BioCloud: A Network Cloud-Computing Method for Predicting RNA Secondary Structure

Ahmad Habboush, Mohammad Al Rawajbeh, Ahmed M. Manasrah, and Ra'ed M. Al-Khatib

***Abstract*—**Predicting RNA secondary structure becomes an important issue, due to the useful functions of RNA in designing antiviral drugs for AIDS and malignant diseases like cancer. Many computational methods have been proposed to predict RNA secondary structure from a given single sequences. In this paper, the *BioCloud* method is proposed, which is more accurate cloud-computing method for predicting the secondary structure of RNA. The proposed *BioCloud* method runs via a global networks as a cloud computing to predict the RNA secondary structure. Practically, the proposed method uses a Minimum Free Energy (MFE) algorithm with the Dynamic Programming (DP) prediction method to predict the needed structure of RNA molecules. The experimental results show that the *BioCloud* method runs faster and predict more accurate RNA secondary structure compared to the state-of-the-art prediction methods exist in the literature. The proposed *BioCloud* method performs efficiently in a short running time and available globally in a cloud-computing system.

***Keywords*—**Ribonucleic acid (RNA), Secondary Structure, Tertiary Structure, Cloud Computing.

I. Introduction

BIOINFORMATICS is a new disciplinary comes from the combination between Computer Science and Biology. The main focus of researchers is to design computational methods that are essentially introduced to manage, analyse and predict structure of the biological data components like (Protein, DNA & RNA) [1]. Ribonucleic acid (RNA) plays important roles in living cells. The messenger RNA (mRNA) runs as an intermediate to carry the genetic information from DNA to produce a protein in protein synthesis [2]. Central dogma proves that determining and studying the RNA structures assists to scrutinise the RNA functions. Therefore, predicting the RNA secondary structure by using the computational methods is the main step to know their useful functions, which will be helpful to build antiviral drugs for AIDS and malignant diseases like cancer [3], [4], [5].

The biologists use experimental methods (Nuclear Magnetic Resonance (NMR) and X-ray crystallography) to determine the tertiary (3D) structures of RNA molecules [6], [7]. These two experimental methods however are time consuming, tedious and difficult to be accomplished based on experiments in biology side [8]. Because of these constraints in performing the experimental methods, the computer scientists proposed many Bioinformatics methods to predict the secondary structures of RNA from the input single sequences. Then, the demanded 3D tertiary structure can be scrutinized from the predicted RNA secondary structure by applying visualizing tools like [9], [10], [11]. These visualization tools can display the needed 3D structure by using the predicted pseudoknots RNA secondary structure to generate the tertiary structure of RNA molecule, which is similar to the real 3D tertiary structure in biology side [9].

To solve the problem of the RNA secondary structure prediction, several Dynamic Programming (DP) method [12], [13] methods and the heuristic methods [14], [4], [15] have been proposed. In GenBank, a big gap is still between the huge discovered number of genome-sequences and the known structures like RNA structures [16]. Figure 1 illustrates the growth of the genome databases, while the zoomed-in shows the gap in growth of determining structures. So, it explains the different growth rate between the big number of primary genome-sequences and the limited number of known structures.

Basically, RNA comes into two different structural types. The RNA secondary structure without pseudoknots and the secondary structure with *pseudoknots* structure. The RNA without pseudoknots type as shown in figure 2 (a), is predicted by MFE model with DP algorithms, and they mostly require $O(n^3)$ for time and $O(n^2)$ for space complexity [12], [4]. The RNA with pseudoknots type as shown is figure 2 (b), have recently been emphasised in most classes of the RNA molecules. The RNA with pseudoknots type are confirmed to be an important structural functions [17], [18]. The existing methods in the literature only predict a part of the RNA with pseudoknots type; Due to the insufficiency of knowing the MFE lookup tables for pseudoknots structural parts. The prediction of RNA secondary structure with pseudoknots by using DP algorithms, are to be an NP-hard problem [19], [20].

The proposed *BioCloud* method in this paper is built and qualified for predicting secondary structure of RNA with pseudoknots from a given single sequence. The *BioCloud* performs in a cloud computing fashion via a global network

Ahmad Habboush is with Faculty of Computer and information technology, Jerash Private University, Jerash, Jordan

Mohammad Al Rawajbeh is with Faculty of sciences and information technology, Al Zaytoonah university of Jordan, Amman, Jordan

Ahmed M. Manasrah and Ra'ed M. Al-Khatib are Faculty of Information Technology and Computer Science, Yarmouk University, Irbid 21163, Jordan

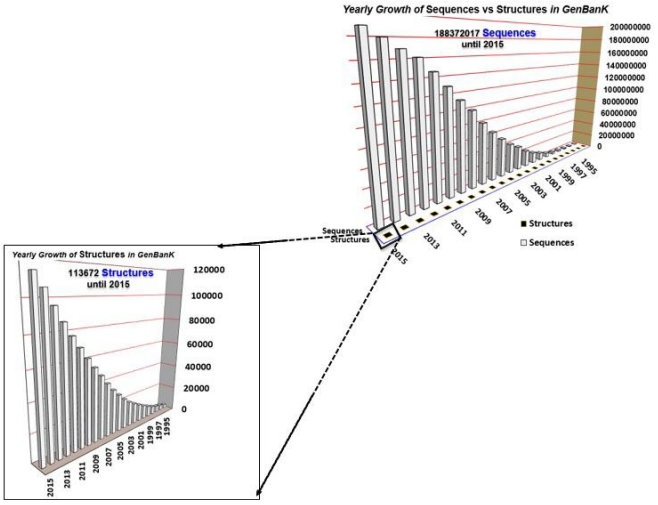

Fig. 1 Exponential growth of biological data in GenBank and the small growth of their known Structures

system. The single sequence of RNA is firstly downloaded from GenBank. After filtering this downloaded RNA sequence, the proposed algorithm divides it into proper hits by using the GUUGle algorithm [21]. Then, the cloud-computing algorithm of proposed method sends these hits to the pknotsRG algorithm [22], to predict the possible pseudoknots structural parts. Thereafter, the proposed *BioCloud* method globally predicts the free energy structure for the remaining parts by using the power of UNAFold web server [23]. Finally, the proposed method compiles and collects the global RNA secondary structure with pseudoknots. And it also builds the 3D structures of the given RNA by PseudoViewer3 as a visualization tool [10]. From the comparative results, it can be concluded that the proposed *BioCloud* method is the most accurate prediction method for building the RNA secondary structure with pseudoknots.

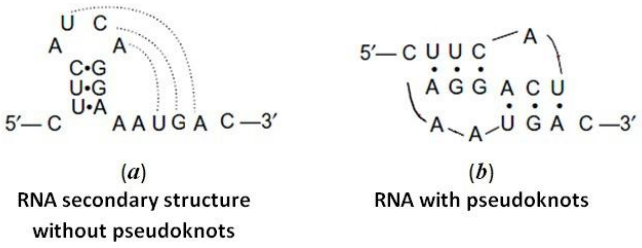

Fig. 2 Two types of RNA secondary structures: (a) without pseudoknots. (b) with pseudoknots structural parts

The rest of the paper is organised as follows: Section 2 presents the related work and the background of the genome RNA data that are tested in the comparative process. Section 3 describes the proposed *BioCloud* method with its workflow to predict the RNA secondary structure with pseudoknots. The experimental results and discussions are fully introduced in Section 4. Finally, in the conclusion the upcoming future works are briefly summarized.

II. Related Work

RNA like DNA is a nucleic acid but it is a linear molecule. RNA is composed of four ribonucleotide bases; adenine (A), cytosine (*C*), guanine (*G*), and uracil (*U*). Each ribonucleotide base contains ribose sugar, phosphate group and a nitrogenous base. The adjacent ribose nucleotide bases are connected together in chemical bonds to form the RNA chain. So, the single sequence of RNA, is known as a primary structure and it is a single chain or a single-stranded of the four bases (*A*, *C*, *G*, or *U*), as shown in figure 3 (a). Then, the RNA secondary structure is come from the chemical compound of base pairs that are formed according to the Watson-Crick two pairs *C*-*G* and *A*-*U*, and the Wobble pair *G*-*U* [24], [25], [7], as depicted in figure 3 (b). However, the 3D tertiary structure of RNA can be scrutinized from visualizing the pseudoknots structure of RNA as shown in figure 2 (b).

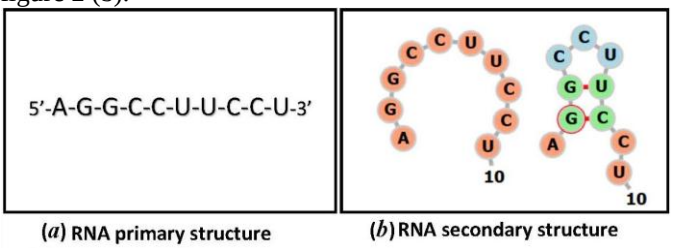

Fig. 3 The main two structures of RNA: (a) RNA primary structure as a single-stranded. (b) RNA secondary structure generated by using the (*forna*) visualization tool [26]

The Data of RNA with pseudoknots is downloaded from GenBank datasets and the PseudoBase [27], [28]. These RNA databases are used in this study to obtain the RNA sequences for different kinds of RNA molecules; like *mRNA*, *tmRNA* and *rRNA*. The pseudoknots secondary structures of these kinds of RNA-molecules are taken out from the PseudoBase [27], as real known-structures in order to perform the experimental process.

III. Material and Methoms

This section describes the details of proposed *BioCloud* method and its workflow. The algorithm of proposed method basically runs in a cloud-computing behavioral via global World Wide Web, to predict RNA secondary structure with pseudoknots. The basic implementation of the *BioCloud* proposed method is that there are multiple sites running globally through the cloud computing system. Our proposed method predict the needed RNA secondary structures of the biological data from a certain input RNA genome-sequence. The complete process of the bio-cloud computing algorithm of proposed *BioCloud* method, practically does the bio-cloud computing process to obtain the bio-structural outputs. This process of proposed *BioCloud* method, is purely invisible in a cloud based technology. Basically, the proposed method submits the input RNA bio-sequence as a service over the cloud computing scenario. Then, it submits the intermediate

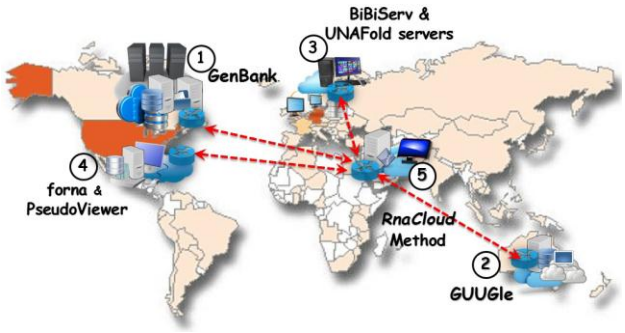

Fig. 4 A workflow of the proposed *BioCloud* method.
The numbers from (1 .. 5) represent the followings:
(1) GenBank at url: *www.ncbi.nlm.nih.gov/Genbank*.
(2) GUUGle algorithm at url: *www.uwa.edu.au*.
(3) BiBiServ server: *bibiserv.techfak.uni-bielefeld.de/* & UNAFold server: *unafold.rna.albany.edu/*.
(4) *forna* & PseudoViewer at: *http://nibiru.tbi.univie.ac.at/forna/* & *http://pseudoviewer.inha.ac.kr/*,respectively.
(5) Site of proposed *BioCloud* Method: *http://www.yu.edu.jo/en*

outputs based on virtual resources from invisible cloud sites [29]. Figure 4 illustrates the workflow of the proposed *BioCloud* method with detailed descriptions of the convergence of various virtual computers that are provisioned and connected as a single unified computing resources based on the cloud infrastructure technologies.

Because of the exponential growth and big scalable of RNA genome biological data, and instead of migrating this huge data around the world, we make the proposed *BioCloud* method available globally for cloud-based computing algorithmic. The environmental services of our proposed method are implemented as web-based algorithmic, which deployed by using the cloud computing background technology. For integrity and scalability, the proposed *BioCloud* method is computing in multiple stages and its ubiquitously accessible running provision over the cloud computing scenario as explained in Figure 5.

In the first stage, the proposed *BioCloud* downloads the primary sequence of RNA data in a FASTA format from GenBank (*www.ncbi.nlm.nih.gov/Genbank*). Then, there is purification procedure in *BioCloud* method to filter any noisy bases that can be found in the initial primary sequence before sending it to the GUUGle algorithm at the server of University of Western Australia (*www.uwa.edu.au*). In the second stage, the GUUGle algorithm divides the input sequence to many hits according to the suitable chemical components, which can be joining together forming a proper base pairs conferring with the complementary bases (Watson-Crick pairs *C-G* and *A-U*, and Wobble pair *G-U*). The proposed method use the power of GUUGle algorithm to extract these initial hits. Then the method sends these hits by using a cloud-based infrastructures to the predicted two approaches pknotsRG and UNAFold at (BiBiServ & UNAFold web servers), respectively.

So, the proposed method as a cloud-computing algorithmic, predicts the RNA secondary structure with pseudoknots.

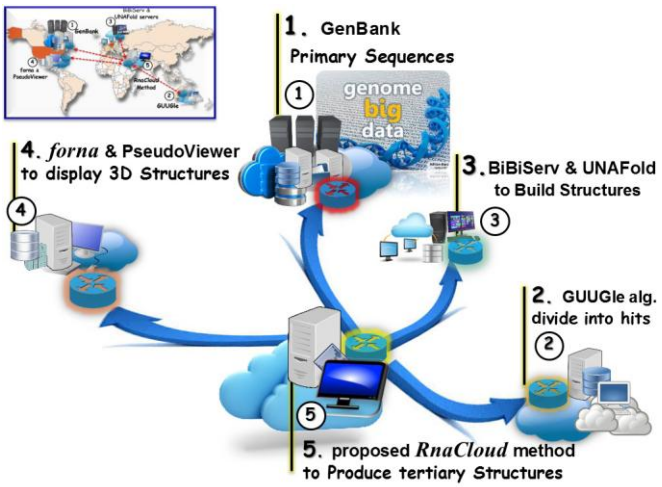

Fig. 5 Layout and architecture five stages of the proposed *BioCloud* method is built and designed over multi-recourses in cloud-based computing technology.
The details of these stages from (1 .. 5) are explained as follows:
(1) Primary Sequences of Genome data in FASTA format are downloaded from GenBank. www.ncbi.nlm.nih.gov/Genbank .
(2) Dividing the input Sequence into proper hits and subsequences using GUUGle algorithm. *www.uwa.edu.au/* .
(3) Predicting Secondary Structures of Genome data by most suitable prediction algorithmic by BiBiServ web server for pseudoknots parts *http://bibiserv.techfak.uni-bielefeld.de/* & UNAFold web server for non-pseudoknots part *http://unafold.rna.albany.edu/* .
(4) Building and Visualizing the 3D tertiary Structures from predicted Secondary structures by the visualization tools (*forna* & PseudoViewer servers) at url's: http://nibiru.tbi.univie.ac.at/forna/ and *http://pseudoviewer.inha.ac.kr/* .
(5) Collecting & producing final 3D tertiary Structures of the initial sequence for a given Genomedata by a cloud computing algorithms of our proposed *BioCloud* methods. *http://www.yu.edu.jo/en/*

Practically, *BioCloud* method performs the computational process in a global cloud-based computing technique by combining the final output of the RNA structure from the GUUGle algorithm (uwa server, Australia) [30], the pknotsRG method (BiBiServ, Germany) [22], and the UNAFold MFE method (UNAFold web server, NY, USA) [23]. The proposed method is functionally structured as a cloud-computing predictor, which is represented in Equation 1.

$$S = F_{BioCloud}(X, f_{guugle}, f_{pknotsRG}, f_{unafold}) \quad (1)$$

The vector *S* is the final predicted RNA secondary structure with pseudoknots types. Where $F_{BioCloud}$ denotes to the general proposed as a cloud-computing method. It predicts the final pseudoknots of the RNA secondary structure S by using the power of f_{guugle} , $f_{pknotsRG}$, and $f_{unafold}$. These functions represent the GUUGle algorithm, the pknotsRG method, and the UNAFold MFE method, respectively. These three algorithmic functions are managed globally by our proposed method as cloud-computing resources. The input RNA sequence is denoted by vector $X := \{x_i \mid i = 1,2,..,n\}$,

which includes the initial input RNA primary sequence of length *n*, as a single-stranded RNA molecule.

Finally, the proposed *BioCloud* method reconnects the hits and their structures. Then, it combines the predicted structure from different resources in a cloud-computing manner, to produce the complete final secondary structures for the first input RNA primary sequence. At the last stage, *BioCloud* method is using the visualization tools like (*forna* & PseudoViewer servers), which are widely utilized to scrutinize the 3D tertiary Structures for RNA genome data.

IV. Results And Discussion

In this section, an evaluation process to the results of proposed *BioCloud* method is discussed. Actually, the RNA structural results that are obtained from proposed *BioCloud* method are compared to the results of existing RNA prediction methods in literature like; FlexStem [31], HotKnots [15], pknotsRG [13], ILM [14] and NUPACK [32]. This performance evaluation was confirmed with standard native structures of RNA molecule that were downloaded from the real pseudoknots structural RNA database PseudoBase [27], [33]. The cloud-computing proposed method predicts the pseudoknots of the RNA secondary structure, and its results are evaluated in this comparison process to the real native RNA structures as a main factor. We use three different factors as accuracy metrics in the comparison process that are coined by [34], to evaluate the performance of the proposed *BioCloud* method. These three accuracy metrics (*Sensitivity* (*SE*), *Specificity* (*SP*) and *F-measure*), are illustrated in the following formulas:

$$Sensitivity(\%)\ SE = \frac{TP}{TP + FN} \times 100 \quad (2)$$

$$Specificity(\%)\ SP = \frac{TP}{TP + FP} \times 100 \quad (3)$$

$$F - measure(\%) = 2 \times SP \times (\frac{SE}{SP + SE}) \times 100 \quad (4)$$

The variables in these formulas are used to calculate the accuracy components of the proposed *BioCloud* prediction methods, and can be further explained as follows:

TP: True Positive (*TP*) base pair indicates the number of the predicted base pairs. Then, the base pairs found in the predicted RNA structure and the native RNA structure.

FN: False negative (*FN*) is the total number of the connected base pairs that are found in the known native structure, but not in the predicted RNA structure.

FP: False Positive (*FP*) is the number of base pairs, which are incorrectly predicted and are not found in the native structure.

F-measure: This measures is an indication of the algorithm's performance, which combines the Specificity (*SP*) and the *Sensitivity* (*SE*) of the algorithm in a single performance measurement.

We basically use the *Sensitivity* (*SE*) and *Specificity* (*SP*) in order to measure the performance of predicted structure; a third one, the *F-measure* joins both *SE* and *SP* to combine in one metric. By using same variables and referring to accuracy and performance measures from [35]. In table I, the results of the proposed *BioCloud* method are reported by using three accuracy metrics to the obtained outputs compared to the other existing RNA prediction methods; such as FlexStem [31], HotKnots [15], pknotsRG [13], ILM [14] and NUPACK [32]. In fact, the comparison process is performed according to the native RNA structures, which are derived from real structures of the PseudoBase database [27], [33].

TABLE I
Comparative Outputs and Results of the Proposed *BioCloud* Method and other RNA Prediction Methods Using the Three Performance Metrics (SE, SP & F-measure)

Prediction Methods	*SE*	*SP*	*F-measure*
***BioCloud* (2016)**	**89.3**	**84.1**	**86.4**
FlexStem (2008)	73.3	72.2	72.6
HotKnots (2005)	53.3	56.1	54.5
pknotsRG (2004)	67.8	74.1	70.7
ILM (2004)	55.8	61.1	58.7
NUPACK (2003)	63.8	61.1	62.7

The whole process of comparative results, shows that the predicted RNA structures by our proposed *BioCloud* method, more accurate than the structural results obtained from other RNA prediction methods. Hence, the performance, accuracy and precision of structural results obtained by the proposed method, are evaluated against five other state-of-the-art pseudoknots RNA prediction methods. As a result, the proposed method was outperformed by 23% totally, and is considered as the best accurate method compared to other RNA prediction methods (FlexStem, HotKnots, pknotsRG, ILM, and NUPACK methods), that exist in literature.

V. Conclusions

The study of this paper presented a novel cloud-computing method for predicting the secondary structure of RNA with pseudoknots. The proposed method is combined within a global manager programmatic called *BioCloud*. The proposed *BioCloud* method, combines the following three algorithms; the GUUGle algorithm (uwa server in Australia), the pknotsRG method (BiBiServ in Germany) and the UNAFold method (its web server in NY, USA), as global resources in a cloud-computing system. The evaluation process and the comparative results are conducted to distinguish the efficiency and performance of the proposed *BioCloud* method, compared to other state-of-the-art existing RNA prediction methods. The accuracy measurements (Sensitivity, Specificity & Specificity) is used to confirm that our proposed *BioCloud* method obtained the best output-results in predicting the RNA secondary structure with pseudoknots from a given input RNA sequence. It is worth noting that the main factor of the experimental comparison in the evaluation results, is the real origin of the RNA structures downloaded from the GenBank data [36] and the PseudoBase [27], [33].

In conclusion, it can be inferred from the obtained results of the proposed *BioCloud* method that the cloud-computing proposed method have high potential with a good performance and more efficiency for solving the problems.

And it introduces a good contribution in affecting the pseudoknots of the secondary structure of the RNA molecular computation within the bioinformatics community. This worthwhile domain can be involved in several future RNA and bioinformatics research directions such as: (1) pursuing on further studies relevant to the network and cloud-computing domain, (2) enhancing different factors of prediction processes and (3) combining many other RNA prominent methods into prediction stages.

ACKNOWLEDGMENT

The authors are gratefully thanks the anonymous reviewers for their insightful observations and their valuable comments. This research was partially supported by Yarmouk University, Al Zaytoonah university of Jordan and Jerash Private University. Particularly, the assistances of Computer Center and these were introduced by the technicians of the high parallel computing research centre at http://www.yu.edu.jo/en/.

REFERENCES

[1] C. Priami, "Algorithmic systems biology," *Communications of the ACM*,vol. 52, no. 5, pp. 80–88, 2009. http://dx.doi.org/10.1145/1506409.1506427

[2] R. Nussbaum, R. R. McInnes, and H. F. Willard, *Thompson & Thompson genetics in medicine*. Elsevier Health Sciences, 2015.

[3] J. S. Mattick, "Challenging the dogma: the hidden layer of non-proteincoding RNAS in complex organisms," *Bioessays*, vol. 25, no. 10, pp. 930–939, 2003. http://dx.doi.org/10.1002/bies.10332

[4] J. Sperschneider and A. Datta, "Knotseeker: Heuristic pseudoknot detection in long RNA sequences," *RNA*, vol. 14, no. 4, pp. 630–640, 2008. http://dx.doi.org/10.1261/rna.968808

[5] K.-T. Shum, J. Zhou, and J. J. Rossi, "Aptamer-based therapeutics: new approaches to combat human viral diseases," Pharmaceuticals, vol. 6, no. 12, pp. 1507–1542, 2013. http://dx.doi.org/10.3390/ph6121507

[6] H.-K. Cheong, E. Hwang, C. Lee, B.-S. Choi, and C. Cheong, "Rapid preparation of RNA samples for NMR spectroscopy and X-ray crystallography," *Nucleic acids research*, vol. 32, no. 10, pp. e84–e84, 2004. http://dx.doi.org/10.1093/nar/gnh081

[7] R. M. Dirks, J. S. Bois, J. M. Schaeffer, E. Winfree, and N. A. Pierce, "Thermodynamic analysis of interacting nucleic acid strands," *Siam Review*, vol. 49, no. 1, pp. 65–88, 2007. http://dx.doi.org/10.1137/060651100

[8] R. M. Al-Khatib, R. Abdullah, and N. Rashid, "A comparative taxonomy of parallel algorithms for RNA secondary structure prediction," *Evolutionary bioinformatics online*, vol. 6, p. 27, 2010. http://dx.doi.org/10.4137/EBO.S4058

[9] K. C. Wiese and E. Glen, "jViz. RNA-an interactive graphical tool for visualizing RNA secondary structure including pseudoknots," *in Computer-Based Medical Systems, 2006. CBMS 2006. 19th IEEE International Symposium on. IEEE*, 2006, pp. 659–664.

[10] Y. Byun and K. Han, "Pseudoviewer3: generating planar drawings of large-scale RNA structures with pseudoknots," *Bioinformatics*, vol. 25, no. 11, pp. 1435–1437, 2009. http://dx.doi.org/10.1093/bioinformatics/btp252

[11] M. Antczak, T. Zok, M. Popenda, P. Lukasiak, R. W. Adamiak, J. Blazewicz, and M. Szachniuk, "Rnapdbeea webserver to derive secondary structures from pdb files of knotted and unknotted RNAS," *Nucleic acids research*, vol. 42, no. W1, pp. W368–W372, 2014. http://dx.doi.org/10.1093/nar/gku330

[12] M. Zuker and P. Stiegler, "Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information," *Nucleic acids research*, vol. 9, no. 1, pp. 133–148, 1981. http://dx.doi.org/10.1093/nar/9.1.133

[13] J. Reeder and R. Giegerich, "Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics," *BMC bioinformatics*, vol. 5, no. 1, p. 104, 2004. http://dx.doi.org/10.1186/1471-2105-5-104

[14] J. Ruan, G. D. Stormo, and W. Zhang, "An iterated loop matching approach to the prediction of RNA secondary structures with pseudoknots," *Bioinformatics*, vol. 20, no. 1, pp. 58–66, 2004. http://dx.doi.org/10.1093/bioinformatics/btg373

[15] J. Ren, B. Rastegari, A. Condon, and H. H. Hoos, "Hotknots: Heuristic prediction of RNA secondary structures including pseudoknots," *RNA*, vol. 11, no. 10, pp. 1494–1504, 2005. http://dx.doi.org/10.1261/rna.7284905

[16] P. Heller, H. J. Tripp, K. Turk-Kubo, and J. P. Zehr, "Arbitrator: a software pipeline for on-demand retrieval of auto-curated nifh sequences from Genbank," *Bioinformatics*, p. btu417, 2014. http://dx.doi.org/10.1093/bioinformatics/btu417

[17] I. Brierley, S. Pennell, and R. J. Gilbert, "Viral RNA pseudoknots: versatile motifs in gene expression and replication," *Nature Reviews Microbiology*, vol. 5, no. 8, pp. 598–610, 2007. http://dx.doi.org/10.1038/nrmicro1704

[18] V. Thiel, K. A. Ivanov, A´ . Putics, T. Hertzig, B. Schelle, S. Bayer, B. Weißbrich, E. J. Snijder, H. Rabenau, H. W. Doerr et al., "Mechanisms and enzymes involved in SARS coronavirus genome expression," *Journal of General Virology*, vol. 84, no. 9, pp. 2305–2315, 2003. http://dx.doi.org/10.1099/vir.0.19424-0

[19] R. B. Lyngsø and C. N. Pedersen, "RNA pseudoknot prediction in energybased models," *Journal of computational biology*, vol. 7, no. 3-4, pp. 409–427, 2000. http://dx.doi.org/10.1089/106652700750050862

[20] T. Akutsu, "Dynamic programming algorithms for rna secondary structure prediction with pseudoknots," *Discrete Applied Mathematics*, vol. 104, no. 1, pp. 45–62, 2000. http://dx.doi.org/10.1016/S0166-218X(00)00186-4

[21] W. Gerlach and R. Giegerich, "Guugle: a utility for fast exact matching under RNA complementary rules including G–U base pairing," *Bioinformatics*, vol. 22, no. 6, pp. 762–764, 2006. http://dx.doi.org/10.1093/bioinformatics/btk041

[22] (2015) Bibiserv - the bielefeld university bioinformatics server. [Online]. Available: *http://bibiserv.techfak.uni-bielefeld.de/*.

[23] (2015) Unafold - the Unafold web server at (the RNA institute. University at Albany, State University of New York.). [Online]. Available: *https://www.rna.albany.edu/*.

[24] D. H. Mathews, "Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization," *RNA*, vol. 10, no. 8, pp. 1178–1190, 2004. http://dx.doi.org/10.1261/rna.7650904

[25] ——, "Predicting a set of minimal free energy RNA secondary structures common to two sequences," *Bioinformatics*, vol. 21, no. 10, pp. 2246–2253, 2005. http://dx.doi.org/10.1093/bioinformatics/bti349

[26] P. Kerpedjiev, S. Hammer, and I. L. Hofacker, "*forna* (force-directed RNA): Simple and effective online RNA secondary structure diagrams," 2015.

[27] F. Van Batenburg, A. P. Gultyaev, C. Pleij, J. Ng, and J. Oliehoek, "Pseudobase: a database with RNA pseudoknots," *Nucleic Acids Research*, vol. 28, no. 1, pp. 201–204, 2000. http://dx.doi.org/10.1093/nar/28.1.201

[28] J. J. Cannone, S. Subramanian, M. N. Schnare, J. R. Collett, L. M. D'Souza, Y. Du, B. Feng, N. Lin, L. V. Madabusi, K. M. M¨uller et al.," The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAS," *BMC bioinformatics*, vol. 3, no. 1, p. 2, 2002. http://dx.doi.org/10.1186/1471-2105-3-2

[29] S. V. Angiuoli, M. Matalka, A. Gussman, K. Galens, M. Vangala, D. R. Riley, C. Arze, J. R. White, O. White, and W. F. Fricke, "Clovr: a virtual machine for automated and portable sequence analysis from the desktop using cloud computing," *BMC bioinformatics*, vol. 12, no. 1, p. 356, 2011. http://dx.doi.org/10.1186/1471-2105-12-356

[30] (2015) Network computing server & simulator at the university of western australia. [Online]. Available: *http://www.uwa.edu.au/*.

[31] X. Chen, S.-M. He, D. Bu, F. Zhang, Z. Wang, R. Chen, and W. Gao, "Flexstem: improving predictions of RNA secondary structures with

pseudoknots by reducing the search space," *Bioinformatics*, vol. 24, no. 18, pp. 1994–2001, 2008. http://dx.doi.org/10.1093/bioinformatics/btn327

[32] C.-H. Huang, C. L. Lu, and H.-T. Chiu, "A heuristic approach for detecting RNA h-type pseudoknots," *Bioinformatics*, vol. 21, no. 17, pp. 3501–3508, 2005. http://dx.doi.org/10.1093/bioinformatics/bti568

[33] (2015) Pseudobase: is a collection of RNA pseudoknots & their real secondary structures. [Online]. Available: *http://www.ekevanbatenburg.nl/PKBASE/PKB.HTML*.

[34] P. Baldi, S. Brunak, Y. Chauvin, C. A. Andersen, and H. Nielsen, "Assessing the accuracy of prediction algorithms for classification: an overview," *Bioinformatics*, vol. 16, no. 5, pp. 412–424, 2000. http://dx.doi.org/10.1093/bioinformatics/16.5.412

[35] H. Zhao, Y. Yang, and Y. Zhou, "Structure-based prediction of RNA binding domains and RNA-binding sites and application to structural genomics targets," *Nucleic acids research*, p. gkq1266, 2010.

[36] (2015) The Genbank database. [Online]. Available: *http://www.ncbi.nlm.nih.gov/genbank/*.

Ahmad Habboush is an associate professor in the Department of Computer Science, Jerash Private University, Jerash, Jordan. He received his BS, MS and PhD degree in Computer Engineering from Kharkov State Poly-technical University, Kharkov, Ukraine. His research interest includes security, parallel processing, networking & mobile computing, artificial intelligence and software engineering.

Mohammad Al Rawajbeh is an associate professor at the Computer Network Department, Faculty of Science and Information Technology, AL Zaytoonah University of Jordan. Received the High Diploma Degree in Computer Technology from Yarmouk University, Jordan in 1991. Received B.SC and Master Degree in computers and networks from Kharkov National Polytechnic University, Ukraine from 1992-1996. Obtained his Ph.D degree in Computer Network Management from Kharkov National University of Radio-Electronic, Ukraine. His research interests are in the fields of network management, performance evaluation, and internet technology.

Ahmed M. Manasrah is an assistant professor at the Faculty of Information Technology and Computer Sciences, Yarmouk University, Irbid, Jordan since 2010. Dr. Manasrah served as the Deputy Director (Research and Innovation) and the Head of iNetmon project at the National Advanced IPv6 Centre (NAv6) in Universiti Sains Malaysia. Since 2009. Dr. Manasrah started his career as a web developer at Telaterra LLC from 2003 until 2004. Between 2005 until 2008, Dr. Ahmed worked as senior research officer in NRG, Universiti Sains Malaysia. He is interested in network monitoring and security.

Ra'ed M. Al-Khatib is a senior lecturer in the Department of Computer Science, Faculty of Information Technology and Computer Science, Yarmouk University, Irbid 21163, Jordan. He received his BS degree in Computer Sciences from Mut'ah University in 1993, his MS degree in Computer Engineering from Yarmouk University in 2006, and his PhD degree in Computer Sciences from Universiti Sains Malaysia, Penang, Malaysia in 2012. His research interest includes high parallel computing, computer networks, artificial intelligence, computer algorithms and software engineering.