

Predicting the minimum free energy RNA Secondary Structures using Harmony Search Algorithm

Abdulqader M. Mohsen, Ahamad Tajudin Khader, Dhanesh Ramachandram and Abdullatif Ghallab

Abstract—The physical methods for RNA secondary structure prediction are time consuming and expensive, thus methods for computational prediction will be a proper alternative. Various algorithms have been used for RNA structure prediction including dynamic programming and metaheuristic algorithms. Musician's behavior-inspired harmony search is a recently developed metaheuristic algorithm which has been successful in a wide variety of complex optimization problems. This paper proposes a harmony search algorithm (HSRNAFold) to find RNA secondary structure with minimum free energy and similar to the native structure. HSRNAFold is compared with dynamic programming benchmark mfold and metaheuristic algorithms (RnaPredict, SetPSO and HelixPSO). The results showed that HSRNAFold is comparable to mfold and better than metaheuristics in finding the minimum free energies and the number of correct base pairs.

Keywords—Metaheuristic algorithms, Dynamic programming algorithms, Harmony search optimization, RNA folding, Minimum Free Energy.

I. INTRODUCTION

THE importance of Ribonucleic Acid (RNA) has increased in the recent years. It was found that RNA performs a central role within the living cells such as carrying genetic information (mRNA), interpreting the code (ribosomal RNA), and transferring genetic code (tRNA). It also performs different functions include catalyzing chemical reactions [1], [2], directing the site specific modification of RNA nucleotides, controlling gene expression, modulating protein expression and serving in protein localization [3], [4]. The function of RNA molecules determines many diseases caused by RNA viruses. Identifying the secondary structure of an RNA molecule is the fundamental key to understand its biological function and predict its tertiary structure [5], [6].

The physical methods to determine the secondary structure such as x-ray diffraction and NMR spectroscopy are difficult, time consuming and expensive. Therefore the computational approach to predict the secondary structure of RNA molecule is an appropriate alternative.

RNA secondary structure prediction is not a hard problem. It has been estimated that the number of secondary structures modeled from the input of n nucleotides is greater than

Abdulqader M. Mohsen is a Ph.D. candidate at Universiti Sains Malaysia (USM); e-mail: amayemen@yahoo.com.

Ahamad Tajudin Khader is an Associate Professor in the School of Computer Sciences, Universiti Sains Malaysia e-mail: tajudin@cs.usm.my.

Dhanesh Ramachandram is a Senior Lecturer in the school of Computer Sciences, USM, e-mail: dhaneshr@cs.usm.my.

Abdullatif Ghallab a Ph.D. candidate at Universiti Sains Malaysia (USM); e-mail: ghallab@cs.usm.my.

1.8^n [7]. For example, *Saccharomyces cerevisiae* (X67579) 5S rRNA with 118 nucleotides in length has an estimated of 1.3×10^{30} secondary structure models. The larger RNA, such as the *Sulfolobus acidocaldarius* (D14876) 16S rRNA, with 1493 nucleotides, has an estimated total of 1.3×10^{381} possible secondary structure models.

Currently, two different approaches study the RNA secondary structures. The first one is the single sequence approach which predicts the secondary structure by searching the minimum free energy (MFE). The second is the comparative sequence analysis. In this approach, the iterative process takes a sequence, applies accurate sequence alignments data and analyzes the structure that is common to all the sequences in the database. Most of the developed methods which based on the free energy minimization either apply dynamic programming (DP) or a metaheuristics on the domain.

This paper proposes a version of a Harmony Search (HS) algorithm calls HSRNAFold to predict the RNA secondary structure with MFE and similar to the known structure. The performance of HSRNAFold is compared with both DP and metaheuristic algorithms using standard sets of RNA test molecules. Section II provides a short overview of RNA secondary structure algorithms. RNA secondary structure prediction is presented in Section III. HSRNAFold is introduced in Section IV. The experimental results are presented in Section V. Conclusions and future work are given in Section VI.

II. RELATED WORK

The dynamic programming algorithms which based on free energy minimization of a single RNA sequence has been studied since the early 1970s. Mathews [8] provided a review on the revolutions that have occurred in the development of a number of algorithms. Ruth Nussinov et al. [9] predicted the RNA secondary structure using DP method by maximizing the number of base pairs. Then, in 1980, they adapted of their original method to use a simple nearest-neighbor energy model to enhance the results [10]. Michael Zuker and Patrick Stiegler in [11] proposed using a slightly refined DP approach that models the nearest neighbor energy interactions which directly incorporates stacking into the prediction.

Later, Zuker et al. proposed the DP algorithm which called mfold. It is still a popular algorithm to find MFE pseudoknot-free secondary structure of an RNA molecule. Moreover, it has become the benchmark for predicting the RNA secondary

structure. mfold takes the primary RNA sequence as input, and uses a complex thermodynamic model to evaluate the free energy of the structures by seeking the pseudoknot-free secondary structure with the MFE [12] [13]. RNAFold from the ViennaRNA [14] package predicts the RNA secondary structure through energy minimization. It reads an RNA sequence as input and provides three kinds of DP algorithms to predict the structure: i) the MFE algorithm to find a single optimal structure; ii) the partition function algorithm to calculate the base pair probabilities in the thermodynamic ensemble; iii) the suboptimal folding algorithm to generate all suboptimal structures based on MFE.

On the other hand, many metaheuristics algorithms were proposed such as Genetic Algorithms (GAs), Simulated Annealing (SA) and Particle Swarm Optimization (PSO). GAs have been shown to achieve higher prediction rates of base pairs than DP [15]. The more recent GAs works in this area are RnaPredict and its parallelized version P-RnaPredict. They showed that the quality is comparable to mfold [16] [17]. SARNAPredict has been introduced as a SA algorithm. [5] [18]. It has attempted to predict the RNA secondary structures with a low free energy and a high number of correctly predicted base pairs when compared to known native structures. Recently, SetPSO is a PSO algorithm was proposed by Neethling and Engelbrecht [6]. Then, HelixPSO, another version of POS, was proposed by Michael Geis and Martin Middendorf [19]. Both Algorithms were used to find secondary structures with MFE.

DP algorithm as mathematical technique can hit the global optima in the small problems. However, in real world problems there are some drawbacks. For examples, when the number of variables increases, the number of evaluations of the recursive function will also increase exponentially. For RNA Secondary structure prediction the large number of structure alternatives make it difficult to determine which one is more correct [20]. On the other hand, the drawbacks of the most existing metaheuristics approaches on the RNA secondary structure domains are: i) they required more mathematics requirements; ii) they need initial value settings for the decision variables; ii) derivative information is also necessary and iii) they not consider all the existing solutions when creating a new one [21].

Harmony Search (HS) algorithm is an optimization technique that was developed by Geem [21]. It mimics the musicians' improvisation process. The HS algorithm has been successful in a variety of optimization problems in several optimization fields such as: continuous engineering optimization, vehicle routing, combined heat and power economic dispatch, water pump switching problem, optimal scheduling of multiple dam system and transport energy modeling [22].

III. RNA SECONDARY STRUCTURE PREDICTION

RNA molecule consists of a single stranded sequence of four nucleotides: adenine (A), guanine (G), cytosine (C), or uracil (U). This linear sequence is the primary structure of RNA molecule.

The RNA strand can fold back upon itself. During the folding process, the hydrogen bonds between the different

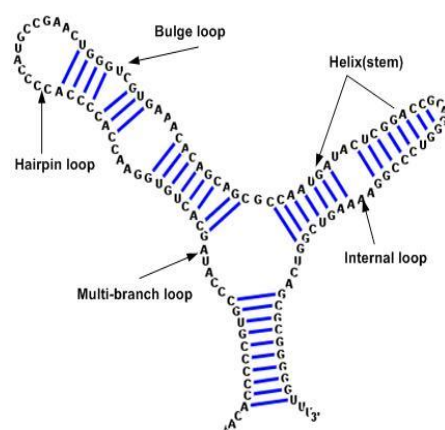


Fig. 1. RNA secondary structure components: stems (helices), interior loops, hairpin loop, multi loops and bulges loops. This figure is created using jViz.RNA [24] for the *Deinococcus radiodurans* organism.

nucleotides form the base pairs. These hydrogen bonds which occur mostly between G and C, or A and U are called the Watson-Crick base pairs and the bond between G and U is called the wobble base pair. These base pairs; GC, AU, and GU, and their mirrors, CG, UA, and UG are called the canonical base pairs. The RNA secondary structure is defined by a set of base pairs which satisfy the following constraints [17], [23], [12]:

- 1) for (i,j) , it must be canonical base pairs;
- 2) each base cannot share more than one base;
- 3) pairing bases must be at least three bases apart $i - j > 3$ and
- 4) two base pairs must not cross, i.e.: $i, j \cap i', j' = \Phi$ or for all $(i, j), (i', j')$ either $i < i' < j' < j$ or $i' < i < j < j'$ holds.

i) ii) iii) ;

The stability of the RNA secondary structure is quantified as the amount of free energy being released or used by the forming base pairs. The stability increases according to the number of GC versus AU and GU base pairs and the number of base pairs in a hairpin loop region. The number of unpaired bases decreases the stability of the structure such as interior loops, hairpin loop or bulges.

Since RNA folding is subject to the laws of thermodynamics, there is an assumption that the correct structure is a low energy structure [?]. The stability of the secondary structure depends on the amount of free energy released to form the base pairs. Thus, the more negative the free energy of a structure is, the more stable a particular sequence is formed. This structure is called the MFE secondary structure [25].

As many researchers have predicted the RNA secondary structure, several metaheuristic algorithms start by computing the set H of all the potential helices of an RNA molecule [19] [20] [26]. A helix is specified by three constraints. 1) Each helix must have at least three stacked canonical base pairs. 2) The sequence or loop connecting the two strands must be at least three nucleotides long. 3) Each helix must not share its base with others. This is done by iterating over

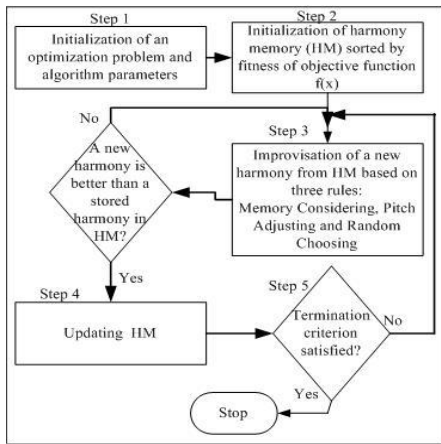


Fig. 2. The five steps of HS algorithms.

all pairs of bases and checking if they can be extended to a helix. If the helix satisfies the constraints, it must be added to H . After that the algorithm tries to find a subset of H that defines an optimal secondary structure with the MFE [19] [20] [26]. The helix size and the type of base pairs contribute to decrease the free energy of an RNA secondary structure [19]. The native structure usually has a free energy of about 5-10% from the MFE of the sequence. There are many functions that compute the free energy of RNA secondary structure based-on the different thermodynamic models.

IV. HSRNAFOLD

Geem et al in [21] introduced metaheuristic algorithm inspired by a music performance process involving search for a better harmony.

Similarities between the music improvisation and the optimization algorithms are summarized as follows [27]: Music improvisation seeks the best state (fantastic harmony) by aesthetic estimation, as the optimization algorithms seek the best state (global optimum) by objective function evaluation. Aesthetic estimation is determined by a set of pitches played by the joined instruments, just as objective function (function evaluation) is determined by a set of values of decision variables. Aesthetic sound quality (better aesthetic) can be improved by practice after practice; objective function value can be improved by iteration after iteration [21] [22] [28] [29].

HS has five steps [21] [22], as shown in Figure 1 and as explained below: Initialize the problem and algorithm parameters.

- Initialize the harmony memory (HM).
- Improvise a new harmony from HM.
- Update the harmony memory.
- Check the stopping criterion.

The descriptive detail of these steps is presented in Figure 2 and illustrated the following subsections:

A. Initialize the Problem and Algorithm Parameters

Mathematically, the general form of optimization problem can be specified as follows:

$$\begin{cases} \text{Minimize } f(x) \\ \text{Subject to } g(x) > 0, x = \{x_1, x_2, \dots, x_n\} \\ h(x) = 0 \end{cases} \quad (1)$$

Where $f(x)$ is the objective function, and $g(x)$ and $h(x)$ are the inequality and equality constraint functions respectively; x is the set of each decision variable x_i and N is the number of decision variables (music instruments).

HS algorithm has four parameters to control the solution procedure and these parameters must be specified in this step as follows:

- The harmony memory size (HMS) which represents the number of solution in the harmony memory (HM). The harmony memory is a memory location where the entire solution vectors are stored. This HM is similar to the genetic pool in the GA [12].
- Harmony memory consideration rate (HMCR) which represents the probability of picking up values from HM to the variables.
- Pitch adjusting rate (PAR) which represents the probability of further adjusting the pitch with neighboring pitches.
- The number of improvisations (NI) that represents the number of iterations to be used during the solution process, or stopping criterion.

In this step, a set of helices H is built from a predefined base pairs pool. The algorithm presented by [19] [20] [26] was used to generate these helices. The objective of helices generation is to construct the RNA stems.

B. Initialize the Harmony Memory

Initialize the HM matrix ($N * M$) where N is the length of RNA nucleotide and M is HMS. Then fill the HM randomly by generating the feasible solution vectors. These solutions are randomly created by subsets of helices from H . Thus, these solutions must satisfy all RNA secondary structure constraints. The HMS and their corresponding fitness function values are shown below:

$$HM = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_{N-1}^1 & x_N^1 \\ x_1^2 & x_2^2 & \dots & x_{N-1}^2 & x_N^2 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ x_1^{HMS-1} & x_2^{HMS-1} & \dots & x_{N-1}^{HMS-1} & x_N^{HMS-1} \\ x_1^{HMS} & x_2^{HMS} & \dots & x_{N-1}^{HMS} & x_N^{HMS} \end{bmatrix} \begin{matrix} \Rightarrow f(x^1) \\ \Rightarrow f(x^2) \\ \Rightarrow \vdots \\ \Rightarrow f(x^{HMS-1}) \\ \Rightarrow f(x^{HMS}) \end{matrix} \quad (2)$$

Where each x_i vector and $f(x)$ represent a feasible RNA secondary structure and it's corresponding free energy function respectively.

Each structure is evaluated by using RNAeval algorithm from the ViennaRNA and all solutions in the HM are sorted out based on the free energies progressively.

TABLE I
TEST RNA SEQUENCES WITH THEIR ORGANISM, CLASS, ACCESSION NUMBER AND SIZE

Organism	Accession No.	RNA class	Size
Saccharomyces cerevisiae	X67579	5S rRNA	118
Haloarcula marismortui	AF034620	5S rRNA	122
Aureocoumbra lagunensis	U40258	Group I intron, 16S rRNA	468
Drosophila virilis	X05914	16S rRNA	784
Xenopus laevis	M27605	16S rRNA	945
Sulfolobus acidocaldarius	D14876	16S rRNA	1493

C. Improvise a New Harmony

Generating a new harmony is called 'improvisation' [21]. A new harmony vector $x' = (x'_1 x'_2 \dots x'_N)$, is generated based on three parameters: memory consideration, pitch adjustment and random selection. It is generated as follows: i) for each component x'_i , pick the component of x'_i randomly from any of the values in the specified HM range ($x'_i - x'_i^{HMS}$) with the probability of P_{hmcr} ; ii) the rest of the components of x'_i are picked by random value with the probability of $1 - P_{hmcr}$. For example, a HMCR of 0.95 indicates that the probability of HS algorithm to choose the decision variable value from historically stored values in the HM is 95 % and the probability of choosing a new random value in the allowed range is (100- 95) % and iii) change x'_i has the probability of P_{par} with small amount (bw) of changes taking place, for pitch adjustments: $x'_i \leftarrow x'_i \pm bw * rand()$. It has been noted that the P_{par} will be tiny and leads to zero, because changing the position in the secondary structure usually yields infeasible structure.

D. Harmony Memory Update

Evaluate the new harmony (structure) $x' = (x'_1 x'_2 \dots x'_N)$ by calculating its energy. If the value of its objective function is better than the objective function of the worst harmony in the HM, the new harmony is included in the HM and the existing worst harmony is excluded from the HM. Subsequently, the vectors are sorted out based-on their free energies.

E. Termination criterion check

The process repeats step IV-C and IV-D until the maximum number of iterations (number of improvisations) is reached.

The pseudo code of the modified HS algorithm (HSR-NAFold) is shown in Figure 3.

V. EXPERIMENTAL RESULTS

The implementation of the proposed algorithm was implemented by C#. For the experiments, six different RNA sequences lengths were used. Since these sequences have already been used by other authors, HSRNAFold can be compared with HelixPSO, RnaPredict, SetPSO and mfold.

The test sequences were taken from the comparative RNA website [30].

These sequences represent good variety of sequence lengths, organisms, and RNA types. The tested RNA sequences are listed in Table I together with their organism, accession

```

Generate all possible base-pairs pool.
Generate all possible helices from the base-pairs pool.
Initial HS parameters: HMCR, RSR and PAR
for i= 1 to HMS
  Generate feasible structure[i] randomly from the helices
  FreeEnergy[i] = Evaluate (structure[i])
  Sort the generated structures according to the freeEnergy
for (i=1 to number_of_ iterations)
  Generate new feasible newstructure depending on the HMCR,
  RSR and PAR
  newFreeEnergy = Evaluate (newstructure);
  If (newFreeEnergy < FreeEnergy[HMS])
    structure [HMS]= newstructure
  Sort the generated structures according to the freeEnergy
Return structure with minimum free energy

```

Fig. 3. The pseudo code of HSRNAFold algorithm.

number, class and size. Each chosen sequence has a known structure available for comparison. These structures were determined by comparative methods.

Table II shows the suitable parameters setting of HSRNAFold after 50 runs and the parameters that Neethling et al. and Geis et al. in [6] and [19] used to implement SetPSO, HelixPSO, and to re-implement RnaPredict.

For the short sequences (lengths 118 and 122) and medium sequences (lengths 468 and 784) the values for HMSR were 0.95 and 0.90 are used respectively. For the long RNA sequence of length (945 and 1492) the values of HMSR were 0.90 and 95 respectively. For PAR parameter the value was 0.50.

HSRNAFold ran over 20 times for all the test sequences in the search space including the maximal number of helices. Free energies of the best secondary structure that were found by both HSRNAFold and RnaPredict algorithms after 280000 generations are shown in Table III. It can be seen that HSRNAFold found a free energy of secondary structure between (-9.35 and -52.16) kcal/mol. This is better than the free energy found by RnaPredict.

The free energies of the best secondary structure that was found by both algorithms, SetPSO and HSRNAFold after 35000 generations are shown in Table IV. It can be seen that HSRNAFold found secondary structures which have free energy between (-0.5 and -80.46) kcal/mol is better than the secondary structures found by SetPSO.

Comparing HSRNAFold with HelixPSO, it is obvious that HSRNAFold achieved better results in term of finding the MFE between (-7.25 and -47.26) kcal/mol. Free energies of the

TABLE II

THE SETTING OF SIZE AND NUMBER OF ITERATION PARAMETERS FOR RNAPREDICT, SETPSO, HELIXPSO, AND HSRNAFOLD

	RnaPredict	SetPSO	HelixPSO	HSRNAFold
Size	Population size 700	Swarm size 50	Swarm size 500	If No. of helices $j = 500$ then HMS= H.size Else HMS= 500
No of Iter.	280,000	35000	280,000	280,000

TABLE III

FREE ENERGY (ΔG IN KCAL/MOL) FOUND BY RNAPREDICT AND HSRNAFOLD AFTER THE GENERATION OF 280000 SECONDARY STRUCTURES; THE BEST RESULTS ARE SHOWN IN BOLD.

[!h]

RNA size	RnaPredict	HSRNAFold
118	-	-53.9
122	-	-56.62
468	-124.4	-133.75
784	-124.3	-151.26
945	-207.8	-259.96
1493	-633.8	-662.3

TABLE IV

FREE ENERGY (ΔG IN KCAL/MOL) FOUND BY SETPSO AND HSRNAFOLD AFTER THE GENERATION OF 35000 SECONDARY STRUCTURES; THE BEST RESULTS ARE SHOWN IN BOLD.

RNA Size	SetPSO	HSRNAFold
118	-53.4	-53.9
122	-48.42	-56.62
468	-	-133.75
784	-105.8	-151.26
945	-173.3	-253.76
1493	-	-662.3

best secondary structure that were found by both algorithms after 280000 generations are shown in Table V.

TABLE V

FREE ENERGY (ΔG IN KCAL/MOL) OF SECONDARY STRUCTURES FOUND AFTER 280000 SECONDARY STRUCTURES HAVE BEEN GENERATED BY HELIXPSO AND HSRNAFOLD; THE BEST RESULTS ARE SHOWN IN BOLD.

RNA Size	HelixPSO	HSRNAFold
118	-	-53.9
122	-	-56.62
468	-126.5	-133.75
784	-125.7	-151.26
945	-212.7	-259.96
1493	-653.5	-662.3

On the other hand, HSRNAFold achieved comparable results compared with the DP benchmark mfold. The free energies of the best secondary structure that were found by both algorithms at the end of a run are shown in Table VI.

Figures V and V show that the secondary structure predicted by HSRNAFold and SetPSO have been compared with the known secondary structure for the *Saccharomyces cerevisiae* sequence. The dark grey lines represent the base pairs for both the known and predicted structure. The light grey lines represent the predicted base pair which is not found in the known structure. The black lines indicate the base pairs in the known structures which have not been predicted.

TABLE VI

FREE ENERGY (ΔG IN KCAL/MOL) OF SECONDARY STRUCTURES FOUND AFTER 280000 SECONDARY STRUCTURES HAVE BEEN GENERATED BY MFOLD AND HSRNAFOLD; THE BEST RESULTS ARE SHOWN IN BOLD.

RNA length	mfold	HSRNAFold
118	-53.5	-53.9
122	-56.44	-56.62
468	-140.5	-133.75
784	-146.3	-151.26
945	-250.6	-259.96
1493	-803.3	-662.3

It is noted that HSRNAFold could find 33 base pairs out of 37, which is 89.2% of the known base pairs while SetPSO 28 could only find 28 base pairs out of 37(75.7%). The result with regards to mfold was 89.2%. From Figure V, it

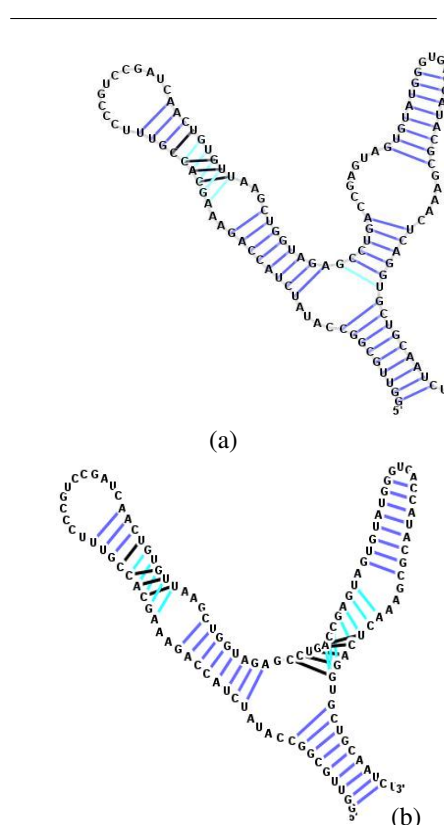


Fig. 5. The secondary structures of *Saccharomyces cerevisiae* 5S rRNA (X67579) predicted by HSRNAFold and SetPSO compared to the native structure. (a) HSRNAFold structure to native structure. (b) SetPSO structure to native structure.

can be observed that HSRNAFold algorithm performance is influenced by the length of RNA sequence. That means finding structure with MFE only needs less iterations for

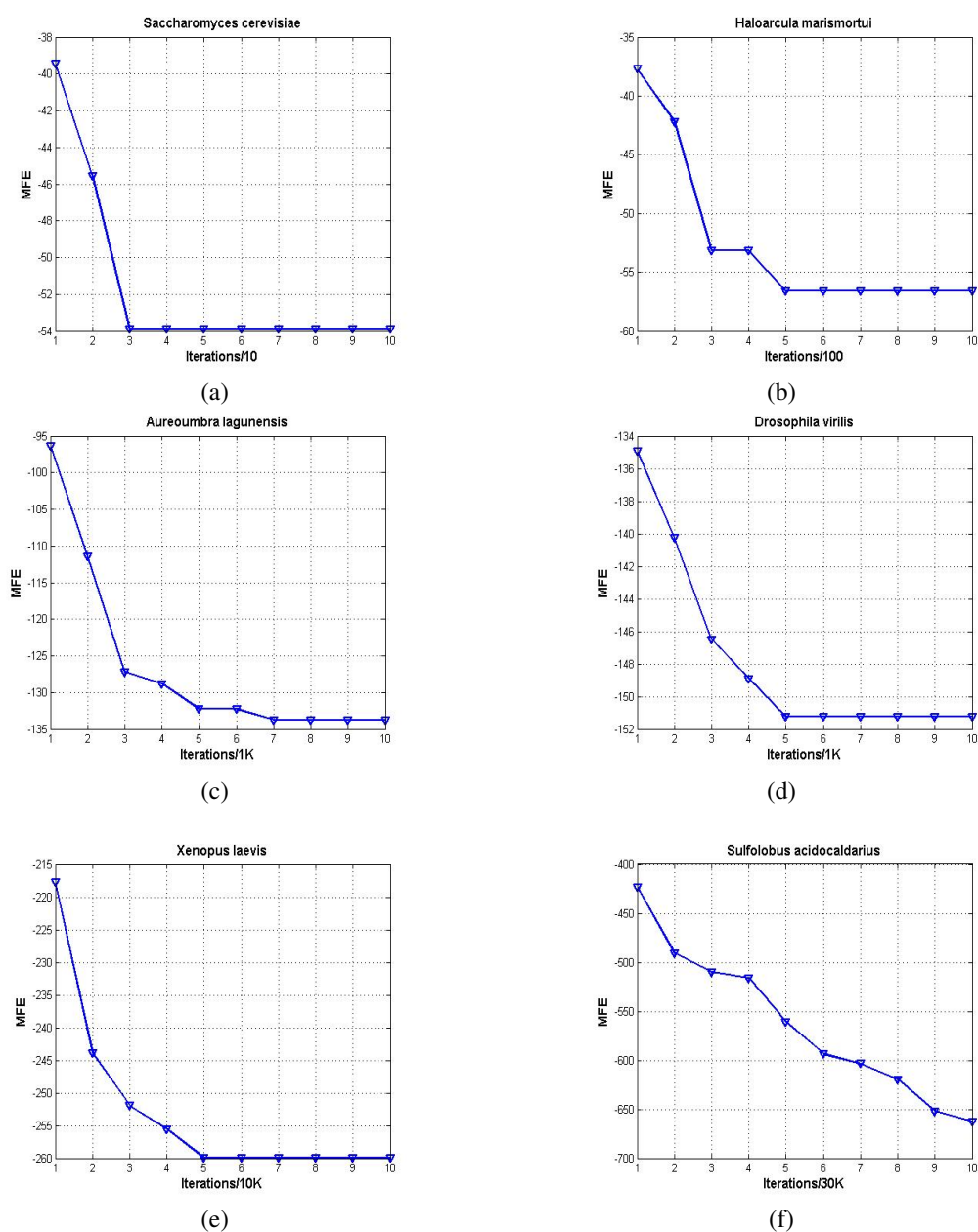


Fig. 4. The optimization behavior of HSRNAFold with all test sequences. (a) *Saccharomyces cerevisiae* over 1000 generations; (b) *Haloarcula marismortui* over 1000 generations; (c) *Aureoumbra lagunensis* over 10000 generations; (d) *Drosophila virilis* over 10000 generations; (e) *Xenopus laevis* over 100000 generations; and (f) *Sulfolobus acidocaldarius* over 300000

the short sequences, whereas for the long RNA sequence, it needs a more iterations. For all sequences, they required short time (iterations) to find the optimum solution except *Sulfolobus acidocaldarius*. For example, the number of iterations required to find the minimum free energies of *Saccharomyces cerevisiae* and *Sulfolobus acidocaldarius* are around 30 and 300000 iterations respectively. As noted in Figure 5, the trend goes down gradually towards the optimum solution in the first iterations. Then, it slowly approaches the optimum solution in the last iterations. This behavior can be demonstrated in the initial iterations whereby the diversity is seen to be high. Consequently, the opportunity to generate

better solution can be optimally attained. In last iterations, the convergence in the search space is seen to be high. Therefore, the opportunity to generate good solution than the worst solution in the search space is less. As a result, generating a better solution will lead to the increase of the number of iteration to reach the optimum solution.

VI. CONCLUSION

In this paper, a harmony search algorithm (HS) called HSRNAFold is presented to find the RNA secondary structure with MFE. Compared to DP and metaheuristics algorithms,

HSRNAFold is found to be comparable with mfold. It achieved better results than GA and PSO in predicting the RNA secondary structure with MFE and similar to the known structure.

On the other hand, HSRNAFold requires long time to reach the optimum solution with the large sequences compared to the small sequences.

For future work, refinement of the helix generation algorithm and study on the effect of the HS parameters to enhance the results should be investigated. Code optimization, effect of parameters adaptation and the possibility to hybridize HS with other optimization algorithms to enhance the HS performance is needed to be studied.

REFERENCES

- [1] J. A. Doudna and T. R. Cech, "The chemical repertoire of natural ribozymes," *Nature*, vol. 418, no. 6894, pp. 222–228, 2002.
- [2] J. L. Hansen, T. M. Schmeing, P. B. Moore, and T. A. Steitz, "Structural insights into peptide bond formation," in *Proceedings of the National Academy of Sciences*, vol. 99, 2002, pp. 11 670–11 675.
- [3] J. Bachelier, J. Cavaill, and A. Httenhofer, "The expanding snorna world," *Biochimie*, vol. 84, no. 8, pp. 775–790(16), August 2002.
- [4] G. Meister and T. Tuschl, "Mechanisms of gene silencing by double-stranded rna," *Nature*, vol. 431, pp. 343–349, 2004.
- [5] H. Tsang and K. Wiese, "Sarna-predict: A study of rna secondary structure prediction using different annealing schedules," in *Proceedings of the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, 2007, pp. 239–246.
- [6] M. Neethling and A. Engelbrecht, "Determining rna secondary structure using set-based particle swarm optimization," in *IEEE Congress on Evolutionary Computation(CEC2006)*, 2006, pp. 1670–1677.
- [7] K. J. Doshi, J. J. Cannone, C. W. Cobaugh, and R. R. Gutell, "Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for rna secondary structure prediction," *BMC Bioinformatics*, vol. 5, no. 105, pp. 1471–2105, 2004.
- [8] D. H. Mathews, "Revolutions in rna secondary structure prediction," *Journal of Molecular Biology*, vol. 359, p. 526532, 2006.
- [9] R. Nussinov, G. Pieczenik, J. R. Griggs, and D. J. Kleitman, "Algorithms for loop matchings," *SIAM Journal on Applied Mathematics*, vol. 35, no. 1, pp. 68–82, 1978.
- [10] R. Nussinov and A. Jacobson, "Fast algorithm for predicting the secondary structure of single-stranded rna," *Proc Natl Acad Sci U S A*, vol. 77, no. 11, pp. 6309–6313, 1980.
- [11] M. Zuker and P. Stiegler, "Optimal computer folding of large rna sequences using thermodynamics and auxiliary information," *Nucl. Acids. Res.*, vol. 9, no. 1, pp. 133–148, 1981.
- [12] M. Zuker, "Prediction of rna secondary structure by energy minimization," in *Computer Analysis of Sequence Data*, ser. Methods in Molecular Biology, M. G. Annette and G. G. Hugh, Eds. Humana Press, 1994, pp. 267–294, 10.1385/0-89603-276-0:267.
- [13] —, "Mfold web server for nucleic acid folding and hybridization prediction," *Nucleic Acids Research*, vol. 31, no. 13, p. 34063415, 2003.
- [14] I. Hofacker, W. Fontana, P. Stadler, S. Bonhoeffer, M. Tacker, and P. Schuster, "Fast folding and comparison of rna secondary structures (the vienna rna package)," *Monatshfte f. Chemie*, vol. 125, pp. 167–188, 1994.
- [15] A. Gultyaev, F. van Batenburg, and C. Pleij, "Dynamic competition between alternative structures in viroid rnas simulated by an rna folding algorithm," *Journal of Molecular Biology*, vol. 276, no. 1, pp. 43–55, 1998.
- [16] K. Wiese, A. Deschnes, and A. Hendriks, "Rnapredict - an evolutionary algorithm for rna secondary structure prediction," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 5, no. 1, pp. 25–41, 2007.
- [17] K. C. Wiese and A. Hendriks, "Comparison of p-rnapredict and mfold algorithms for rna secondary structure prediction," *Bioinformatics*, vol. 22, no. 8, p. 934942, 2006.
- [18] H. Tsang and K. Wiese, "The significance of thermodynamic models in the accuracy improvement of rna secondary structure prediction using permutation-based simulated annealing," in *Proceedings of the IEEE Congress on Evolutionary Computation*, Singapore, 2007, pp. 3879–3885.
- [19] M. Geis and M. Middendorf, "A particle swarm optimizer for finding minimum free energy rna secondary structures," in *Swarm Intelligence Symposium, 2007. SIS 2007. IEEE*, 2007, pp. 1–8.
- [20] H. H. Tsang, "Sarna-predict: A permutation-based simulated annealing algorithm for rna secondary structure prediction," Ph.D. dissertation, Simon Fraser University, 2007.
- [21] Z. W. Geem, J. H. Kim, and G. Loganathan, "A new heuristic optimization algorithm: Harmony search," *SIMULATION*, vol. 76, no. 2, pp. 60–68, 2001.
- [22] M. Mahdavi, M. Fesanghary, and E. Damangir, "An improved harmony search algorithm for solving optimization problems," *Applied Mathematics and Computation*, vol. 188, no. 2, pp. 1567–1579, 2007.
- [23] D. H. Mathews, "Predicting rna secondary structure by free energy minimization," *Theoretical Chemistry Accounts:Theory, Computation, and Modeling*, vol. 116(1-3), pp. 160–168, 2006.
- [24] K. C. Wiese, E. Glen, and A. Vasudevan, "jviz.rna - a java tool for rna secondary structure visualization," *IEEE Transactions on NanoBio-science*, vol. 4, no. 3, pp. 212–218, 2005.
- [25] D. M. Layton and R. Bundschuh, "A statistical analysis of rna folding algorithms through thermodynamic parameter perturbation," *Nucleic Acids Research*, vol. 33, no. 2, p. 519524, 2005.
- [26] A. Hendriks, "A parallel evolutionary algorithm for rna secondary structure prediction," Ph.D. dissertation, Simon Fraser University, May2005.
- [27] B. Alatas, "Chaotic harmony search algorithms," *Applied Mathematics and Computation*, vol. 216, no. 9, pp. 2687 – 2699, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/B6TY8-4YPPPWP-6/2/489c124258d50da8d1930a2be60e777a>
- [28] K. S. Lee and Z. W. Geem, "A new meta-heuristic algorithm for continuous engineering optimization: harmony search theory and practice," *Computer Methods in Applied Mechanics and Engineering*, vol. 194, no. 36-38, pp. 3902–3933, 2005.
- [29] Z. W. Geem, "Harmony search in water pump switching problem," *ICNC*, vol. 3, pp. Pp 751–760, 2005.
- [30] J. J. Cannone, S. Subramanian, M. N. Schnare, J. R. Collett, L. D'Souza, L. N. Du Y, Feng B, M. L.V, M. K.M, P. N, Y. N. Shang Z, and G. R.R., "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. 35, pp. 169–172, 2002.