The Prediction of Tertiary Structure of RNA Based upon Secondary Structure by Using the Relaxation Method and Texture Mapping Method

C. C. Lin and R. C. T. Lee
Department of Computer Science and Information Engineering,
National Chi-Nan University, Puli, Nantou Hsien, 545, Taiwan, ROC
{s2321905, rctlee}@ncnu.edu.tw

ABSTRACT

In this paper, we propose a method to predict tertiary structure of an RNA sequence. Our prediction method would first find the secondary structure of the given RNA sequence. We then use the relaxation method to predict the 3-space structure of it. Experimental results show that our method is quite feasible. We tested our approach on 731 RNA's. 179 of them got a complete match, 391 of them got 90% match and 161 of them got 80% match.

Keywords: RNA, RNA Secondary Structure, RNA Tertiary Structure, RNA Structure Prediction and Relaxation Method.

1. INTRODUCTION

The prediction of tertiary structure of an RNA sequence is important because drug design is sometimes based on the physical structure of an RNA sequence. There are also some papers discussing the prediction of the tertiary structure of an RNA sequence [1]-[9]. These are based on the assumption that similar secondary structures reflect structures that also share similarities in tertiary structure. For a query sequence of RNA, they predict its secondary structure as the first step. There are two types in next steps for secondary structure to tertiary structure of RNA. In type 1, they measure the similarity between the predicted secondary structure and the secondary structures which exist in their data base [1]-[8]. Suppose the predicted secondary structure of an RNA *A* is similar to the secondary structure of a known RNA *B*. Then they conclude that the tertiary structure of RNA *A* is the same as that of RNA *B*. In type 2, given an unknown RNA sequence, they assign initial positions of this sequence randomly to start and adjust the positions by using the simulated annealing algorithm [10],[11] so that the tensional angles and bond lengths coincide as much as possible with those of a known RNA whose sequence is similar to that of the unknown RNA [9].

In both types, they compare an unknown sequence *A* with known sequences and find a known sequence *B* which is similar to this unknown sequence. Then they determine the tertiary structure of *A* based upon that of the known sequence *B*.

In this paper, we present a method to predict the tertiary structure of an RNA sequence. We transform the secondary structure of an RNA into a distance matrix in such a way that the distances reflect the secondary structure. Then, we assign some initial starting points in the 3-space for each nucleotide in the sequence. Finally, we use the relaxation method [12] to move all nucleotides into the 3-space such that their distances in the distance matrix are preserved as much as possible.

2. THE RELATIONSHIP BETWEEN SECONDARY STRUCTURE AND TERTIARY STRUCTURE

An RNA sequence of length *n*, can be represented as $R = r_1 r_2 r_3 \cdots r_n$, where $r_i \in \{A, U, G, C\}$. If r_i and r_j are paired in *R*, we call them a base pair and denote this pair as (r_i, r_j). A secondary structure of *R* is a set *M* of base pairs, such that $M = \{(r_i, r_j) \mid 1 \le i < j \le n\}$. The prediction of the secondary structure can be found in [13]-[26].

We shall call a sequence of base pairs, denoted as $(r_i, r_j)(r_{i+1}, r_{j-1}) \cdots (r_{i+k}, r_{j-k})$ where $1 \le k \le (j - i - 3)/2$, as a segment. Consider Fig. 1, we shall say that this sequence consists of one segment and one turn. Every segment in an RNA sequence itself is in the form of a helical structure. It takes eleven nucleotides to form a complete helix [27].

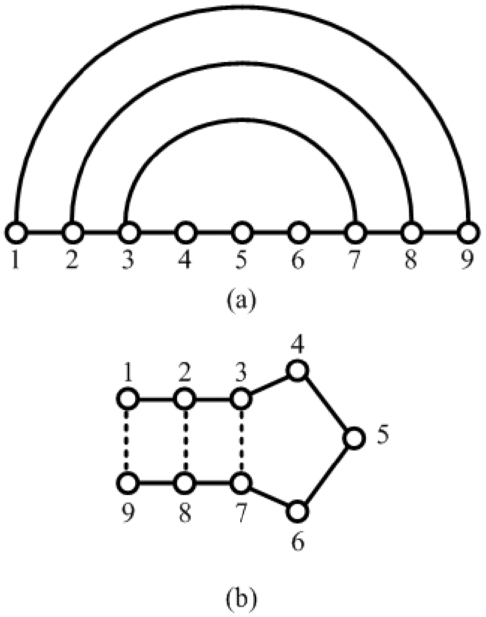

Fig. 1 (a) A secondary structure, (b) The tertiary structure of (a)

Consider the cylinder with radius *r* and height *h* with a line curved on the surface of it as shown in Fig. 2(a). Suppose we spread out the surface of the cylinder in the 2-space. It will be as shown in Fig. 2(b).

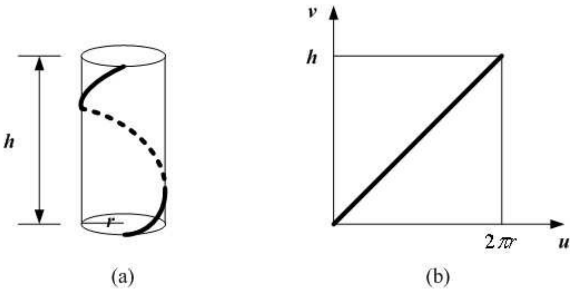

Fig. 2 (a) A cylinder with radius *r* and height *h* with a line curved on its surface, (b) the surface of the cylinder spread out in the 2-space

As we indicated before, a complete helix consists of eleven base pairs. The height *h* of a complete helix takes $34\overset{\circ}{\mathrm{A}}$, and the radius *r* is around $8.5\overset{\circ}{\mathrm{A}}$ [27]. In the surface of cylinder, the angle θ between $\overline{p_{1,k} p_{1,k+1}}$ and *x*-axis is $\tan^{-1}(\frac{h}{2\pi r}) = \tan^{-1}(\frac{34}{17\pi}) = 32.48^\circ$, as shown in Fig. 4-3. Therefore, the length of *L* of RNA is $\frac{h}{\sin\theta} = 63.32\overset{\circ}{A}$. The distance between two adjacent nucleotides is around $\frac{63.32\overset{\circ}{A}}{10} = 6.332\overset{\circ}{A}$.

Given a sequence of total length n, we divide it into two subsequences. For each point $p_{1,k}$ and $p_{2,k}$ where $1 \le k \le \frac{n}{2}$, there are x and y coordinates, denoted as $(x_{1,k}, y_{1,k})$ and $(x_{2,k}, y_{2,k})$ as follows:

$$\begin{aligned} x_{1,k} &= (k-1)(L/10)\cos\theta, \\ y_{1,k} &= (k-1)(L/10)\sin\theta, \\ x_{2,k} &= \pi r + x_{1,k}, \\ y_{2,k} &= y_{1,k}, \\ &\text{where } 1 \le k \le \frac{n}{2}. \end{aligned} \qquad Eq.(1)$$

Consider a point *T*(*u*, *v*) on the 2-space surface of the face of the cylinder. Its 3-space location *P*(*x*, *y*, *z*) denotes as follows[28]:

$$x = r\cos(2\pi \times \frac{u}{2\pi r}),$$
$$y = r\sin(2\pi \times \frac{u}{2\pi r}), \qquad Eq.(2)$$
$$z = v.$$

Let $(X_{a,k}, Y_{a,k}, Z_{a,k})$ denote the position of $p_{a,k}$ in the 3-space, where $1 \le k \le \frac{n}{2}$ and a=1,2. $X_{a,k}, Y_{a,k}$ and $Z_{a,k}$ are found by using Eq.(2) as follows:

$$X_{a,k} = r\cos(\frac{x_{a,k}}{r}),$$
$$Y_{a,k} = r\sin(\frac{x_{a,k}}{r}), \qquad Eq.(3)$$
$$Z_{a,k} = y_{a,k},$$

where $1 \le k \le \frac{n}{2}$ and a=1,2.

3. PRINCIPLE OF OUR PREDICTION METHOD

The basic principle of our prediction method is as follows: Suppose we are given a set of points and suppose we are also given the distances among them. We initially arbitrarily assign these points onto some positions in the 3-space and we use some iterative method to move the points around in such a way that their resulting inter-distances are approaching their real distances. In our case, we use the relaxation method introduced in [12].

The basic idea of the relaxation method presented in [12] is rather straightforward. If two points already mapped are too far away with respect to their distance in the distance matrix, we try to move these two points in such a way to reduce their distance. On the other hand, if two points are too close to each other with respect to their distance in the distance matrix, we try to move these two points in such a way to increase their distance.

Let P_i and P_j denote two points in the 3-space. Let d_{ij} denote the Euclid distance between P_i and P_j. Let D_{ij} denote the distance between P_i and P_j in the desired distance matrix. If $d_{ij} > (<) D_{ij}$, we shorten (enlarge) the distance between P_i and P_j by moving P_i and P_j towards (away) each other along the line connecting P_i and P_j, respectively to P'_i and P'_j.

The above operation is continued until some terminating criteria are met. The final resulting locations constitute the predicted tertiary structure of the RNA sequence.

The above steps are presented mathematically as follows: Let $p_{ix}(p_{jx})$, $p_{iy}(p_{jy})$ and $p_{iz}(p_{jz})$ denote a location of point $i(j)$ in the 3-space. Let $p'_{ix}(p'_{jx})$, $p'_{iy}(p'_{jy})$ and $p'_{iz}(p'_{jz})$ denote a new location of point $i(j)$ in the 3-space. Let d_{ij} denote the Euclid distance between i and j in the 3-space. The formulas for moving the points in the relaxation method are as follows.

$$d_{ij} = \sqrt{(p_{ix} - p_{jx})^2 + (p_{iy} - p_{jy})^2 + (p_{iz} - p_{jz})^2},$$
$$p'_{ix} = p_{ix} - \frac{1}{2}\left[\frac{1 - (D_{ij} / d_{ij})}{1 + D_{ij}^2}\right](p_{ix} - p_{jx}),$$
$$p'_{iy} = p_{iy} - \frac{1}{2}\left[\frac{1 - (D_{ij} / d_{ij})}{1 + D_{ij}^2}\right](p_{iy} - p_{jy}),$$
$$p'_{iz} = p_{iz} - \frac{1}{2}\left[\frac{1 - (D_{ij} / d_{ij})}{1 + D_{ij}^2}\right](p_{iz} - p_{jz}), \qquad Eq.(4)$$
$$p'_{jx} = p_{jx} + \frac{1}{2}\left[\frac{1 - (D_{ij} / d_{ij})}{1 + D_{ij}^2}\right](p_{ix} - p_{jx}),$$
$$p'_{jy} = p_{jy} + \frac{1}{2}\left[\frac{1 - (D_{ij} / d_{ij})}{1 + D_{ij}^2}\right](p_{iy} - p_{jy}),$$
$$p'_{jz} = p_{jz} + \frac{1}{2}\left[\frac{1 - (D_{ij} / d_{ij})}{1 + D_{ij}^2}\right](p_{iz} - p_{jz})$$

for $1 \le i < j \le n$.

4. THE DISTANCES BASED UPON THE SECONDARY STRUCTURE

In the above section, we pointed out that our method is a mapping method. We are given a distance matrix of a set of points. We initially assign the points arbitrarily into the 3-space and gradually adjust these points in such a way that their resulting inter-point distances are as close to their desired distances as possible. Thus, whether we have a good desired distance is critical to the success or failure of our method.

To obtain the desired distance matrix, we rely upon the secondary structure of the RNA sequence. Let D_{ij} denote the desired distance between r_i and r_j, where $1 \le i < j \le n$. We first state that Rule 1 is an initialization rule that creates a desired inter-nucleotide distance matrix for an RNA sequence. In this rule, *r* is set to be 8.5.

Rule 1: This rule is used for any RNA sequence in general to initialize the distance matrix.

(1) If r_i is adjacent to r_j, $D_{ij} = 7.04$,

(2) If r_i is paired with r_j, $D_{ij} = 2r$,

(3) If r_i and r_j are neither base pairs nor adjacent to each other in the subsequence without pseudoknot of an RNA sequence, $D_{ij} = DO$ which is a relatively large number; otherwise $D_{ij} = DO/2$.

For a segment without a turn, Rule 2 is used.

Rule 2: This rule is used for a segment of base pairs without a turn.

Step 1: The relative positions of the nucleotides in the 3-space are determined by using Eq.(1) and (3) as follows:

Let (P_{ix}, P_{iy}, P_{iz}) denote the relative position of the *i*th nucleotides in the segment in the 3-space.

Step 2: Use the following formula to find D_{pq} for all inter-nucleotide distances of this segment.

$D_{pq} = \sqrt{(P_{px} - P_{qx})^2 + (P_{py} - P_{qy})^2 + (P_{pz} - P_{qz})^2}$, where $i \le p \le i + k$, $j - l \le p \le j$ and $j - l \le q \le j$.

Rule 3: This rule is used for a segment with a turn.

Step1: Divide the nucleotides of the segment together with the turn into two sequences. For each sequence, use Eq.(1) to determine the relative positions of all nucleotides in the 2-space.

Step2: For the nucleotides in the turn, construct a desired inter-point distance matrix by using Rule 1.

Step3: Apply the relaxation method to the nucleotides of the turn, using their positions in the 2-space determined in the Step1 as initial positions.

Step4: Based upon the above results, use Eq.(3) to determine the relative positions of nucleotides in both segment and loop.

Step5: Based upon the results obtained above, construct the desired distance matrix of all nucleotides in both segment and turn by using the following equation:

$D_{pq} = \sqrt{(P_{px} - P_{qx})^2 + (P_{py} - P_{qy})^2 + (P_{pz} - P_{qz})^2}$, where $i \le p < q \le j$.

After the desired distance matrix of an RNA sequence is constructed, we apply the relaxation method to predict its 3-space structure. We first divide all of the *n* nucleotides into two sequences. Then, we use Eq.(1) to determine the relative positions of all nucleotides of the two sequences in the 3-space. Finally, we use the relaxation method to map them in such a way that the distances are preserved as much as possible.

5. PREDICTING THE RNA TERTIARY STRUCTURE WITH HELICAL DISTANCE ALGORITHM

In the following, we present the entire algorithm.

Algorithm: The algorithm for predicting the tertiary structure of an RNA sequence
Input: An RNA sequence $R = r_1 r_2 \cdots r_n$ and a parameter *k*.
Output: The tertiary structure of an RNA sequence in the 3-space.

Step 1: Predict the secondary structure of *R* by using the algorithm in [13].

Step 2: Transform the secondary structure into the desired distance matrix.

Step 3: If *n* is an odd number, we divide the first (*n*-1) elements of the input sequence into two equal sequences; otherwise, divide the input sequence into two equal sequences. Then, assign initial locations for all nucleotides in the 2-space by using Eq.(1). For *n* being odd, do the following extra work:

$$x_{odd} = \pi r - (L/10)\cos\theta,$$
$$y_{odd} = -(L/10)\sin\theta.$$

Step 4: Use Eq.(3) to assign all nucleotides into initial locations in 3-space.

Step 5: Utilize the relaxation method [12] to adjust all nucleotides in the 3-space, iteratively *k* times.

Step 6: Output the 3-space locations as the tertiary structure of an RNA.

6. THE VALIDITY OF OUR PREDICTED STRUCTURES

In the previous section, we showed our method to predict an RNA 3-space structure from an RNA sequence. We shall verify the validity of our method by checking the number of turning points of our predicted RNA structure and those of the RNA tertiary structure. In the following, we shall discuss how we define a turn in the 3-space. We are given a sequence of a 3-space points $(X_1, Y_1, Z_1), (X_2, Y_2, Z_2), \cdots, (X_n, Y_n, Z_n)$. Let P_{i-1}, P_i and P_{i+1} be three consecutive points in this sequence. Let ϕ_i denote the angle between $\overline{P_i P_{i-1}}$ and $\overline{P_i P_{i+1}}$. In our experiments, we say the there is a turn, if ϕ_i is smaller than 120°.

Let set Tp be the set of turning points of our predicted RNA 3-space structure. Let Set Tr be the set of turning points of the RNA tertiary structure from PDB. If $Tp = Tr$, there is a perfect matching between our predicted RNA 3-space structure and the RNA tertiary structure. But, this is rather difficult to obtain. We must allow some deviation.

Let *d* denote our allowable deviation. That is, we say that a turning point i in Tp is correct with respect to some j in Tr if $i - d \leq j \leq i + d$ where d is a pre-specified. Let C denote the number of the correct points, and Rdd denote the number of the redundant points. Let $|Tp|(|Tr|)$ denote the size of Tp (Tr). We define Sr_d , be a similarity ratio based upon our allowable deviation d , as follows,

$$Sr_d = \frac{2C - Rdd}{|Tp| + |Tr|}.$$

7. EXPERIMENTAL RESULTS

Our program was written in the C language. The entries of RNA were taken from RNABase [29] which is a classification database of 1118 RNA’s whose physical structures are stored either in Nucleotide Acid Database (NDB)[30] or Protein Data Bank (PDB)[31]. We used 731 of them to test our approach.

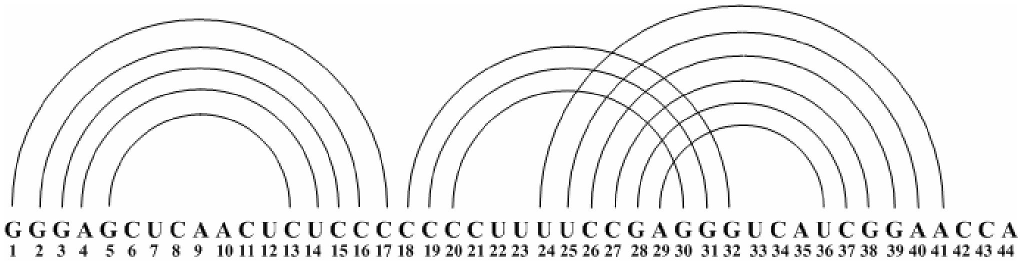

Fig. 3 The secondary structure of RNA 1A60 sequence

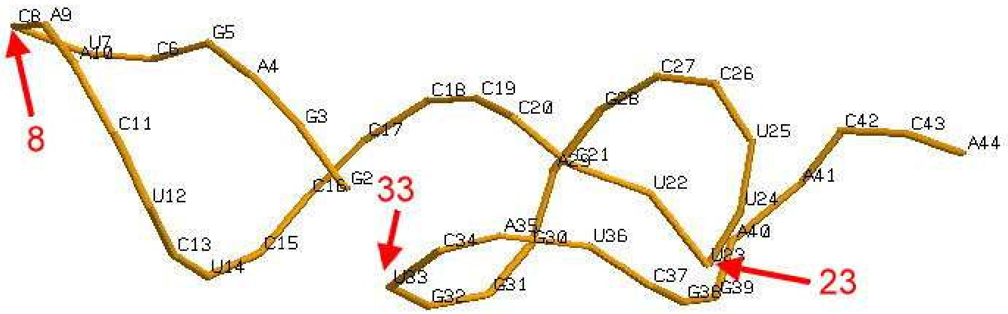

Fig. 4 The tertiary structure of RNA 1A60 sequence obtained from NDB

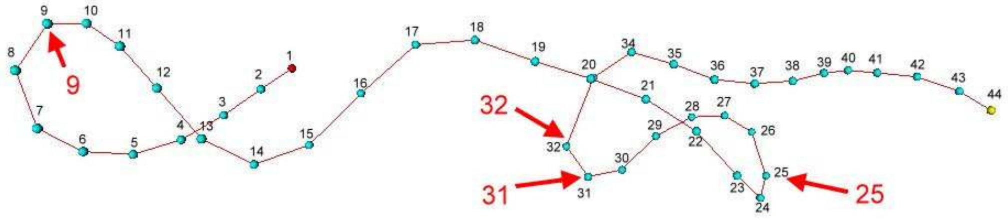

Fig. 5 The tertiary structure of RNA 1A60 sequence predicted by out algorithm

In our experiment, we set five parameters, r and θ of Eq.(1) as $r = 8.5$ $\theta = 32.48$, $DO = 50$ in Rule 1 of our algorithm, the number of iterations k = 2000 and our allowable deviation d = 2. For each sequence, after obtaining the predicted result, we used the definition of a turn introduced in Section 6 to find out all of the turns predicted. Limited by space, in the following, we only introduce an experimental results is detail. Finally, we also show the summary of our experimental results.

Experiment: Predicting the Tertiary Structure of RNA 1A60

In this experiment, we selected the RNA 1A60 sequence from NDB. The sequence is as follows:

G G G A G C U C A A C U C U C C C C C C
C U U U U C C G A G G G U C A U C G G A
A C C A

The length of the RNA 1A60 sequence is 44. Its secondary structure with pseudoknots predicted by using the algorithm in [A2000] is shown in Fig. 3. The tertiary structure of RNA 1A60 sequence obtained from NDB is shown in Fig. 4. From Fig. 4, we can see that Points 8, 23 and 33 are turning points. None of them is a helical turn. The tertiary structure predicted by us is in Fig. 5. As shown in Fig. 5, Points 9, 25, 31 and 32 are turning points. Although it appears that there are more turning points in the predicted result, actually points 31 and 32 in our predicted 3-space result are so close to point 33 in the tertiary structure that we consider them as one point and the similarity ration is again equal to 1.

We totally tested 731 RNA sequences in the experiment, and the results are shown in Fig. 6. In Fig. 6, the *x*-axis denotes the similarity ratio between our predicted result and physical RNA, and the *y*-axis denotes the number of our predicted results which fall into a case within the similarity ratio. As shown in Fig. 2, 179 cases of our predicted results fall into the cases within 100%, and furthermore 391 and 161 cases of our predicted results fall into the cases within 90% and 80% respectively. These experimental results show that our method is quite feasible.

The web-site based upon our algorithm can be found in [http://alg.csie.ncnu.edu.tw/rnapredict.htm].

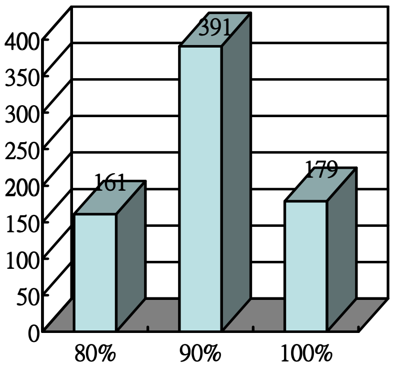

Fig. 6 the summary of our predicted results

8. CONCLUSIONS

In this paper, we proposed an algorithm for predicting the tertiary physical structure of an RNA sequence based upon its secondary structure. Through the experiment results, we conclude that our approach can be used to predict the tertiary structure quite well. At present, our algorithm can not handle rather long sequences with multiple helical structures. We believe that we have to use more complicated optimization handle this kind of sequences.

9. REFERENCES

[1] T. Macke and D. Case, “Modeling unusual nucleic acid structures. In: N. Leontes and J.J. SantaLucia,” **Molecular Modeling of Nucleic Acids**, American Chemical Society, Washington, DC, pp. 379-393, 1998.

[2] C. Zwieb and F. Muller, “Three-dimensional comparative modeling of RNA,” **Nucleic Acids Symp Ser**, vol. 36, pp. 69-71, 1997.

[3] C. Massire and E. Westhof, “MANIP: an interactive tool for modeling RNA,” **Journal of Molecular Graph and Modeling**, vol. 16, pp. 197-205, 255-257, 1998.

[4] F. Major,”Building three-dimensional ribonucleic acid structures,” **IEEE Computing in Science and Engineering**, vol. 5, pp. 44-53, 2003.

[5] DC. J. E. Barreda and Y. Shigenobu, “RNA 3D Structure Prediction: (1) Assessing RNA 3D Structure Similarity from 2D Structure Similarity, “ **Genome Informatics**, vol. 15, No. 2,, pp. 112-120, 2004.

[6] Y.G. Yingling and B.A. Shapiro,”The prediction of the wild-type telomerase RNA pseudoknot structure and the pivotal role of the bulge in its formation,” **Journal of Molecular Graph and Modeling**, vol. 25, pp. 261-274, 2006.

[7] Y. Shigenobu and C. A. D. Carpio, “A Bioinformatic Approach for RNA 3D Structure Prediction: Development of a Knowledge-Base for 2D-to-3D Structural Elements Compatibility Analysis,” **Genome Informatics**, vol. 12, pp. 360-361, 2001.

[8] K. Yamaguchi and C. A. Del Carpio, “A Genetic Programming Based System for the Prediction of Secondary and Tertiary

Structures of RNA," **Genome Informatics**, vol. 9, pp. 382-383, 1998.

[9] Y. Shigenobu and C. A. D. Carpio, "Development of a Bioinformatic System for Determination of the 3D Structure of RNA from Secondary Structure Constrains," **Genome Informatics**, vol. 11, pp. 305-306, 2000.

[10] D. S. Johnson, C. R. Aragon, L. A. McGeoch and C. Schevon, "Optimization by simulated annealing: an experimental evaluation," **Part I, graph partitioning, Operations Research**, vol. 37, pp. 875-892, 1989.

[11] D. S. Johnson, C. R. Aragon, L. A. McGeoch and C. Schevon, "Optimization by simulated annealing: an experimental evaluation," **Part II, graph coloring and number partitioning, Operations Research**, vol. 39, pp. 378-406, 1989.

[12] Chang, C. L. and Lee, R. C. T., "A Heuristic Relaxation Method for Nonlinear Mapping in Cluster Analysis," *IEEE Trans. Syst., Man, Cyber.*, vol. SMC-3, pp. 197-200, 1973.

[13] T. Akutsu, "Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots," **Discrete Applied Mathematics**, vol. 104, pp. 45-62., 2000.

[14] J. S. Deogun, R. Donis, O. Komina and F. Ma, "RNA Secondary Structure Prediction with Simple Pseudoknots," **APBC2004**, vol. 29, pp.239-246, 2004.

[15] S. Ieong, M. Y. Kao, T. W. Lam, W. K. Sung and S. M. Yiu, "Predicting RNA Secondary Structures with Arbitrary Pseudoknots by Maximizing the Number of Stacking Pairs," **2nd IEEE International Symposium on Bioinformatics and Bioengineering**, pp. 183-190, 2001.

[16] E. Rivas and S. R. Eddy, "A Dynamic Programming Algorithm for RNA Structure Prediction Including Pseudoknots," **Journal of Molecular Biology**, vol. 285, pp. 2053-2068, 1999.

[17] F. Tahi, M. Gouy and M. Regnier, "Automatic RNA secondary structure prediction with a comparative approach," **Computers and Chemistry**, vol. 26, pp. 521-530, 2002.

[18] D. H. Turner, N. Sugimoto and S. M. Freier, "RNA structure prediction," **Ann. Rev. Biophys. and Biophys. Chem.**, vol. 17, pp. 167-192, 1988.

[19] M. S. Waterman and T. F. Smith, "D ynamic Programming Algorithms for R NA Secondary Structure," **Advances in Applied Mathematics**, vol. 7, pp. 455 -464, 1986.

[20] M. Zuker, "Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information," **Nucleic Acids Res.**, vol. 9, pp.133-148, 1981.

[21] R. B. LyngsØ and C. N. S. Pedersen, "Pseudoknots in RNA Secondary Structures," **RECOMB**, pp. 201-209, 2000.

[22] V. Jyan and C. Wilson, "RNA secondary structure prediction based on free energy and phylogenetic analysis," **Journal of Molecular Biology**, vol. 289, pp. 935-947, 1999.

[23] R. B. LyngsØ, M. Zuker and C. N. S. Pedersen, "Internal loops in RNA secondary structure prediction ," **RECOMB**, pp. 260-267, 1 999.

[24] M. Zuker, "The Use of Dynamic Programming Algorithms in RNA Secondary Structure Prediction," **Mathematical Methods for DNA Sequences**, Waterman, M. S. Ed., CRC Press Inc., Boca Raton Florida, Chapter 7, pp. 159-184, 1989.

[25] M. Zuker and D. Sankoff, "RNA Secondary Structures and Their Prediction," **Bulletin of Mathematical Biology**, vol. 46, pp. 591-621, 1984.

[26] M. Zuker, "Mfold web server for nucleic acid folding and hybridization prediction," **Nucleic Acids Res.**, Vol. 31, No.13, pp. 3406-15, 2003.

[27] D. L. Nelson and M. M. Cox, "Lehninger Principles of Biochemistry Third Edition," **Worth Publishers**, 2000.

[28] A.H. Watt, “3D Computer Graphics 3rd edition,” Chapter 8, *Addison Wesley*, 1999.

[29] V. L. Murthy and G. D. Rose, “RNABase: an annotated database of RNA structures,” **Nucleic Acids Research**, vol. 31, no. 1, pp. 502-504, 2003.

[30] H. M. Berman, W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan and B. Schneider, “The Nucleic Acid Database: A Comprehensive Relational Database of Three-Dimensional Structures of Nucleic Acids,” **Biophys. Journal**, vol. 63, pp. 751-759, 1992.

[31] F.C. Bernstein, T.F. Koetzle, G.J. Williams, E.F. Meyer Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi,“The Protein Data Bank. A computer-based archival file for macromolecular structures,” ***Eur J Biochem***, vol. 80, pp. 319–324, 1977.