A Hybrid Genetic Algorithm for RNA Structural Alignment

Abdesslem Layeb
University Mentouri Constantine
Route Ain El Bey, Algeria

Imen Bensetira
University Mentouri Constantine
Route Ain El Bey,Algeria

Kenza Bouaroudj
University Mentouri Constantine
Route Ain El Bey, Algeria

ABSTRACT
The RNA structural alignment is one of the most challenging tasks in bioinformatics. However, finding the accurate conserved structure of a set of RNA sequences is still being a difficult task. In this work, the problem is cast as an optimization problem for which a new framework relaying on hybrid genetic algorithm is proposed. The contribution consists in using a new objective function based on the Structure Conservation Index (SCI). In order to enhance the Genetic Algorithms (GA) performances, a Simulated Annealing (SA) procedure has been used. The proposed algorithm is composed on two phases.The first phase consists of applying a genetic algorithm.In the second phase, the simulated annealing procedure is applied in order to improve the final population given by the genetic algorithm. Experiments on a wide range of data sets have shown the effectiveness of the proposed framework and its ability to achieve good quality solutions comparing to those given by others techniques.

General Terms
Algorithms,bioinformatic, metaheuristic, optimization problem.

Keywords
RNA Structure Prediction, Genetic Algorithm, Simulated Annealing,Structure Conservation Index.

1. INTRODUCTION
The ribonucleic acids (RNA) areamong the molecules stimulating the interest of the biologists. The RNA is now regarded as a potential target, very interesting in pharmacology. In fact, RNA plays a multiple and fundamental roles in all cellular process[1].The RNA plays a direct role in the catalytic processes like the synthesis of proteins. It plays also a fundamental role in regulation processes of the DNA replication, DNA transcription and translation [2]. There are close connections between RNA structures and their catalytic function. Indeed, it has been showed that a catalytic RNA becomes functional only when it has adopted its structure. Consequently, it is very important to know the secondary structure and when it is possible the tertiary structure of RNA molecules. Satisfactory prediction of RNA secondary and tertiary structures is an important issue in bioinformatics and requires coupling experimental and modeling approaches [3].

Considering obtaining the structures of large RNA molecules by using nuclear magnetic resonance (NMR) spectrum is often difficult, the reliable forecast of RNA structures of their primary sequences is strongly required. Two main approaches are currently used to predict RNA secondary structures. The first is the comparative sequence analysis [4]. The basic idea is to examine homologous sequences to identify potential helices which maintain complementarities in sequences. The second approach is the thermodynamic optimization. In this method we use thermodynamics to determine structures with minimum or near minimum free energies [5].

One of the iterative methods that have been developed recently to solve this type of problem which considered as an optimization problem is Genetic Algorithms (GA). It is a stochastic iterative algorithm which maintains a population of individuals. GA adapts nature optimizing principles like mechanics of natural selection and natural genetics. Each individual represents a potential solution in the search space of the problem. Basically, a genetic algorithm consists of three essential operations: selection, crossover, and mutation. The selection operator consists in selecting an intermediate population from the current one in order to create the future population by using crossover and mutation operators. The crossover operator merges two individuals to provide new ones. The mutation operator allows moving each solution to one of its neighbors in order to maintain a good diversity during the optimization process. GA allows guided search that samples the search space. Although GAs have been showed to be appropriate for solving many bioinformatics problems [6], their computational cost seems to be a dissuasive factor for their use on large instances. To overcome this drawback and in order to get better speed and quality convergence, their implicit parallelism is exploited.

The Stochastic Local Search (SLS) methods were demonstrated to be useful in solving many complex problems. Among these methods, one distinguishes the Simulated Annealing algorithm [7] that is found to be useful in many hard combinatorial optimization problems. The Simulated Annealing (SA) is a local search technique inspired by a process used in metallurgy [8]. This process of alternating cycles of slow cooling and reheating or annealing, which tend to minimize the energy of the material. The simulated annealing method uses the Metropolis algorithm; which starts from a given configuration, and it's subjected to an elementary modification system. If the perturbation has the effect of reducing the objective function (or energy) of the system, it is accepted. Otherwise, it is accepted with the probability $exp\ (\Delta E\ /\ T)$. The iterative application of this rule will generate a sequence of configurations that tend to thermodynamic equilibrium.

Within this issue, we propose in this paper a new framework to cope with RNA secondary structure alignment problem. The aim of our approach called GARSRNA is to improve the alignment accuracy in terms of secondary structure information using the secondary conservation index SCI [9,10] as objective function. To foster the process, a simulated annealing method has been used to improve the results given by the genetic algorithm. To assess the efficiency and accuracy of the proposed appraoch, several experiments were designed. We have used different data

sets taken from the BRALiBASE benchmark base [11]. The obtained results are very encouraging and show the feasibility and effectiveness of the proposed hybrid approach.

The remainder of the paper is organized as follows. In section 2, the RNA secondary prediction is presented. In section 3, a formulation of the tackled problem is given. In section 4, the proposed framework is described. Experimental results are discussed in section 5. Finally, a conclusion is drawn.

2. RNA SECONDARY PREDICTION

The prediction of RNA structures has become a significant task in bioinformatics. To find the primary structure of RNA molecules, we use techniques of sequencing. However the task is more difficult for secondary and tertiary structure. The secondary structure of RNA sequence is a set S of base pairs (r_i,r_j) over the alphabet {A, C, G, U} satisfying the following criteria [12]:

1. $\forall\ (r_i,r_j) \in S, (r_i,r_j) \in \{(A,U), (U,A), (G,C), (C,G), (G,U), (U,G)\}$
2. $1 \le i < j \le |S|$
3. $\forall\ (r_i,r_j), (r_{i'},r_{j'}) \in S,\ i = i' \Leftrightarrow j = j'$
4. $(r_i,r_j) \in S \Rightarrow |j - i| \ge 4$

There are two approaches to determine the structure of RNA molecules. The first is an exact method which uses experimental techniques such as nucleic magnetic resonance (NMR) and X-ray crystallography. This method is long, difficult and expensive. However, the second method predicts the secondary structure starting from the primary structure by using secondary prediction algorithms [13]. One distinguishes two methods in this approach: comparative method and thermodynamic method. The comparative method is based on the homology of sequences. Homologous sequences share the same structure. These methods are made up of the following steps [5]:

1. Construction of a multiple alignment.
2. Detections of the positions correlated by using the mutual information of columns i and j of the multiple alignment.

This method requires that the alignment of homologous sequences is known in advance. On the other hand, the thermodynamic method is based on following assumptions [14]:

1. A quantity of free energy corresponds to each configuration of the molecule.
2. The most stable configuration is that which minimizes the free energy.
3. The molecule, while being folded up, adopts the most stable configuration.

Consequently, the RNA structure prediction problem consists in finding the most stable configuration. The energy of the molecule is the sum of energies of each pair of bases**.** The free energy of a structure S is given by following formula:

$$E(S) = \sum_{(r_i,r_j)\in S} a(r_i, r_j) \qquad (1)$$

$a(r_i, r_j)$ is the free energy of the pair (r_i,r_j). This method presents some limits like the relevance of the energy function and biological assumptions are not always true. Indeed, there is no reliable tool for detecting functional RNAs in multiple sequence alignment. One of the most used score scheme to assess the precision of the conserved secondary structure information contained within the alignment, is the structural conservation index SCI [9]. It is based on the RNAalifold consensus folding algorithm (MFE) [15, 16] which is based upon the sum of a thermodynamic and a covariance term. The Structural conservation index is computed using the following function:

$$SCI = E_A / E' \qquad (2)$$

Where E_A is the consensus minimum free energy (MFE) of the alignment and E' is the average of the individual MFEs. The SCI is close to zero if RNAalifold identifies no common RNA structure in the alignment, while a set of perfectly conserved structures has an SCI≈1. An SCI >1 shows that there is a conserved RNA secondary structure which is, in addition, supported by compensatory and/or consistent mutations [9].

3. MULTIPLE STRUCTURAL ALIGNMENT PROBLEM FORMULATION

Let $S = \{s_1, s_2, ..., s_n\}$ a set of n sequences with $n \ge 2$. Each sequence s_i is a string defined over an alphabetΛ. The lengths of the sequences are not necessarily the same. The multiple structural alignment problem can be defined by specifying implicitly a pair (Ω, C) where Ω is the set of all feasible solutions that is potential alignments and C is a mapping $\Omega \rightarrow R$ called structure conservation index. Each potential structural alignment is viewed as a set $S' = \{s'_1, s'_2, ..., s'_n\}$ satisfying the following criteria:

Each sequence s'_i is an extension of s_i and is defined over the alphabet $\Lambda' = \Lambda \cup \{-\}$. The symbol "–" is a dash denoting a gap. Gaps are added to s_i in a way when deleted from s'_i, s_i and s'_i are identical.

For all i,j length(s'_i)=length(s'_j).

A score of an alignment S' denoted by *SCI* (S') is defined as:

$$SCI(S') = \frac{E_{Align}(y^{MFE}_{S'})}{\frac{1}{n}\sum_{x\in S'} E(y^{MFE}_x)} \qquad (3)$$

Where n is the number of sequences in the alignment S'. For a single sequence x, E(y) denotes the free energy of a secondary structure $y \in S(x)$, and $y^{MFE}_x = \arg\min_{y\in S(x)} E(y)$ is defined to be the MFE structure of x calculated by RNAfold[17].Similarly, for an alignment S', $E_{Align}(y)$ is the free energy of a consensus structure $y \in S(x)$, and

$y_{S'}^{MFE} = \arg\min_{y \in S(S')} E_{Align}(S')$ the consensus MFE structure of S'.

The free energy of a consensus structure is defined as the average of the energy contributions of the single sequences plus covariance scores for bonuses of compensatory and consistent co-mutation in the alignment [18].

The addressed task is clearly a combinatorial optimization problem. Although the computing power available has been increasing steadily at a rapid rate, it is still practically impossible to find globally optimal solutions to combinatorial optimization problems. The main reason is that the required computation grows exponentially with the size of the problem. Therefore, it is often desirable to find near optimal solutions to these problems. Efficient heuristic algorithms offer a good alternative to reach this goal such as the progressive methods.

4. THE PROPOSED APPROACH

In order to show how evolutionary framework have been tailored to the problem at hand, we need first to derive a representation scheme which includes the definition of an appropriate representation of potential alignments and the definition of evolutionary operators. Then, we describe how these defined concepts have been integrated in a genetic algorithm.

4.1 Genetic representation of alignment

To successfully apply genetic operators on multiple RNA structural alignment, we have needed to map potential solutions into a chromosome representation that could be easily manipulated by genetic operators. The multiple structural sequence alignment Aln= $\{S_1,S_2,\ldots,S_n\}$ is viewed as an alphabetic matrix AM where:

- Each line *i* represents a sequence S_i'.
- The character "–"denotes a gap.

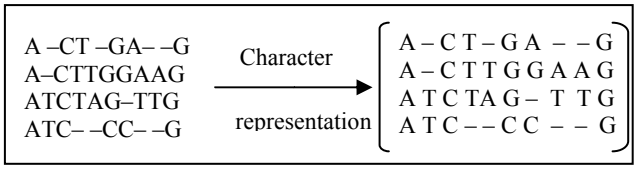

Fig 1: Alphabetic representation of multiple sequence alignment.

4.2 Population creation

To solve the multiple structural alignment problem, our algorithm starts by generating an initial population of chromosomes which encode the potential sequence alignments. The initial solution is very significant; a good initial solution can effectively converge faster and consequently cut the computational cost.Therefore, it is better to start with a good population which contains aligned solutions created by a progressive alignment method such as ClustalW[19].

4.3 Selection

Selection is a genetic operator that chooses a chromosome from the current generation's population for inclusion in the next generation's population. We can find in the literature, a large number of selection methods which are more or less adapted to the problems they treat. In this research, we choose the Elitism selection [20].This technique can promote the best individuals of the population, so the most promising ones will participate in the improvement of our population. Elitism method can increase the convergence of genetic algorithm, because it always preserves the best solutions in every generation [20].Nevertheless, elitism method suffers from the local optimum problem, that's why we have chosen some less efficient individuals in order to give them a chance to be improved in the future generations.

4.4 Mutation

Formally, the problem of structural alignment consists in well placing gaps in different RNA sequences. A wrong placement of gaps appears when gap seriesof the same size occur in different positions (Figure 2), or when an island of characters is surrounded by gaps (Figure 3).

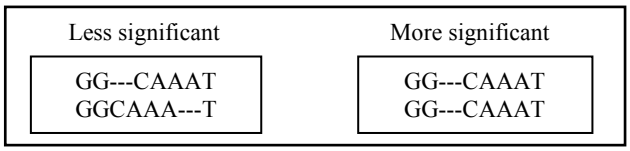

Fig 2: gaps of the same size in different positions.

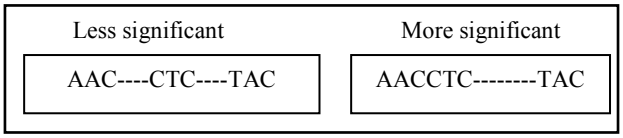

Fig 3: The presence of islands in a sequence of an alignment.

To resolve this problem, we have used a simple mutation based on changing randomly the position of gaps. Unfortunately, we noticed that this kind of mutation does not improve the solution quality. For that, we have used four adaptive mutations. The mutation operators operate on a gap, series of gaps, gap column and gap blocs.

4.4.1 Gap Mutation

In this kind of mutation, we chose an isolated gap close to a suite of gap in a sequence and we merge it with this suite. (Figure 4)

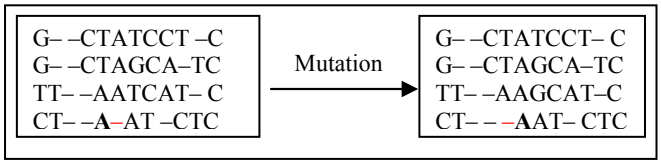

Fig 4: Single gap mutation.

4.4.2 Gap sequence Mutation

A sequence of gaps is moved to the left or to the right as it's shown in Figure 5.

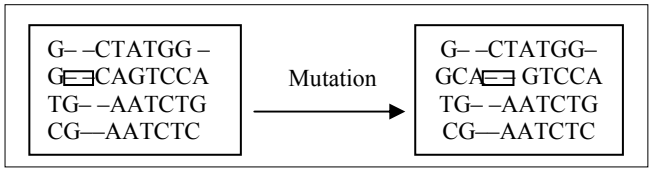

Fig5: Gap series mutation.

4.4.3 Gap column Mutation

This kind of mutation affects a set of sequences of an alignment.It consists in taking a column of gaps and moves it to the left or to the right (Figure 6).

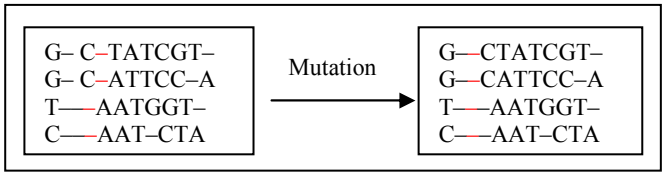

Fig6: Gap column mutation.

thenucleotid composition [21]. Therefore, several comparative approaches have been proposed[22, 23, 24, 25, 26]. For purpose, Washietand al [27] have proposed the Structure Conservation Index (SCI) as a feature to measure the evolutionary conservation in terms ofsecondary structures of multiple sequence alignment.Assuming that MFE for the consensus secondary structure is close to that for each sequence if a given multiple alignment is structurally conserved.The SCI of an alignment is given as the fraction of the consensus folding free energy ($E_{consensus}$) to the average of the folding free energies of the single sequences. We denote with S(x) the entire folding space of a single sequence x, and we denote with S(A) the entire consensus folding space of an alignment A. The SCI is defined as:

$$SCI(A) = \frac{E_{Align}(y_A^{MFE})}{\frac{1}{n}\sum_{x \epsilon A} E(y_x^{MFE})} \quad (4)$$

4.4.4 Gap bloc Mutation

A gap bloc is moved left or right. This kind of mutation affects many sequences (Figure 7).

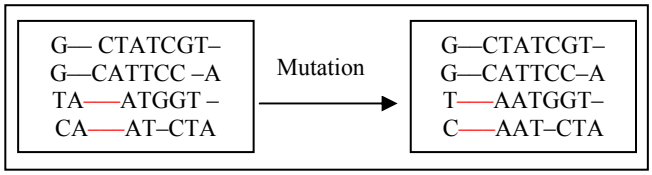

Fig 7: Gap bloc mutation.

4.5 Crossover operator

Crossovers are important for promoting the exchange of high quality blocks within the population by exchanging subparts of two chromosomes. In our case, the application of a simple crossover operator may give completely an incorrect solution. Therefore, we used a new crossover more adapted to the multiple structural alignment problem. The idea is to take two alignments and then a vertical cut is applied on each alignment. The next step of the crossover operator is to create new individuals by interchanging the parent parts (Figure 8).

4.6 Fitness evaluation

The objective function is used to evaluate the alignment quality, and it isthe center of the optimization process.Generally, the objective function is the mathematical tool used to measure the degree to which two or more sequences are similar. Therefore, the definition of an adequate affinity function is a crucial biological task in bioinformatics. The secondary structure with the minimum free energy (MFE) has been regarded as the most reliable prediction of RNA secondary structures[14]. However, MFE alone could not be an appropriate measure for identifying a certain kind of RNA since the free energy is heavily biased by

Where n is the number of sequences in the alignment A. For a single sequence x, E(y) denotes the free energy of a secondary structure $y \epsilon\, S(x)$, and$y_x^{MFE} = \arg\min_{y \in S(x)} E(y)$is defined to be the MFE structure of x calculated by RNAfold [17]. Similarly, for an alignment A,$E_{Align}$is the free energy of aconsensus structure $y \epsilon\, S(A)$;

$y_A^{MFE} = \arg\min_{y \in S(A)} E_{Align}(A)$is the consensus MFE structure of A. The free energy of a consensus structure is defined as the average of the energycontributions of the single sequences and the covariance scores for bonuses of compensatory and consistentco-mutation in the alignment[18].

For a multiple alignment which is not structurally conserved, the SCI will be near to 0, that's means there is no common RNA structures between different sequences. The SCI should be close or greater than 1for an alignment that is structurally conserved. If the alignment is structurally well conserved and compensatory and consistent mutation often occurs, the SCI maybe above 1.

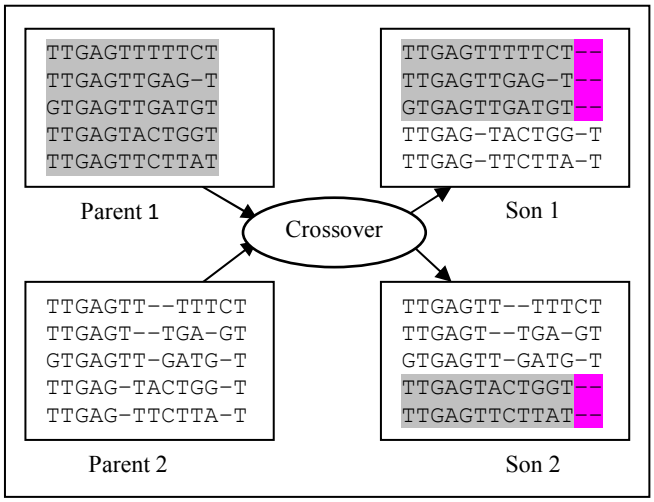

Fig 8: Crossover operator.

4.7 Local search

Recently, studies have shown that the hybridization of evolutionary algorithms and population-based metaheuristics with local search techniques is very effective for the improvement of the results. In our approach we have used the simulated annealing method as local search method.The simulated annealingprocedure has been applied successfullyto solve many optimization problems. It should be noted that our algorithm is flexible, so we can use other local search methods such as Tabu search.In the problem at hand, the concept of the simulated annealing energy is replaced by the SCI measurement, while the elementary modification brought to the individuals is madeby the application of mutation operator. The temperature schema should be well selected in order to avoid that simulated annealing is trapped in a local minimum. In more detail, the

proposed simulated annealing for RNA structural prediction is as follow:

Input: An alignment *Aln*

Initialisation:
Select an initial temperature T=T0.
Set Aln_{best}=*Aln* and SCI_{best} = *C (Aln)*.
Repeat
(1) *Aln'*=apply mutation operator to *Aln*.
(2) Calculate the objective function of *Aln'*.
(3) If *C(Aln')*>SCI_{best}then
Aln = Aln'
Aln_{best} = *Aln'*
SCI_{best} = *C(Aln')*
(4) Else
Generate a random number r
If$r < e^{\frac{C(Aln)-C(Aln')}{T}}$then *Aln=Aln'*
Update T.
Until a termination criterion is reached.

Output: Aln_{best} and SCI_{best}.

4.8 Outline of the proposed framework

The proposed approach called GARSRNA is divided in two-phase algorithm.In the first stage, we apply the genetic algorithm. First, an initial population is created as was explained previously. In the second step, we refine iteratively the alignment in order to improve the quality of the conserved secondary structure information of the alignment. In each generation of our program, we perform a selection operation to constitute the mature population. Then, the crossover and mutation operators are applied which allow exploring other solutions, only one type of mutations is selected randomly. The mature population is evaluated using the SCI objective function. The global best solution is then update if better one is found and the whole process is repeated until having satisfaction of stopping criterions.

In order to improve the efficiency of the genetic algorithm, we have introduced the local search method after the genetic algorithm. In our approach the hybridization between the two metaheuristics is done sequentially. First,we execute the genetic algorithm, and then we apply the simulated annealing on the result population. However, we can also integrate the simulated annealing in the core of the genetic algorithm as mutation operator. But, this kind of hybridization will increase considerably the runtime execution because it carries out several evaluations of the objective function. In more detail, the proposed hybrid evolutionary algorithm for RNA structural prediction is as follow:

Input:A set of sequences SEQ

(1) Generate population of n chromosomes, POP.
Repeat
(2) Select a subset of the population using the selection operator.
(3) Apply a crossover operation.
(4) Apply a mutation operation.
(5) Evaluate the current population.
(6) If $SCI(Aln_{best}) < SCI(Aln_i)$ then $Aln_{best} = Aln_i$and$SCI_{best} = SCI(Aln_i)$.
(7) Apply the replacement operator
Until a termination criterion is reached.
(8) Apply the simulated annealing algorithm.

Output: Aln_{best} and $CSI(Aln_{best})$

5. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed approach was implemented in MATLAB language. The choice of this language is motivated by the easy and effective manipulation of vectors and matrices by MATLAB while most data in GARSRNA are in matrix form. Another advantage is the existence of a Bioinformatics toolbox which contains predefined functions for sequence analysis.
To assess the experimental performance of our approach in the RNA secondary structure prediction, we have used several tests obtained from the BRALiBASE benchmark base [11], which contains sets of RNA structures. These tests are divided into five references: Intron, rRNA, SRP, tRNA and U5 which are characterized by their sequence length.Moreover, the found results were statistically evaluated using the Friedman test.The table 1 summarizes the found results; the first column contains the used tests, the second column contains the results of pure genetic algorithm, and the last column contains the results of our approach.
The results of our method illustrate clearly the effectiveness of merging the genetic algorithm and the simulated annealing to deal with the multiple RNA structural alignment. The statistical test of Friedman (Figure 9) illustrates obviously that the use of the simulated annealing increases the efficiency of the genetic algorithm, there is clear difference between GARNA and GARSRNA.

Table1.Comparison between GARNA and GARSRNA.

		GARNA	**GARSRNA**
intron	Aln 20	0.59	0.65
	Aln 25	0.70	0.78
	Aln 75	0.54	0.91
	Aln 76	0.81	0.81
	Aln 11	0.65	0.65
rRNA	Aln 12	0.60	0.60
	Aln 20	0.82	0.82
	Aln 25	1.02	1.02
	Aln 34	0.91	0.91
	Aln 50	0.98	0.98
	Aln 6	0.29	0.69

SRP	Aln 15	0.81	0.81
	Aln 34	0.24	0.24
	Aln 50	0.20	0.39
	Aln 61	0.81	0.95
tRNA	Aln1	0.46	0.69
	Aln34	1.09	1.09
	Aln 50	1.08	1.08
	Aln 60	1.16	1.16
	Aln 46	0.86	1.03
U5	Aln 63	0.90	0.90
	Aln 70	0.47	0.52
	Aln 80	0.69	0.79
	Aln 90	0.70	0.71
	Aln 100	0.60	0.60

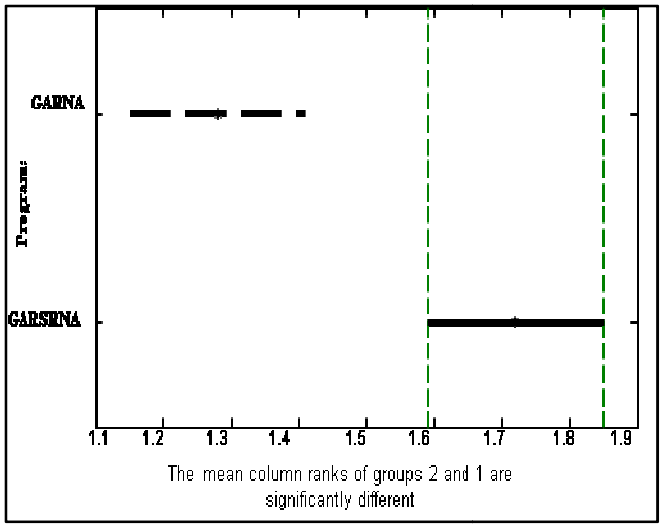

Fig9:Friedman test (0.05) comparesGARNA and GARSRNA.

6. CONCLUSION

In this paper, we have proposed a novel approach to solve the RNA secondary structure prediction problem. The objective of this study is to demonstrate the efficiency of the genetic algorithm and its hybridization with a local search method to deal with the problem at hand.The obtained results are very encouraging and show the effectiveness of the method. In all tests, the use of the simulated annealing has improved the quality of the results compared to those found by a pure genetic algorithm. The proposed framework provides an extensible platform for evaluating different objective functions. It would be an interesting attempt to study this issue as ongoing work.

7. References

[1] Eddy,S. R. 2001. Non-coding RNA genes and the modern RNA world. Nat Rev Genet 2(Dec 2001), 919-929.

[2] Mattick,J. S.,Makunin,I. V. 2006. Non-coding RNA. Hum Mol Genet 15 Spec N°1, R17(Apr 2006).

[3] Gorodkin,J., Heyer,L., Brunak,S. and Stormo,G. 1997. Displaying the information contents of structural RNA alignments. CABIOS, Vol. 13, 583–586.

[4] Gutell, R., Power, A., Hertz, G., Putz, E. &Stormo, G. 1992. Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. Nucleic Acids Res, Vol. 20 (21), 5785–595.

[5] Zuker, M., Jaeger, J. & Turner, D. 1991. A comparison of optimal and suboptimal RNA secondary structures predicted by free energy minimization with structures determined by phylogenetic comparison. Nucleic Acids Res,Vol. 19 (10),2707–214.

[6] Han,K.-H. andKim,J.-H. 2000. Genetic quantum algorithm and its application to combinatorial optimization problem. Proc. 2000 Congr. Genetic Computation, vol. 2, La Jolla, CA, 1354–1360.

[7] Kirkpatrick, C.D. Gelart, and P.M. Vecchi. 1983. Optimization by Simulated Annealing. Science, Vol. 220, 671-680.

[8] Balaji, A.N., Jawahar, N. 2010. A Simulated Annealing Algorithm for a two-stage fixed charge distribution problem of a Supply Chain. International Journal of Operational Research, Vol. 7, No.2, 192 – 215.

[9] Washietl, S., Hofacker, I. and Stadler, P. 2005. Fast and reliable prediction of noncoding RNAs. Proc. Natl Acad. Sci. Vol. 102, 2454–2459.

[10] Gruber, A.R., Bernhart, S.H., Hofacker, I.L., Washietl, S. 2008. Strategies for measuring evolutionary conservation of RNA secondary structures. BMC Bioinformatics 9, 122.

[11] Gardner, P., Wilm, A. and Washietl, S. 2005. A benchmark of multiple sequence alignment programs upon structural RNAs. Nucleic Acids Research, Vol. 33(8) 2433–2439.

[12] Sankoff, D. and Kruskal, J. B. 1983.Time warps, string edits, and macromolecules: The theory and practice of sequence comparison.Addison Wesley.

[13] Layeb, A, Meshoul, S., andBatouche, M. 2008. Quantum Genetic Algorithm for Multiple RNA Structural Alignment in the IEEE proceedings of the 2nd Asia International Conference on Modelling& Simulation, pp. 873-877.

[14] Washietl, S. 2010. Sequence and structure analysis of noncoding RNAs. Methods in Molecular Biology, Vol. 609, 285-306.

[15] Washietl, S. and Hofacker,I. 2004. Consensus folding of aligned sequences as a new measure for the detection of functional RNAs by comparative genomics. J. Mol. Biol., Vol. 342, pp. 19–30.

[16] Hofacker, I. L.,Fekete,M., andStadler,P. F. 2002. Secondary structure prediction for aligned RNA sequences. J. Mol. Biol., Vol. 319, 1059–1066.

[17] Hofacker, I. L. 2003. Vienna RNA secondary structure server. Nucleic Acids Res, Vol. 31, 3429-3431.

[18] Okada,Y., Sato, K., and Sakakibara, Y. 2010. Improvement of structure conservation index with centroid estimators. Pacific Symposium on Bio computing, 15:88-97.

[19] Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson,TJ., Higgins, DG., Thompson, JD. 2003. Multiple sequence alignment with the Clustal series of programs.Nucleic Acids Res., Vol. 31, 3497-3500.

[20] Thierens,D.1997.Selection schemes, elitist recombination and selection intensity. in International conference of genetic algorithm, pp. 152-159.

[21] Ziv-Ukelso, M. 2010. A faster algorithm for simultaneous alignment and folding of RNA. Journal of Computational Biology, 17(8), 1051-1065.

[22] Hamada, M., Kiryu, H., Sato, K., Mituyama, T., and Asai, K. 2009. Bioinformatics, 15; 25(4):465-473.

[23] Gesell,T., and Washietl, S. 2008.Dinucleotide controlled null models for comparative RNA gene prediction. BMC Bioinformatics, 9:248.

[24] Tahi, F.,Engelen, S., andRégnier,M. 2003. A Fast Algorithm for RNA Secondary Structure Prediction Including Pseudoknots. Third IEEE Symposium on BioInformatics and BioEngineering (BIBE'03), pp.11.

[25] Engelen, S., andTahi, F. 2010. Tfold: efficient in silico prediction of non-coding RNA secondary structures.Nucleic Acids Res; 38(7):2453-2466.

[26] Engelen, S., andTahi, F. 2007.Predicting RNA secondary structure by the comparative approach: how to select the homologous sequences. BMC Bioinformatics; 8:464.

[27] Washietl, S., Hofacker, I.L., Stadler, P.F.2005. Fast and reliable prediction of noncoding RNAs.ProcNatlAcadSci, 102(7):2454-2459.