup> in a parallelization design. The new paradigm enabled the researchers to predict larger RNA sequences. While, the two previous methods<sup>25,34</sup> were not able to act as efficient as the improved parallel one.<sup>17</sup> The *CompPknots* parallel design was implemented on the Beowulf cluster based on the master-slave architecture. This parallel design obtained the good results in the pseudoknotted RNA secondary structure detection with the higher accurate prediction. Also, the method ran in a shorter execution time. Moreover, the authors in<sup>17</sup> introduced a new automatic comparison approach, which allowed the end users to compare their final results directly with previous ones. This automatically comparison process takes a shorter time than the traditional manual methods that they were using tools such as the Pseudoviewer3.<sup>50</sup>

### Parallel on FPGA co-processors

Different researches found and confirmed that the traditional RNA prediction methods can be put into operation by using fine-grained hardware implemented on the FPGA. Recently, the bioinformatic researchers found that the modern computers, parallel or multi-core, do not show a greater than 50% parallel usefulness and efficiency.<sup>18</sup> But the researchers achieved better accelerations when using the FPGA co-processors, which grows to be a hopeful approach for the main RNA prediction algorithms like RNAalifold,<sup>18</sup> Nussinov's algorithm<sup>19</sup> and Zuker's algorithm.<sup>20</sup>



### (i) Fine-grained parallel implementation on the FPGA for the RNAalifold method

The most popular computational approach for RNA secondary structures based on using the MFE-model is the *RNAalifold* folding method, which was presented by Hofacker et al in.<sup>35</sup> The *RNAalifold* was used for computing the consensus of RNA structures and it had been implemented as an extension of Zuker's algorithm.<sup>13</sup> Also, the *RNAalifold* algorithm was considered the thermodynamic energy minimization stability with an average energy matrix and sequence co-variation score matrix together. The *RNAalifold* approach has the worst case execution time and storage space complexities are respectively  $O(m \times n^4 + n^3)$  and  $O(n^2)$ , as  $n$  is the length of RNA sequence and  $m$  is the number of RNA sequences in the alignment. Due to these strong reasons Xia et al in<sup>18</sup> adopted the *RNAalifold* algorithm and re-implemented it on FPGA chips.

The researchers in<sup>18</sup> supplied an innovative accelerated approach for the *RNAalifold* algorithm. They proposed a systolic array structure with one master Processing Element (PE) and many slave PEs for fine-grained hardware implementation on the FPGA. Their goals from this re-implementation, was to parallelize the original *RNAalifold* algorithm. The master PE loads the energy matrices from the outer memory "DRAM", while the other slave PEs remain waiting to take data from the master PE. The execution time of the parallel prediction algorithm was grown to be more than 12× on one FPGA with 16PEs. This results compared with the previous results for *ViennaRNA-1.6.5* software,<sup>51</sup> the original *RNAalifold* algorithm.

### (ii) Fine-grained parallelization of Nussinov's algorithm implemented on FPGA

Many RNA folding methods utilized empirical models for predicting the RNA secondary structures based on the MFE estimation by using the Dynamic Programming (DP) algorithm. Jacob et al in<sup>19</sup> pointed out that in all these empirical models, algorithms for predicting RNA structures are just DP recurrences from the original Nussinov's algorithm.<sup>12</sup> There are two features that make this algorithm computationally more applicable for predicting RNA secondary structures among the other prediction methods; Firstly, Nussinov's algorithm used the length of RNA sequence as a proxy for the MFE and it computed

the most maximum of RNA base pairs. Secondly, the original Nussinov<sup>12</sup> one runs in  $O(n^3)$  time and requires  $O(n^2)$  storage space complexities.

Jacob et al in<sup>19</sup> adapted one of the most recent parallel FPGA architectures (*Virtex-II 6000* FPGA) for implementing the normal Nussinov's algorithm. Also, the researchers calculated and built it on two classic two-dimensional (2D) systolic arrays to achieve optimal string parenthesization and to deal with the maximum length of RNA primary sequence. Hence, Jacob et al<sup>19</sup> implemented this design of accelerating the original Nussinov RNA structural prediction method by utilizing the 2D systolic arrays on FPGA implementation. This parallel design obtained good throughput results on the *Virtex-II 6000* FPGA. This output was compared with the same implemented method on the modern x86 CPU. The results obtained better factor up to 39× speed-ups in execution time.

### (iii) Parallel fine-grained implementation for the Zuker algorithm on FPGA

The most well-liked and admired computational approach for the RNA secondary structure based on using the MFE is the Zuker's algorithm.<sup>13</sup> It was confirmed to be the most stable RNA secondary structure prediction method based on calculating MFE. The Zuker's algorithm runs in  $O(n^4)$  execution time and resides in the  $O(n^2)$  storage memory requirements complexities. Recently, different developments and many optimal algorithms have been derived to simplify and reduce the Zuker's algorithm computational complexities. Also, most biology researchers have proved that, the RNA molecule existing in large sequences is more than what was expected before. In other words, biological experiments have shown that RNA molecules fold in thousands of bases.<sup>15</sup> Also, new discoveries make a huge increase of the RNA data sources; the data in GenBank almost doubling every year.<sup>52</sup> The original polynomial time complexity of the Zuker's algorithm is  $O(n^4)$ . This complexity could not be able to deal with these sophisticated instances introduced earlier, in a sequential manner.

The promising way to make Zuker's algorithm tolerable and balanced in calculation is a parallel structural design implementation. Recently, Dou et al in<sup>20</sup> proposed an innovative parallel design for accelerating Zuker's algorithm. The investigators in<sup>20</sup> utilized and built their techniques by dividing the Zuker algorithm

matrix in the usual mode. Then, they submitted and distributed the subtasks as a multithreading procedure that can be calculated independently. This parallel scheme was implemented on the FPGA fine-grained hardware by proposing one master PE and multiple slave PEs with systolic array structure. In addition, the RNA researchers in<sup>20</sup> projected new methods for reducing the RNA energy lookup table size by 85%. These lookup tables should be loaded in the memory, as the RNA prediction algorithm requires to use them. Consequently, they implemented their parallel algorithm on 16 PEs on FPGA co-processors. The experimental results of this parallel scheme explained enhancement up to factor of 14× speed-up time comparing with the *ViennaPackage*.<sup>51</sup>

## RNA Parallel Taxonomy

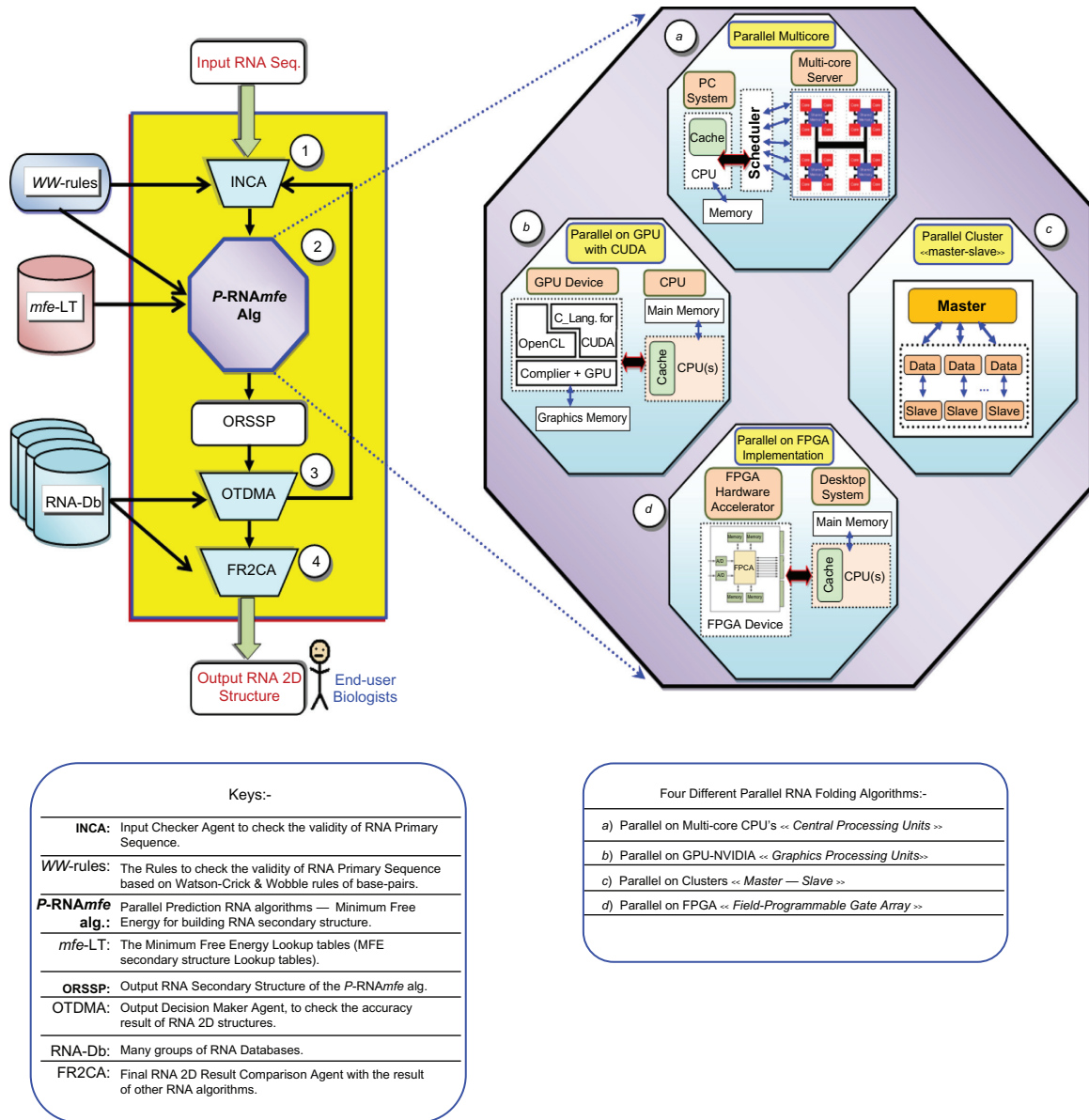
### Parallel taxonomy of RNA folding

In particular, the most popular RNA secondary structure prediction methods are the dynamic programming algorithms based on the MFE models. Herein, we proposed and described a general taxonomy to apply on the existing RNA parallel methods. This proposed taxonomy for the RNA parallel secondary structure methods is using four main phases. Three of them working as a different agents and the fourth one is the main RNA prediction algorithm, as shown in Figure 5:

- The Input Checker Agent (INCA) receives the RNA primary sequence from the input device or reads it from a text file. Subsequently, the INCA checks the validity of this RNA by applying the valid nucleotides RNA bases in Watson-Wobble rules (*WW*-rules).<sup>10,24</sup>
- INCA transfers the valid RNA sequence to the parallel prediction RNA MFE algorithm (*P-RNA<sub>mfe</sub>*), which starts calculating the RNA secondary structure by using *WW*-rules to combine the valid RNA base-pairs. The *P-RNA<sub>mfe</sub>* algorithm simultaneously predicts the optimal and most stable RNA secondary structure by using the MFE lookup tables (*mfe-LT*).<sup>15,28</sup> These *mfe-LT*s are loaded to the shared cache memory by loader just before the *P-RNA<sub>mfe</sub>* algorithm start calculating and predicting the RNA secondary structures. Essentially, based on this investigation, there are four parallel techniques was used to re-implement the previous RNA prediction methods. These four popular parallel RNA

prediction methods were illustrated by zooming in the “*P-RNA<sub>mfe</sub>*”, as shown in Figure 5 (a, b, c and d). A brief explanation of these four implementation parallel methods are as follows:

- *First*, a parallel design to harness the power of the multi-core CPUs for predicting the RNA secondary structures. The *GTfold* is a parallel multi-core algorithm.<sup>15</sup> It computed a larger RNA sequences, as shown in Figure 5 (a).
- *Second*, a parallel paradigm on the latest modern GPU-NVIDIA cards. This parallel design was utilizing the CUDA programming code in the C-language environment. This new scheme was used in<sup>16</sup> to accelerate the execution time for the previous RNA folding method.<sup>48</sup> This design was explained in Figure 5 (b).
- Another parallel design is a combination of two existing pseudoknotted RNA secondary structures prediction methods.<sup>25,34</sup> This design<sup>17</sup> used a parallel master-slave techniques based on the Beowulf cluster/hardware, as depicted in Figure 5 (c).
- The last parallel scheme elaborated the 2D systolic array parallel by using FPGA in fine-grained hardware. This parallel paradigm was illustrated in Figure 5 (d). Actually, there were three various parallel RNA detection methods.<sup>18–20</sup> These original RNA prediction methods were lately re-implemented on FPGA.
- The Output Decision Maker Agent (OTDMA) checks the accuracy of the initial Output RNA Secondary Structure Production (ORSSP). It performs this checkable process by testing and comparing the first output “ORSSP” with the diverse known RNA structures (as a standard benchmark for testing and evaluating the RNA prediction method) in the RNA Databases (RNA-Db).<sup>53–57</sup> If the OTDMA finds the quality of the new ORSSP is poor or low, it feeds the RNA sequence back to the INCA to re-start another round of detection with new and more intelligent constraints. Or else, the OTDMA finds the accuracy of the output RNA structure is high, it transfers the new ORSSP to the last agent, which called Final Result 2D Comparison Agent (FR2CA).
- The FR2CA compares the final RNA secondary structure result with the known and existing structures in RNA-Db, to report the accuracy scale



**Figure 5.** Parallel taxonomy of RNA folding algorithms. (1) INCA: Agent to check the validity of input RNA primary sequence. (2) P-RNAmfe Alg: Parallel RNA secondary structure prediction algorithm based on MFE. It zooms out in [a, b, c or d]. (3) OTDMA: Agent to compare the first result with existing online RNA databases. (4) FR2CA: Agent to measure the performance of the RNA structural prediction method with the standard benchmarks.

of predicted RNA.<sup>58</sup> In addition, the FR2CA agent measures the execution time and space complexities for the RNA prediction algorithm.

### Evaluating existing RNA algorithms using comparative taxonomy

By using this proposed comparative taxonomy, as a classified comparison procedure for the existing parallel RNA prediction methods. The authors found that some of these RNA parallel prediction

methods ran through all the comparative taxonomy phases. While, the other RNA prediction methods went through some steps of this proposed comparative taxonomy.

From the methodological point of view, the researchers explained and investigated a comparison procedure among these RNA parallel methods according to the proposed taxonomy in Table 2. The selected group of the previous RNA algorithms focuses on the parallel implementation of the existing RNA secondary structure methods.

**Table 2.** The evaluating taxonomy for the parallel RNA secondary structure prediction approaches.

| <b>Taxonomy Phase</b>  |   |   |   |   |  |  |
|--|---|---|---|---|--|--|
| <b>RNA Parallel Method</b>                                     | <b>GTfold RNA Alg.<sup>15</sup> on Multi-core</b>       | <b>Accelerated RNA Alg. on GPU<sup>16</sup></b> | <b>compPKnots: RNA Alg.<sup>17</sup> on Cluster «Master-Slave»</b>            | <b>Parallelize RNAalifold Alg. On FPGA<sup>18</sup></b>       | <b>Accelerating Nussinov Alg. on FPGA<sup>19</sup></b>         | <b>Parallel Zuker Alg. on FPGA<sup>20</sup></b>    |
| Check valid RNA Seq. by INCA                                   | ✓   | ✓   | ✓   | ✓   | ✓  | ✓  |
| Applying canonical WW-rules                                    | ✓   | ✓   | ✓   | ✓   | ✓  | ✓  |
| Using mfe-Lookup tables "mfe-LT"                               | ✓   | ✓   | ✓   | ✓   | ✓  | ✓  |
| Comparing the 1st output with existing RNA structure via OTDMA | ✓   | ✓   | Automatically comparing   | comparing to ViennaRNA  | Only with original Nussinov                                    | comparing to ViennaRNA                             |
| Compare accuracy by FR2CA                                      | mfold <sup>29,46</sup> and RNAfold <sup>14</sup>        | Unafold <sup>48</sup>                           | Pknots-RE <sup>25</sup> and Pknots-RG <sup>34</sup>                           | RNAalifold <sup>35</sup>                                      | Nussinov Alg. <sup>12</sup>                                    | Zuker Alg. <sup>13</sup>                           |
| Types of RNA including prediction                              | Stem-loop RNA Structure                                 | Stem-loop RNA Structure                         | Pseudoknotted RNA Structure   | Stem-loop RNA Structure                                       | Stem-loop RNA Structure  | Stem-loop RNA Structure                            |
| Parallel Framework   | Parallel Multicore and Scalable Program by using OpenMP | CUDA programming on GPU card                    | Master-slave paradigm using MPICH library                                     | 16 PE's on FPGA chips to accelerate RNAalifold RNA alg.       | 2D systolic array design implemented on a Virtex-II 6000 FPGAs | 16 PE's on FPGA chips to accelerate Zuker RNA alg. |
| Parallel Improvements  | Speed-up factor of 1.6x on execution time               | Achieving a factor of x17 on Speed-up           | Achieving results in a shorter amount of time Avg. $O(n^4)$ speed-up for both | A factor of 12.2 x Speed-up over RNAalifold (ViennaRNA-1.6.5) | Achieving Speed-up up to 39x over a recent x86-family CPU      | Speed-up of more than 14x over the ViennaRNA-1.6.5 |



## Discussion and Comparison

Scores of RNA researchers have introduced several evolutionary parallel blueprints for solving the RNA secondary structure problem. In this paper, the researchers discussed and compared these parallel RNA methods, as shown in Table 3. They have discussed in<sup>22</sup> an intensive RNA detection algorithms, as a first phase of the RNA secondary research in a Bioinformatics domain. The authors classified and compared the RNA approaches in two main groups and they presented the results in the two tables. In this study, the researchers extracted the well-known RNA sequential prediction methods in Table 1.

Then, in this paper, the authors focused their analysis on discussing and comparing the latest parallel RNA prediction efforts that have been pioneered in the RNA secondary structure folding domain. Consequently, from the comparison of the parallel RNA methods in Table 3, it could be noticed that only one research<sup>17</sup> presented a parallel design for the RNA pseudoknots type. While the others, proposed parallel designs for the RNA detection on stem-loops types. The comparison had also shown that the latest parallel algorithm<sup>16</sup> applied on the GPU, by utilizing the power of the NVIDIA card with the CUDA program. This method was implemented on the GPU to solve the RNA secondary structure problems. The high accessibility of the GPU card in the contemporary machines and the provided features for the developer to utilize GPU using high level languages like a C language environment; are the important motivations for the RNA research community to exploit the CUDA on the GPU.

Finally, the authors concluded that, implementing the RNA secondary structure prediction methods on parallel architectures has several significant benefits. The main advantage is reduced complexity, in both time and memory storage, comparing with the original RNA structure prediction algorithms. Also, the parallel RNA algorithms provided accurate results with better performance.

## Conclusion

In this paper for the solving RNA secondary structure prediction problem, the authors presented the state-of-the-art of the RNA parallel methods. They introduced an intensive investigation on exhaustive up to date parallel RNA secondary structure

prediction algorithms. Indeed, recently various methods and techniques for predicting RNA secondary structures have emerged. Most of these computational methods have faced with some complexities, from sequential implementation viewpoint. Therefore, re-implementing the existing RNA prediction methods by using parallelization architectures would result and contribute in solving the faced difficulties. The study concluded and showed that all of RNA parallel methods obtained better results, when compared to the sequential methods in terms of accuracy in one side and time/space complexities, in another side.

This research study was comprised in three trends: (i) RNA in biological context. (ii) RNA computational prediction methods. (iii) A comparative taxonomy of RNA parallel methods. In the first part, the researchers explained the experimental method's difficulties that the biologists are facing, in determination of RNA 3D structures. Exploratively, the researchers concluded the reasons that force and allow the RNA molecule to fold and pair back on itself forming a double helices, in a chemical and biological context. Secondly, in the methods and research findings, the authors listed the sequential RNA secondary structure methods, in a schematic classification diagram. Then, they compared their contributions and complexities. Also, the authors investigated comprehensively the state-of-the-art of the RNA parallel prediction methods.

In addition, the researchers performed a scientific comparison among these enhanced RNA parallel methods with the previous existing methods. In the third part, the researchers proposed a new parallel taxonomy. Then, they applied the existing parallel methods using this taxonomy. Lastly, the researchers conducted a comparison procedure to evaluate these RNA parallel methods based on the proposed taxonomy, in terms of the taxonomy steps.

Consequently, this study of the RNA existing parallel methods proved that the parallelization performance of the algorithm is proportional to the method of the parallelization itself. Particularly, the comparison showed that the proposed RNA methods utilizing GPU capabilities result more promising outputs. Besides that, from the implementation point of view, the available open source Application Programming Interface (API) in a high level language in C environment could be considered as a positive



**Table 3.** Comparison parallel algorithms for RNA secondary structure prediction.

| Parallel paradigm  | RNA existing algorithms  | Original complexities   |                   | Major contribution of the parallelization method   | The enhancement on complexities by using parallelization design  |                   |
|--|--|-------------------------|-------------------|--|--|-------------------|
|  |  | Execution time          | Space requirement |  | Speed-up time  | Space requirement |
| RNA prediction algorithms on multicore Parallelization "GTFold"                    | <i>mfold</i> <sup>29,46</sup> and <i>RNAfold</i> <sup>14</sup> | $O(n^4)$                | $O(n^2)$          | <i>GTFold</i> <sup>15</sup> combine with enhancement for RNA prediction algorithms ( <i>mfold</i> & <i>RNAfold</i> ) on CPU multicore Parallelization  | $O(n^3)$ with a factor of 1.6x on execution time                 | –                 |
| Accelerating RNA secondary structure algorithms on GPU                             | Unafold package "hybrid-ss-min" <sup>34,8</sup>                | $O(n^3)$                | $O(n^2)$          | Adapting parallel function on <i>hybrid-ss-min</i> for RNA prediction and re-implementing on GPU   | A factor of 17x on speed-up time                                 | –                 |
| Parallel RNA predictions alg. on Beowulf cluster «Master–Slave»                    | <i>Pknots-RE</i> <sup>25</sup>                                 | $O(n^6)$                | $O(n^4)$          | <i>compPknots</i> <sup>17</sup> : a parallel framework by using combination of both existing methods ( <i>Pknots-RE</i> , <i>Pknots-RG</i> ) and running both alongside for prediction RNA structure with more accuracy and shorter time | $O(n^4)$   | Avg. $O(n^2)$     |
|  | <i>Pknots-RG</i> <sup>34</sup>                                 | $O(n^4)$                | $O(n^2)$          |  |  |                   |
| Parallelizing RNA secondary structure algorithms on FPGA chips                     | RNAalifold alg. <sup>35</sup>                                  | $O(m \times n^4 + n^3)$ | $O(n^2)$          | A systolic array structure using fine-grained parallel on FPGAs <sup>18</sup> to accelerate RNAalifold algorithm   | A factor of 12x on speed-up time                                 | –                 |
| Parallelizing Nussinov RNA structural algorithms on FPGA co-processors             | Nussinov's alg. <sup>12</sup>                                  | $O(n^3)$                | $O(n^2)$          | A parallel systolic arrays on FPGA <sup>19</sup> co-processors for accelerating Nussinov's RNA algorithm   | A factor of 39x on speed-up time                                 | –                 |
| Accelerating Zuker's algorithm for RNA structural by Parallel fine-grained on FPGA | Zuker's algorithm <sup>13</sup>                                | $O(n^4)$                | $O(n^2)$          | A parallel systolic arrays on FPGA <sup>20</sup> co-processors for accelerating Zuker's RNA algorithm  | Up to factor of 14x speed-up comparing with <i>ViennaPackage</i> | –                 |



point. These parallelizing RNA secondary structure prediction methods, showed a promising area for future RNA studies and for computational RNA bio-informatic researches.

## Acknowledgment

This research was partly supported by the Universiti Sains Malaysia (USM) Fellowship awarded to the corresponding author. Also, it has been funded by USM Research Grant for the “A PARALLEL SWARM INTELLIGENCE BASED ALGORITHM FOR PROTEIN CONFORMATIONAL SEARCH” [304/PKOMP/638123]. The authors would like to gratefully acknowledge the help of Dr. Michael Liau Tet Loke (Editor, in the Postgraduate Academic Support Services (USM-P.A.S.S)) for proof-reading this paper.

## Disclosures

This manuscript has been read and approved by all authors. This paper is unique and is not under consideration by any other publication and has not been published elsewhere. The authors and peer reviewers of this paper report no conflicts of interest. The authors confirm that they have permission to reproduce any copyrighted material.

## References

- Hogeweg P. Simulating the growth of cellular forms. *Simulation*. 1978 September 1;31(3):90-a-96.
- Castanotto D, Rossi JJ. The promises and pitfalls of RNA-interference-based therapeutics. *Nature*. 2009;457(7228):426–33.
- Osborne R. Companies jostle for lead in RNAi, despite uncertainties. 2007.
- Thomas D, Frank S. *RNAi and microRNA-oriented therapy in cancer: rationales, promises and challenges*; 2009.
- Buratti E, Dhir A, Lewandowska MA, Baralle FE. RNA structure is a key regulatory element in pathological ATM and CFTR pseudoexon inclusion events. *Nucl Acids Res*. 2007 July 26;35(13):4369–83.
- Reeder J, Steffen P, Giegerich R. pknotsRG: RNA pseudoknot folding including near-optimal structures and sliding windows. *Nucleic acids research*. 2007;35(Web Server issue):W320.
- Cheong HK, Hwang E, Lee C, Choi BS, Cheong C. Rapid preparation of RNA samples for NMR spectroscopy and X-ray crystallography. *Nucl Acids Res*. 2004 June 15;32(10):e84-.
- Felden B. RNA structure: experimental analysis. *Current Opinion in Microbiology*. 2007;10(3):286–91.
- Draper DE. A guide to ions and RNA structure. *RNA*. 2004 March;10(3):335–43.
- Akutsu T. Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots. *Discrete Applied Mathematics*. 2000;104:45–62.
- Waterman M, Smith T. RNA secondary structure: A complete mathematical analysis. *Math Biosci*. 1978;42(3–4):257–66.
- Nussinov R, Pieczenik G, Griggs J, Kleitman D. Algorithms for loop matchings. *SIAM Journal on Applied Mathematics*. 1978:68–82.
- Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic acids research*. 1981;9(1):133–48.
- Hofacker I, Fontana W, Stadler P, Bonhoeffer L, Tacker M, Schuster P. Fast folding and comparison of RNA secondary structures. *Monatshfte Chemie*. 1994;125:167–88.
- Mathuriya A, Bader DA, Heitsch CE, Harvey SC. GTfold: a scalable multicore code for RNA secondary structure prediction. *Proceedings of the 2009 ACM symposium on Applied Computing*. Honolulu, Hawaii: ACM; 2009.
- Rizk G, Lavenier D. GPU Accelerated RNA Folding Algorithm. *Computational Science—ICCS*. 2009:1004–13.
- Estrada T, Licon A, Taufer M. CompPknots: A Framework for parallel prediction and comparison of RNA secondary structures with pseudoknots. *Frontiers of High Performance Computing and Networking—ISPA 2006 Workshops*; 2006:677–86.
- Xia F, Dou Y, Zhou X, Yang X, Xu J, Zhang Y. Fine-grained parallel RNAa-lifold algorithm for RNA secondary structure prediction on FPGA. *BMC Bioinformatics*. 2009;10(Suppl 1):S37.
- Jacob A, Buhler J, Chamberlain R. Accelerating Nussinov RNA secondary structure prediction with systolic arrays on FPGAs, 2008.
- Dou Y, Xia F, Zhou X, Yang X. Fine-grained parallel application specific computing for RNA secondary structure prediction on FPGA, 2008.
- Sharma S, Ding F, Dokholyan N. iFoldRNA: three-dimensional RNA structure prediction and folding. *Bioinformatics*. 2008;24(17):1951.
- Al-Khatib RM, Abdullah R, Abdul Rashid N. A survey of compute intensive algorithms for ribo nucleic acids structural detection. *Journal of Computer Science*. 2009;5(10):680–9.
- Andronescu M, Aguirre-Hernandez R, Condon A, Hoos HH. RNAsoft: a suite of RNA secondary structure prediction and design software tools. *Nucl Acids Res*. 2003 July 1;31(13):3416–22.
- Poolsap U, Kato Y, Akutsu T. Prediction of RNA secondary structure with pseudoknots using integer programming. *BMC Bioinformatics*. 2009;10(Suppl 1):S38.
- Rivas E, Eddy SR. A dynamic programming algorithm for RNA structure prediction including pseudoknots. *Journal of molecular biology*. 1999;285(5):2053–68.
- Lemieux S, Major F. RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucl Acids Res*. 2002 October 1;30(19):4250–63.
- Dirks RW, Molenaar C, Tanke HJ. Visualizing RNA molecules inside the nucleus of living cells. *Methods*. 2003;29(1):51–7.
- Lyngso RB, Zuker M, Pedersen CNS. Internal loops in RNA secondary structure prediction. *Proceedings of the third annual international conference on Computational molecular biology*. Lyon, France: ACM; 1999.
- Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of molecular biology*. 1999;288(5):911–40.
- Lyngso RB, Pedersen CNS. RNA pseudoknot prediction in energy-based models. *Journal of Computational Biology*. 2000;7(3–4):409–27.
- Ruan J, Stormo GD, Zhang W. An Iterated loop matching approach to the prediction of RNA secondary structures with pseudoknots. *Bioinformatics*. 2004 January 1;20(1):58–66.
- Dirks RM, Pierce NA. A partition function algorithm for nucleic acid secondary structure including pseudoknots. *Journal of Computational Chemistry*. 2003;24(13):1664–77.
- Sperschneider J, Datta A. KnotSeeker: Heuristic pseudoknot detection in long RNA sequences. *Bioinformatics*. 2008;14:630–40.
- Reeder J, Giegerich R. Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics. *BMC Bioinformatics*. 2004;5(1):104.



35. Hofacker IL, Fekete M, Stadler PF. Secondary structure prediction for aligned RNA sequences. *Journal of molecular biology*. 2002;319(5):1059–66.
36. Gardner P, Giegerich R. A comprehensive comparison of comparative RNA structure prediction approaches. *BMC Bioinformatics*. 2004;5(1):140.
37. Fang XY, Luo ZG, Wang ZH. Predicting RNA secondary structure using profile stochastic context-free grammars and phylogenetic analysis. *J Comput Sci Technol*. 2008;23(4):582–9.
38. Ziv-Ukelson M, Gat-Viks I, Wexler Y, Shamir R. A Faster algorithm for RNA co-folding. *Algorithms in Bioinformatics*. 2008:174–85.
39. Jens K, Andrew ET. RNA secondary structure prediction using a self-consistent mean field approach. *Journal of Computational Chemistry*. 2009;9999(9999):NA.
40. Chen X, He SM, Bu D, et al. FlexStem: improving predictions of RNA secondary structures with pseudoknots by reducing the search space. *Bioinformatics*. 2008 September 15;24(18):1994–2001.
41. Hengwu L, Daming Z, Zhendong L, Hong L. Prediction for RNA planar pseudoknots. *Progress in Natural Science*. 2007;17(6):717–24.
42. McCaskill JS. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*. 1990;29(6–7):1105–19.
43. Ren J, Rastegari B, Condon A, Hoos HH. HotKnots: Heuristic prediction of RNA secondary structures including pseudoknots. *RNA*. 2005 October;11(10):1494–504.
44. Sankoff D. Simultaneous solution of the RNA folding, alignment and protosequence problems. *SIAM Journal on Applied Mathematics*. 1985;45(5):810–25.
45. Lyngso RB, Zuker M, Pedersen CN. Fast evaluation of internal loops in RNA secondary structure prediction. *Bioinformatics*. 1999 June 1;15(6):440–5.
46. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucl Acids Res*. 2003 July 1;31(13):3406–15.
47. Westerhout EM, Ooms M, Vink M, Das AT, Berkhout B. HIV-1 can escape from RNA interference by evolving an alternative structure in its RNA genome. *Nucl Acids Res*. 2005 February 1;33(2):796–804.
48. Markham NR, Zuker M. DINAMelt web server for nucleic acid melting prediction. *Nucl Acids Res*. 2005 July 1;33(suppl 2):W577–81.
49. Sonzogni VE, Yommi AM, Nigro NM, Storti MA. A parallel finite element program on a Beowulf cluster. *Advances in Engineering Software*. 2002/10//2002;33(7–10):427–43.
50. Byun Y, Han K. PseudoViewer3: generating planar drawings of large-scale RNA structures with pseudoknots. *Bioinformatics*. 2009;25(11):1435–7.
51. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. The Vienna RNA Websuite. *Nucl Acids Res*. 2008 July 1;36(suppl 2):W70–4.
52. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. *Nucl Acids Res*. 2008 January 11;36(suppl 1):D25–30.
53. Mituyama T, Yamada K, Hattori E, et al. The Functional RNA Database 3.0: databases to support mining and annotation of functional RNAs. *Nucl Acids Res*. 2009 January 1;37(suppl 1):D89–92.
54. Wuyts J, Perriere G, Van de Peer Y. The European ribosomal RNA database. *Nucl Acids Res*. 2004 January 1;32(suppl 1):D101–3.
55. Gustafson AM, Allen E, Givan S, Smith D, Carrington JC, Kasschau KD. ASRP: the arabidopsis small RNA project database. *Nucl Acids Res*. 2005 January 1;33(suppl 1):D637–40.
56. Taufer M, Licon A, Araiza R, et al. PseudoBase++: an extension of pseudo-base for easy searching, formatting and visualization of pseudoknots. *Nucl Acids Res*. 2009 January 1;37(suppl 1):D127–35.
57. van Batenburg FHD, Gulyaev AP, Pleij CWA, Ng J, Oliehoek J. Pseudo-Base: a database with RNA pseudoknots. *Nucl. Acids Res*. January 1, 2000 2000;28(1):201–204.
58. Tsang HH, Wiese KC. SARNA-Predict: A simulated annealing algorithm for RNA secondary structure prediction. Paper presented at: computational intelligence and bioinformatics and computational biology, 2006. CIBCB '06. 2006 IEEE Symposium on, 2006.

**Publish with Libertas Academica and every scientist working in your field can read your article**

*“I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely.”*

*“The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I've never had such complete communication with a journal.”*

*“LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought.”*

**Your paper will be:**

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

**<http://www.la-press.com>**