A Review on RNA Secondary Structure Prediction Algorithms

Ian Kings Oluoch
Department of Computer Engineering
Gazi University
Ankara,Turkey
ian.kings.oluoch@gazi.edu.tr

Abdullah Akalın
Department of Computer Engineering
Gazi University
Ankara,Turkey
abdullah.akalin@gazi.edu.tr

Yılmaz Vural
Turkish Data Protection Authority
Ankara,Turkey
yilmazvural@gmail.com

Yavuz Canbay
Department of Computer Engineering
Gazi University
Ankara,Turkey
yavuzcanbay@gazi.edu.tr

***Abstract*— Undoubtedly, RNA has vital functions on organisms. As a single stranded nucleic acid, it tends to bend and twirl and it forms a stable structure of itself. This is what comes to be known as the RNA secondary structure. RNA secondary structure is of use in determining the functionalities of RNA sequences as well as in pharmaceutical developments. Furthermore, predicting the secondary structure of an RNA is a crucial step in predicting its tertiary structure, which is its three dimensional form. For the problem of RNA secondary structure, researchers have developed many algorithms over the years in an attempt to make accurate predictions. In this paper, we review some of the recent works that made use of artificial intelligence in RNA secondary structure prediction. The reviewed articles show that by the power of novel artificial intelligence methods and ensembles of the old techniques, there are promising outcomes for future research.**

Keywords—RNA secondary structures, secondary structure prediction, artificial intelligence, RNA

I. INTRODUCTION

DNA stands for Deoxyribonucleic Acid. It is the molecule that comprises our biological codes which correspond to our idiosyncrasy, thus making everyone of us unique.

DNA is made up of certain building blocks called nucleotides. There are four kinds of nucleotides; Adenine (A), Thymine (T), Guanine (G) and Cytosine (C). DNA codes the organisms using this four lettered alphabet. Nucleotides come together as a sequence. Sequences which contain information about synthesis of some kind of proteins are called genes. The whole collection of genes of an organism is its genome.

DNA can unwind when necessary, that is, when it starts a replication or when protein synthesis occurs.

Protein synthesis occurs with the help of three RNA types. Messenger RNA (mRNA) transcripts the genetic information, that is a portion of the DNA, and transfers it to the ribosome, which is outside the nucleus. Then begins the translation by Transfer RNA (tRNA) and ribosome (in which Ribosomal RNAs (rRNA) exist as a construction element). After these procedures, an amino acid chain is constructed and is then referred to as protein [1], [2], [3], [4].

Genomic data, or DNA/RNA sequences, can be represented as sequences of A, T/U, G and C symbols. These sequences can be considered to be a string made up of these letters [5].

In this paper, we investigate some current approaches to the RNA secondary structure prediction problem.

We propose a review of some recent works that have been done on the RNA secondary structure problem. Srikamdee et al. have modelled this problem as a combinatorial optimization problem. In their method, they have found all possible combinations of RNA secondary structures which have the minimum free-energy. To choose the optimum one, they have adopted an evolutionary algorithm called COIN [6]. They have measured their success using specificity, sensitivity and f-measure metrics [7]. Jetlin et al. have developed a tries data structure based RNA secondary structure prediction. In their work, they have combined genetic algorithm with tries and measured their success with sensitivity metric [8]. A heuristic approach has been proposed by El Fatmi et al. In their work, they have developed an ensemble of generic programming (GA) and greedy randomized adaptive search procedure (GRASP). The proposed work also focuses on pseudo-knotted structures too [9].

This paper is organized as follows: In Section II, necessary genomic background is presented for the understanding of the RNA secondary structure problem. In the same section, performance evaluation measures used in the vast majority of the RNA secondary structure prediction problems are briefly listed. In Section III, several general programming methods are discussed. After reviewing some algorithms that have been in existence in literature for a period of over seven years in Section IV, Section V delves into recent works exploiting artificial intelligence techniques in greater detail. In Section VI, works that are reviewed in V are compared in suitable metrics and Section VII concludes the paper.

II. RNA SECONDARY STRUCTURE PREDICTION PROBLEM

RNA has three types of structures. The first one is the strings of nucleotides in their straight forms. Like AUGCGAUAGCUCGAU... While this structure serves as a means of distinguishing one RNA from another, it conveys little information about its functionality. For this reason, knowing their two dimensional representation gives some insight into the activities they perform like in gene expressions

or as catalysors. RNA secondary structure also shed light on its three dimensional tertiary structure [8].

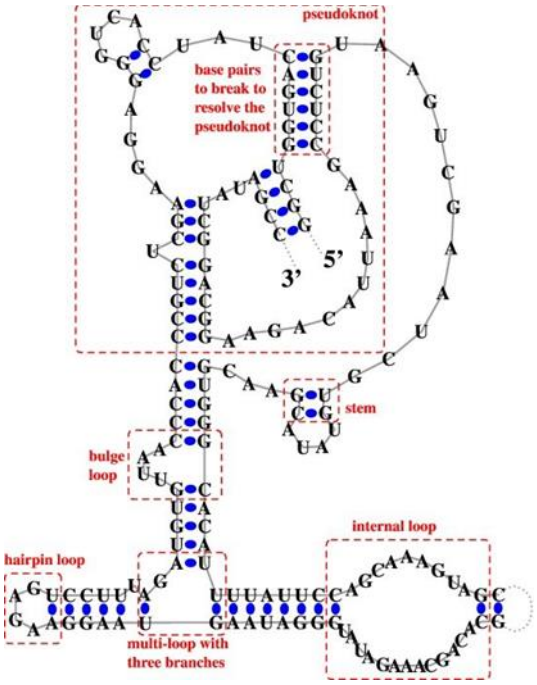

Fig. 1. Visual elements of RNA secondary structure explained in II. (Figure is taken from RNA STRAND v2.0 home page [11])

The problem is predicting from a first order RNA sequence, its second order representation. The problem is considered to be NP-hard. RNA strand length is an important factor on prediction accuracy, generally which gets lower with the increase of molecule size. Another challenge in this problem is pseudo-knots, which makes the problem more complex and RNA strands hard to predict [10].

Second order RNA structure have some visual elements such as stems, hairpin loops, interior loops, bulges and single lines which are presented in Fig. 1.

A stem appears when two or more nucleotides are paired consecutively. Hairpin loops occur when there is a paired couple which has consecutive unpaired nucleotides between two ends of a coupled nucleotide. Interior loops can be thought of as loops that are between two stems or a hairpin loop that has two sides. When there exist some unpaired elements between two nucleotides which are otherwise paired in the stem fashion, a bulge occurs. Single stranded lines are straightforward. Finally, elements which are not paired with any other elements form a single strand [5].

For this problem, there are several physical methods that provide a solution. X-Ray Crystallography and Nuclear Magnetic Resonance are mentioned by most researchers but they are costly and time consuming enough for them to seek computational methods as alternatives [12]. There are three main techniques used in RNA secondary structure prediction, mentioned in [13]. Graphically brute-forcing is one of them, which is not very popular as it becomes less practical with the RNA sequence growth. Another is exploiting thermodynamics, by trying to find a structure that has the minimum free-energy. The last approach is phylogeny focused. If two or more primary RNA structures are similar, then they are likely to be related in their secondary structures and functions too, allowing them to be generalized to one upper-class secondary structure.

Additionally, just as in most prediction problems, the RNA secondary structure prediction problem inclusive, results of the works proposed by the authors are almost always subjected to comparison. In doing so, one assesses a success relative to another. For a healthy comparison to be done, it is crucial for the researchers to provide comparable results with common metrics.

As for the prediction metrics, in which TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative) are affiliated both as numerators or as denominators, we are interested in the ”sensitivity”, ”specificity”, ”accuracy”, ”positive predictive value”, ”Matthews correlation coefficient” and ”negative predictive value” measures.

Sensitivity (SN), or true positive rate, or recall is defined as:

$$SN = \frac{\mathrm{TP}}{TP + FN}$$

Where specificity (SP) or true negative rate is defined as:

$$SP = \frac{\mathrm{TN}}{\mathrm{FP} + \mathrm{TN}}$$

Accuracy (ACC) is defined as:

$$ACC = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}}$$

Positive predictive value $PPV = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$ and negative predictive value $NPV = \frac{\mathrm{NT}}{\mathrm{FN+TN}}$ respectively.

As can be seen on the above equations, none of the SN, SP, PPV or NPV employ all four kinds of the cases stated further above, namely TP, TN, FP and FN. For a more representative measure Matthews correlation coefficient (MCC) is defined as [14],

$$MCC = \frac{\mathrm{TP}.\mathrm{TN} - \mathrm{FP}.\mathrm{FN}}{(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TP} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}$$

The reviewed articles on RNA secondary structure problem in this work mostly use these measures for assessing a proposed algorithm’s success. Hence, this foundation will help the reader interpret the comparisons made.

III. RNA Secondary Structure Prediction Techniques and Programming Methods

In this section, some techniques and programming methods on RNA secondary structure prediction are summarized.

A. Genetic Algorithms

Genetic algorithms are a kind of evolutionary algorithms. In general, genetic algorithms comprise a population which is subject to genetic variance by means of mutation and a selection phase. This is implemented by a programmer as desired for example ”survival of the smallest”, ”biggest” or ”fittest” [15].

A genetic algorithm starts from 0^{th} generation and continues through generations until a termination condition is met. In every generation, some genetic transitions are done and population is filtered by selection

functions. The final result is a population with desired qualities honed and polished [15].

B. Dynamic Programming

Dynamic programming is a programming technique that is used to solve optimization problems. Optimization problems are those in which an optimum solution is sought among numerous other solutions. To apply dynamic programming to a problem, the problem needs to satisfy two conditions: having optimal substructure and overlapping sub-problems [16].

Optimal substructure condition is met when an optimal solution to a problem also solves the sub-problems of that particular problem while overlapping sub-problems condition is met when a recursive algorithm for a problem solves the same sub-problems on every recursion in lieu of producing new ones. By this condition, dynamic programming benefits the sameness of each of the sub-problems by computing them once and recording the results in a table for further look up, which will take constant time when looking up [16].

To solve a dynamic programming applicable problem, first thing to consider is to learn the characteristics of an optimal solution's structure. Then comes finding a definition of an optimal solution's value in recursive manner and computing it. The final step is the construction of an optimal solution with the values computed [16].

C. Greedy Randomized Adaptive Search Procedure

GRASP is an iterative randomized sampling technique where each iteration provides a solution to the problem at hand in this case the prediction of RNA secondary structures. This process is made up of two phases namely: (1) Construction phase (2) Local search phase.

In the construction phase a workable solution is iteratively constructed and the next item to be added is obtained by listing all potential candidates in a list through a greedy function. The most suitable candidate is selected from this list at random although it does not have to be the top candidate, hence allowing various solutions to be found at each iteration. The local search phase similarly works in iterations which replace a current solution by an enhanced solution in the neighborhood of the current solution. In the event that a better solution can no longer be obtained in the neighborhood then this becomes the terminal point of the process.

The quality of solutions obtained by implementing the GRASP process should in principle be relatively satisfactory given that the technique samples all potentially feasible solutions and gives the best as a final result according to Feo and Resende [17]. This makes GRASP a practical technique that can be applied in RNA secondary structure prediction.

IV. RNA Secondary Structure Prediction Algorithms

Some fundamental algorithms on RNA secondary structure prediction are given below.

A. TT2NE

The main idea behind the TT2NE [18] algorithm is its "genus" concept in classifying pseudo-knot topology which allows the algorithm to find secondary structures with pseudo-knots in any kind of topology. The algorithm uses minimum free energy model in determining an optimum secondary structure.

The algorithm is based on maximum weighed independent set formalism (WIS for short). With this point of view, an RNA structure consists of stem-like structures which the authors name them as "helipoints" [18].

TT2NE algorithm firstly builds a graph of helipoints. This graph is weighted and a vertex represents a helipoint. A link between two vertices is formed if and only if those two vertices do not appear in the same secondary structure. Every vertex weighs -1 times its free energy formation. Ultimately the purpose is to find a set in the graph where weights are maximum resulting in a minimum overall free energy.

The genus concept in the proposed work is an additive integer number which denotes the complexity of a pseudo-knot. TT2NE algorithm also offers a tunable parameter to limit maximum genus number (which in turn means pseudo-knot complexity) allowed to improve the performance which will limit the possible pseudo-knot space. Authors also have proposed an optimum genus value of 3 for small RNAs such as RNAs of 250 nucleotides or less.

This proposed algorithm has been compared with McQFold [19], HotKnots [20], ProbKnots [21] and Mfold [22] RNA secondary structure prediction algorithms.

As a result concluded by the authors, the genus penalty improves the success of predicting secondary structures with pseudo-knots and it has an average sensitivity value of 82% alongside an average positive predictive value of 79% on a subset of sequences used in [20]. The algorithm, as stated by the authors, is better than most of the state-of-the-art algorithms.

B. IPknot

This is a computational method that aids in the prediction of RNA secondary structures with pseudo-knots which are usually formed by base pairing of unpaired bases of a loop and those existent outside the loop. It is based on Integer Programming where Pseudo-knotted structures are decomposed into a set of substructures that are free from pseudo-knots. At the same time base pairing probability distribution is approximated to allow the modelling of pseudo-knots [23]. Several methods have been proposed for the prediction of RNA pseudo-knotted structures with only a few providing promising results in terms of the accuracy and speed of the algorithm.

The main advantage of IPknot over the other methods is that it provides quick accurate results despite alignment quality deterioration in RNA. RNA structures are predicted by simply calculating the base pairing probabilities used in the IP objective function and then solving the IP problem [23]. The threshold cut technique is what gives IPknot its speed.

Even though IPknot produces very accurate results when used in the prediction of RNA secondary structures, it only does so when provided with long sequences of RNA. Shorter sequences produce erroneous results that may be worse than the other previously proposed methods of RNA secondary structure prediction. To solve this decrease in accuracy when the algorithm is fed with short RNA sequences, the iterative refinement algorithm (a heuristic algorithm for refining base-pairing probabilities) is used thereby improving the base-pair probabilities.

C. CycloFold

CycloFold [24], is also not restricted in pseudo-knot complexity similar to IV-A. The proposed algorithm takes an input of unfolded sequence and adds substructures to it iteratively, eventually building an RNA secondary structure with minimum free energy.

Firstly, it generates all possible helices that contain 4 or more base pairs according to Watson-Crick [4] and GU-wobble base pairs matching. It then runs 50 simulations to determine the helix with the minimum free energy. In each simulation the candidate substructures are selected from the stem list (generated at the first step) with Boltzmann-weighted probability.

The proposed algorithm has been compared with HotKnots 2.0 [25], pknotsRG [26] and UNAFold [27]. For one of the datasets used in the proposed work, CycloFold had a 83% of MCC whereas pknotsRG, HotKnots 2.0 and UNAFold had MCCs of 82%, 75% and 73% respectively.

Ultimately, providing advantage of no restriction of pseudo-knot complexity, CycloFold shows an MCC prediction success that is comparable to pknotsRG.

D. RNAStructure

RNAStructure is a software package that is also used in RNA secondary structure prediction. It implements free energy minimization structure prediction algorithm based on the nearest-neighbor rule that predicts the stability of a structure as quantified by folding free energy change [28]. The single most probable structure in a folding ensemble is obtained by finding the lowest free energy structure.

This software is meant to make obtaining results of RNA structure prediction from a sequence of RNA bases easier for a user. A user-friendly graphical user interface makes the process less cumbersome for a user than other software that instead have a command line interface. The algorithm for structure prediction is implemented by C++ libraries that are relatively faster compared to other programming languages.

E. SARNA-Predict

This is an algorithm used in the prediction of RNA secondary structures based on Simulated Annealing (SA). According to [29], this process (SA) mimics the process of heating a given material and then cooling it down slowly into a uniform structure. The change in energy between a new and current structure allows SARNA-Predict to either accept or reject a new structure that has just undergone SA. SARNA-Predict employs the INN-HB thermodynamic model but can also have this substituted for the more robust efn2 model that yields better results in terms of accuracy compared to the dynamic programming algorithm mfold implemented in RNA secondary structure prediction [29]. Despite this algorithm producing more accurate predictions, it is limited to the thermodynamic model used in its implementation. When the INN-HB model is used, a deterioration in accuracy is observed with an increase in the length of the RNA sequence. Herbert and Kay [29] in their work suggested that the SARNA-Predict algorithm could in principle be used in prediction of pseudo-knotted structures if an appropriate thermodynamic model with a high accuracy rate is included in the process.

V. Recent works on RNA Secondary Structure Prediction

In order to see the current standing of prediction methods on RNA secondary structure, three recent works have been selected and are reviewed below.

A. A Heuristic Method

El Fatmi et al. [9] have proposed an algorithm for predicting RNA secondary structures containing pseudo-knots. In their work, they combined GRASP and GA (Genetic Algorithm). Their proposed algorithm takes an RNA sequence as input and generates initial population by finding all possible stems. It then calculates the free energy and optimizes the solution to find the minimum. After the optimization, it returns the best solution, which is the secondary structure of the sequence.

They have implemented their algorithm in Java programming language and measured its efficiency by sensitivity and specificity measures. The proposed algorithm produced better performance compared to TT2NE, IPknot, CycloFold and RNAStructure with SP percentage of 100 and SN of 73% on CcTMV sequence, SP of 100% and SN of 100% on Hs PrP sequence.

B. An Evolutionary Method

Srikamdee et al. [7] have modelled this problem as a combinatorial optimization problem. In their work, they have put emphasis on the low average values of sensitivity and specificity of the current state-of-the-art techniques used. They also underline the slowness of some of the current algorithms which tend to find the secondary structure of an RNA through permutating all possible RNA helices. The authors therefore proposed an algorithm that only selects suitable helices during the prediction and deals with a combinatorial optimization problem. They adopted an evolutionary algorithm named, Coincidence Algorithm (COIN) [6] to effectively solve this problem.

COIN stands for Combinatorial Optimization with Coincidence. This algorithm is a genetic algorithm that exploits the bad solutions as well as the good ones in the populations. Optimizing the solutions both by good and bad inferences, it eventually finds an optimum solution. The traditional deduction in the population evaluation phase of genetic algorithms is to select candidates with good solutions for the next generations. COIN algorithm however also makes use of the bad solutions for optimization by preventing undesired results [6].

Table 1: Summarization of the reviewed recent works.

Paper No.	Algorithm	Sequences and Scores	Evaluation
[9]	GRASP+GA	On CcTMV sequence, SP: 100%; SN: 73. On Hs PrP sequence, SP: 100%; SN:100%	Allows pseudo-knots Outperformed TT2NE, Ipknot, CycloFold and RNAStructure
[7]	COIN	On S.cerevisiae, H.marismortui, M.anisopliae, A.lagunensis, H.rubra, A.graffini, C.elegans,	Does not allow pseudo-knots. Outperformed RnaPredict and SARNA-predict.

		D.virilis, X.laevis sequences. Avg. SP:57.2%; Avg. SN: 55.5%	
[8]	TRIES+GA	On S.cerevisiae, H.marismortui, M.anisopliae, A.lagunensis, H.rubra, A.graffini, C.elegans, D.virilis, X.laevis, (H.sapiens and A.fulgens) sequences. Avg. SN: 55.38% (52.05%)	Does not allow pseudo-knots. Score close to [7].

The proposed COIN algorithm consists of 6 steps [7]. First, generator matrix, in which the dual probabilities of helices are written is produced. One cell of the matrix gives information about the probability of the corresponding two helices' appearance in company.

The next step is to generate the population. Until a specified population size is obtained, this step loops over and over to generate candidate structures by selecting a random helix from the generator matrix, and adding it to the population if it is a compatible helix. That is, in the generator matrix the probability is not zero. After this step, comes the evaluation of the population, in which candidates are assessed according to Individiual Nearest-Neighbor Hydrogen Bond energy model. This model is used to calculate the free-energy. Evaluation is made on the basis of the minimum free-energy.

Selecting candidates is the step in which the population is put into two classes, goods and poors. By survival of the fittest principle, first g% candidates are counted as goods which will be used to find the solutions that are desirable, and the last p% as poors. After this processing, the generator matrix is updated. The information gained from the previous step is used to increase the probability of the good solutions on the generator matrix for they form more stable structures. Likewise, probabilities of the poor solutions on the generator matrix is decreased for them to be less likely in the model.

The proposed work has been tested on 10 RNA sequences which are derived from RNA STRAND v2.0 [30]. RNA STRAND is an RNA secondary structure database ready for evaluating computational results of algorithms. This database also offers tools for searching and analyzing RNA data items as well as downloading them. It allows software to score themselves on this carefully selected and collected database. It is also possible to search for certain motifs through RNA secondary structure data. This database is built upon publicly available sources and its specialization on RNA secondary structure is one of its appealing features [30].

For the proposed algorithm, the population size is defined to be 100, the number of generations for the evolution is set to be 100 too. Percentage for the good and poor solutions are both selected as 0.2, while the learning rate of the algorithm is 0.01, the fitness function is finding the minimum free energy via INN-HB.

The results are evaluated on the sensitivity, specificity and f-measure metrics. On sensitivity metric, the proposed algorithm is better than RnaPredict by 3%, and better than SARNA-predict by 5%. On specificity, the proposed method is better than RnaPredict and SARNA-predict by 7% and 8% respectively. For the f-measure, the proposed algorithm is better than the two by 5% and 6% respectively. The average sensitivity, specificity and f-measure values for the proposed algorithm are 55.5%, 57.2% and 56.2% respectively.

C. A Tries Based Genetic Programming Method

This method of RNA secondary structure prediction proposes the combination of GA (Genetic Algorithm) and the Tries based method in the prediction of the most accurate RNA secondary structure. According to [8] GA is based on natural selection where the traits that promote the survival of an organism (the fittest) are the ones that become common in later generations of a population. It is based on selection, crossover and mutation making it efficient for searches in a wide scope that renders other methods impractical [8]. When either binary or real values representation is applied to convert data into a chromosome then the GA method begins leaning more towards application of Artificial Intelligence in problem solving (use of data or examples for problem solution). [8] Suggests that first an array is used to represent stems with the binary values 0 and 1 indicating the presence or absence of a stem. Thereafter binary GA is applied in selection, crossover and finally mutation. An increment of the initial string (array) occurs with each iteration of the procedure. It further suggests that for improved performance of the algorithm the values for stem incrementing mutations and stem deleting mutation are formulated by squaring fitness for a larger stem.

To further understand how GA works, the following was provided by [8]. The creation of an initial population followed by an evaluation of each offspring score after application of a fitness function. In each generation the subjects with the lowest score are eliminated and replaced with newly generated ones (these are referred to as children). For the selection of a child mutation or crossover is used. A parent is selected from a modified alignment taking into consideration the population with no duplicates. Weighted wheel selection is used to pick the fittest parent that is allowed to move over to the next generation.

The proposed method has been analyzed on various organisms. As a result the average sensitivity of the proposed method has been recorded to be 52.05%.

VI. Comparison of Methods

The most recent of the three works investigated here is [9], which supports RNA secondary structures with pseudo-knots. This method has been tested on sequences differing from those used in the other two. Because of this, a neat comparison cannot be done between [9] and the others. However, its success over TT2NE, IPknot, CycloFold and RNAStructure algorithms are remarkable. It has outperformed its counterparts on almost all sequences (CcTMV, Hs PrP, TMV, GCMMV). The significant advantage of [9] over the other two [8] and [7], is it supporting pseudo-knotted structures.

For [7], it has its originality in using COIN algorithm and modeling the problem as combinatorial optimization

problem. It has the average sensitivity of 55.5% on 10 organisms which are also used in [8]. The sequences are of S.cerevisiae, H.marismortui, M.anisopliae, A.lagunensis, H.rubra, A.griffini, C.elegans, D.virilis, X.laevis. [8] has the average sensitivity of 55.38% which is almost the same as that of [7] on the same sequences. However, [8] also tested its proposed algorithm on two more sequences, H.sapiens and A.fulgens, ultimately resulting in 52.05% of average sensitivity. The comparison is summarized in Table 1.

VII. Conclusion

In conclusion, novel methods for RNA secondary structure prediction are still sought for better accuracy and better performance. Pseudo-knotted structures are a challenge to overcome and lengthy RNA molecules secondary structure are yet to be predicted efficiently in shorter times.

Since RNA secondary structure helps predicting RNA tertiary structure, this problem's importance and its gains are twofold. With the growth in both quality and quantity of RNA secondary structure reference datasets, higher prediction accuracies and more reliable results are expected be obtained, which also necessitates extra labor with expensive techniques.

References

[1] N. H. G. R. Institute, "Deoxyribonucleic acid (dna) fact sheet - national human genome research institute (nhgri)." https://www.genome.gov/25520880/ Last access: April 2018.

[2] R. Society, "What is rna?." https://www.rnasociety.org/about/what-is-rna/ Last access: April 2018.

[3] L. H, B. A, and Z. S. et al., Molecular Cell Biology, 4e. New York: W. H. Freeman, 2000. Section 4.4 The Three Roles of RNA in Protein Synthesis.

[4] J. D. Watson and F. H. Crick, "A structure for deoxyribose nucleic acid," Nature, vol. 171, pp. 737–738, April 1953.

[5] E. W. Steeg, "Neural networks, adaptive optimization, and rna secondary structure prediction," Artificial intelligence and molecular biology, pp. 121–160, 1993.

[6] W. Wattanapornprom, P. Olanviwitchai, P. Chutima, and P. Chongstitvatana, "Multi-objective combinatorial optimisation with coincidence al- gorithm," in Evolutionary Computation, 2009. CEC'09. IEEE Congress on, pp. 1675–1682, IEEE, 2009.

[7] S. Srikamdee, W. Wattanapornprom, and P. Chongstitvatana, "Rna secondary structure prediction with coincidence algorithm," in 2016 16th International Symposium on Communications and Information Technologies (ISCIT), pp. 686–690, Sept 2016.

[8] [8] C. P. Jetlin, L. J. Anbarasi, R. Neeraja, and M. Narendra, "Tries based rna structure prediction," in 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), pp. 1–4, Dec 2016.

[9] A. E. Fatmi, A. Chentoufi, M. A. Bekri, S. Benhlima, and M. Sabbane, "A heuristic algorithm for rna secondary structure based on genetic algorithm," in 2017 Intelligent Systems and Computer Vision (ISCV), pp. 1–7, April 2017.

[10] A. Condon, "Problems on rna secondary structure prediction and design," in International Colloquium on Automata, Languages, and Programming, pp. 22–32, Springer, 2003.

[11] R. Soft, "Rna strand v2.0 - the rna sec- ondary structure and statistical analysis database." http://www.rnasoft.ca/strand/secondary structure large.jpg Last access: May 2018.

[12] M. N. Osman, R. Abdullah, and N. AbdulRashid, "Rna secondary structure prediction using dynamic programming algorithm - a review and proposed work," in 2010 International Symposium on Information Technology, vol. 2, pp. 551–556, June 2010.

[13] M. Zuker and D. Sankoff, "Rna secondary structures and their prediction," Bulletin of Mathematical Biology, vol. 46, no. 4, pp. 591 – 621, 1984.

[14] M. Vihinen, "How to evaluate performance of prediction methods? measures and their interpretation in variation effect analysis," in BMC genomics, vol. 13, p. S2, BioMed Central, 2012.

[15] T. Back, Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms. Oxford university press, 1996.

[16] T. H. Cormen, Introduction to algorithms. MIT press, 2009.

[17] [17] T. A. Feo and M. G. Resende, "Greedy randomized adaptive search procedures," Journal of global optimization, vol. 6, no. 2, pp. 109–133, 1995.

[18] M. Bon and H. Orland, "Tt2ne: a novel algorithm to predict rna secondary structures with pseudoknots," Nucleic acids research, vol. 39, no. 14, pp. e93–e93, 2011.

[19] D. Metzler and M. E. Nebel, "Predicting rna secondary structures with pseudoknots by mcmc sampling," Journal of mathematical biology, vol. 56, no. 1-2, pp. 161–181, 2008.

[20] 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.

[21] S. Bellaousov and D. H. Mathews, "Probknot: fast prediction of rna secondary structure including pseudoknots," Rna, vol. 16, no. 10, pp. 1870–1880, 2010.

[22] M. Zuker, "Mfold web server for nucleic acid folding and hybridization prediction," Nucleic acids research, vol. 31, no. 13, pp. 3406–3415, 2003.

[23] K. Sato, Y. Kato, M. Hamada, T. Akutsu, and K. Asai, "Ipknot: fast and accurate prediction of rna secondary structures with pseudoknots using integer programming," Bioinformatics, vol. 27, no. 13, pp. i85–i93, 2011.

[24] E. Bindewald, T. Kluth, and B. A. Shapiro, "Cyclofold: secondary structure prediction including pseudoknots," Nucleic Acids Research, vol. 38, pp. W368–W372, Jul 2010.

[25] M. S. Andronescu, C. Pop, and A. E. Condon, "Improved free energy parameters for rna pseudoknotted secondary structure prediction," RNA, vol. 16, no. 1, pp. 26–42, 2010.

[26] J. Reeder, P. Steffen, and R. Giegerich, "pknotsrg: Rna pseudoknot folding including near-optimal structures and sliding windows," Nucleic acids research, vol. 35, no. suppl 2, pp. W320–W324, 2007.

[27] N. R. Markham and M. Zuker, "Unafold," in Bioinformatics, pp. 3–31, Springer, 2008.

[28] J. S. Reuter and D. H. Mathews, "Rnastructure: software for rna secondary structure prediction and analysis," BMC bioinformatics, vol. 11, no. 1, p. 129, 2010.

[29] H. H. Tsang and K. C. Wiese, "Sarna-predict: Accuracy improvement of rna secondary structure prediction using permutation-based simulated annealing," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 7, pp. 727–740, October 2010.

[30] M. Andronescu, V. Bereg, H. H. Hoos, and A. Condon, "Rna strand: the rna secondary structure and statistical analysis database," BMC bioinformatics, vol. 9, no. 1, p. 340, 2008.