

## Research Article

# A Novel Framework Based on ACO and PSO for RNA Secondary Structure Prediction

Gang Wang,<sup>1,2,3</sup> Wen-yi Zhang,<sup>4</sup> Qiao Ning,<sup>4</sup> and Hui-ling Chen<sup>5</sup>

<sup>1</sup> College of Computer Science and Technology, Jilin University, Changchun 130012, China

<sup>2</sup> Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China

<sup>3</sup> College of GeoExploration Science and Technology, Jilin University, Changchun 130026, China

<sup>4</sup> School of Computer Science and Information Technology, Northeast Normal University, Changchun 130024, China

<sup>5</sup> College of Physics and Electronic Information, Wenzhou University, Chashan University Town, Wenzhou, Zhejiang, 325035, China

Correspondence should be addressed to Hui-ling Chen; [chenhuiling.jlu@gmail.com](mailto:chenhuiling.jlu@gmail.com)

Received 24 June 2013; Accepted 12 August 2013

Academic Editor: William Guo

Copyright © 2013 Gang Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Prediction of RNA structure is a useful process for creating new drugs and understanding genetic diseases. In this paper, we proposed a particle swarm optimization (PSO) and ant colony optimization (ACO) based framework (PAF) for RNA secondary structure prediction. PAF consists of crucial stem searching (CSS) and global sequence building (GSB). In CSS, a modified ACO (MACO) is used to search the crucial stems, and then a set of stems are generated. In GSB, we used a modified PSO (MPSO) to construct all the stems in one sequence. We evaluated the performance of PAF on ten sequences, which have length from 122 to 1494. We also compared the performance of PAF with the results obtained from six existing well-known methods, SARNAPredict, RnaPredict, ACRNA, PSOfold, IPSO, and mfold. The comparison results show that PAF could not only predict structures with higher accuracy rate but also find crucial stems.

## 1. Introduction

RNA functions as an information carrier, catalyst, and regulatory element, perhaps reflecting its importance in the earliest stages of evolution. The structures of RNAs provide insight into the mechanisms behind these functions. Determining sequence is the first step in determining structure, and many billions of nucleotide sequences are now known. The second step is determining secondary structure, and relatively few classes of RNAs currently have known secondary structures [1]. The RNA secondary structure prediction problem is a critical one in molecular biology. Secondary structure as well as tertiary structure can be determined by X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy. Data analysis tools used for prediction of RNA structure are mainly based on dynamic programming [2]. Recently,

the metaheuristic methods are widely used to predict RNA secondary structure. These methods generally include genetic algorithm (GA) [3], particle swarm optimization (PSO) [4], ant colony optimization (ACO) [5], and simulated annealing (SA) [6]. The state-of-the-art methods are introduced as follows.

For SA, Shapiro and Wu [7] later modified the algorithm by introducing an annealing mutation operator. Tsang and Wiese [8] presented SARNAPredict, the permutation-based algorithm for RNA secondary structure prediction based on SA with a simple thermodynamic model and studied mainly its convergence behavior. A performance evaluation of SARNAPredict in terms of prediction accuracy was made via comparison with eight state-of-the-art RNA prediction algorithms. The results presented in this paper demonstrate

that SARNA-Predict can outperform other state-of-the-art algorithms in terms of prediction accuracy. Furthermore, there is substantial improvement of prediction accuracy by incorporating a more sophisticated thermodynamic model.

*For GA.* Benedetti and Morosetti [9] compared the accuracy of an EA against known RNA structures with the objective of finding optimal and suboptimal structures that were similar. Shapiro et al. [10] modified their EA to study folding pathways using a massively parallel genetic algorithm. Wiese et al. designed a serial EA, RnaPredict [11], which encodes RNA secondary structures as permutations. RnaPredict was parallelized via a coarse-grained distributed EA for RNA secondary structure prediction [12].

*For ACO.* Yu et al. [13] put forward ACRNA based on ACO for RNA secondary structure prediction. For a given RNA sequence, the set of all possible stems is obtained, and the energy of each stem is calculated and stored at the initial stage. Furthermore, a more realistic formula is used to compute the energy of multibranch loop in the following iteration. Then a folding pathway is simulated, including such processes as construction of the heuristic information, the rule of initializing the pheromone, the mechanism of choosing the initial and next stem, and the strategy of updating the pheromone between two different stems.

*For PSO.* Geis and Middendorf [14] introduced HelixPSO for finding minimum energy RNA secondary structures. Neethling and Engelbrecht [15] proposed a set-based Particle Swarm Optimization algorithm to optimize the structure of an RNA molecule, using an advanced thermodynamic model. Liu et al. [16] proposed an improved PSO (IPSO). The authors designed an efficient objective function according to the minimum free energy, the number of selected stems, and the average length of selected stems. A promising experimental result was obtained, and the effectiveness and practicability of IPSO for RNA secondary structure prediction was shown. Xing et al. [17, 18] proposed PSOfold based on IPSO. An adaptive parameter controller of PSO based on fuzzy logic is used to improve the balance between exploration and exploitation. Solution conversion strategy (SCS) is designed to enhance the PSO performance in discrete problem such as stem combination.

The metaheuristic methods [19, 20] mentioned previously only paid attention to structure of the sequence but ignored the searching capability of algorithm. PSOfold is our previous work with an attempt to enhance the searching range in PSO, but the crucial stems are not considered. In this paper, we proposed PAF to improve the exploration ability and focused on the influence of the crucial stems.

The main contributions of this paper are described as follows.

- (1) A framework, namely, PAF, was proposed for RNA secondary structure prediction, which includes CSS and GSB.
- (2) In CSS, MACO was proposed to search the crucial stems.

- (3) In GSB, MPSO was designed to construct all the stems in one sequence.

The rest of the paper is organized as follows. Section 2 briefly introduces the theory of ACO and PSO. In Section 3, we present the proposed PAF framework and describe the implementation of CSS and GSB. In Section 4, the experiment results on various public sequences are discussed. Finally, we draw the conclusions of this paper in Section 5.

## 2. Basic Theory

*2.1. ACO.* ACO algorithm is biologically inspired from the behavior of colonies of real ants, and in particular how they forage for food. ACO has been formalized into a metaheuristic for combinatorial optimization problems by Dorigo and coworkers [21].

In ACO, an ant  $k$  being in node  $i$  chooses the next node  $j$  with a probability given by the random proportional rule defined as follows [22].

(a) *State Transition Rule.* Consider

$$P_{ij}(t) = \begin{cases} \frac{[\tau_{ij}(t)] \cdot [\eta_{ij}]^\beta}{\sum_{j \in I_i^k} [\tau_{ij}(t)] \cdot [\eta_{ij}]^\beta} & \text{if } j \in I_i^k, \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where  $I_i^k$  is its feasible neighborhood. The feasible neighborhood excludes nodes already visited in the partial tour of ant  $i$ , and it may be further restricted to a candidate set of the nearest neighbors of a city  $i$ . Once an ant has visited all nodes, it returns to its starting node.

(b) *State Updating Rule.* After all ants have completed their solutions, pheromone evaporation on all nodes is triggered according to (2). The pheromone on each edge is updated according to the following equation:

$$\tau_{ij}(t+1) = (1 - \rho) \cdot \tau_{ij}(t) + \sum_{k=1}^N \Delta\tau_{ij}^k(t), \quad (2)$$

where  $m$  is the number of ants at each iteration and  $\rho \in (0, 1)$  is the pheromone evaporation rate. Consider

$$\Delta\tau_{ij}^k = \begin{cases} \frac{Q}{L_k} & \text{if } \tau_{ij} \in \text{tour done by ant } k, \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

where  $L_k$  denotes the tour length and  $Q$  is a predefined constant.

*2.2. PSO.* PSO originated from the simulation of social behavior of birds in a flock [23, 24]. In PSO, each particle flies in the search space with a velocity adjusted by its own flying memory and its companion's flying experience. All particles

have objective function values which are decided by a fitness function. Consider the following:

$$v_{id}^t = \omega \times v_{id}^{t-1} + c_1 \times r_1 (p_{id}^t - x_{id}^t) + c_2 \times r_2 (p_{gd}^t - x_{id}^t),$$

$$d = 1, 2, 3, 4, \dots, D, \quad (4)$$

where  $c_1$  indicates the cognition learning factor,  $c_2$  indicates the social learning factor, and  $r_1$  and  $r_2$  are random numbers uniformly distributed in  $U(0, 1)$ . Each particle then moves to a new potential solution based on the following equation:

$$x_{id}^{t+1} = x_{id}^t + v_{id}^t, \quad d = 1, 2, \dots, D. \quad (5)$$

Kennedy and Eberhart [25] proposed a binary PSO in which a particle moves in a state space restricted to 0 and 1 on each dimension, in terms of the changes in probabilities that a bit will be in one state or the other. Consider the following:

$$x_{id} = \begin{cases} 1, & \text{rand}() < S(v_{i,d}), \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

$$S(v) = \frac{1}{1 + e^{-v}}.$$

The function  $S(v)$  is a sigmoid limiting transformation and  $\text{rand}()$  is a random number selected from a uniform distribution in  $[0.0, 1.0]$ .

### 3. PAF

#### 3.1. CSS

##### 3.1.1. MACO

(1) *State Transition Rule.* The task of each ant is to build a set of stems. The ants find an RNA secondary structure via a probabilistic decision rule to move through adjacent states. An ant selects a stem as follows:

$$P_i = \begin{cases} \frac{[\tau_i]^\alpha [\eta_i]^\beta}{\sum_{j \in I_i^k} [\tau_j]^\alpha [\eta_j]^\beta} & \text{if } j \in I_i^k, \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

where  $\alpha$  and  $\beta$  are the regulatory factors,  $\tau_i$  is the amount of pheromone trail on stem  $i$ ,  $\eta_i$  is the priori available heuristic information, and  $I_i^k$  is the remaining stems.

(2) *State Update Rule.* The pheromone trails are updated according to (8) and (9). Consider

$$\tau = (1 - \rho) \cdot \tau_i + \sum_{k=1}^n \Delta \tau_i^k, \quad (8)$$

where  $\rho$  is the pheromone trail evaporation rate and  $\tau_i$  is the quantity per unit of length of the trail substance that is laid on stem  $i$  by the  $k$ th ant. Also,

$$\Delta \tau_i^k = \begin{cases} \text{Energy} & \text{if } k \text{ uses stem } i \text{ in its tour} \\ 0 & \text{otherwise,} \end{cases} \quad (9)$$

where Energy represents the quality of an ant's solution. If stem  $i$  is not included, the zero is returned.

3.1.2. *Algorithm of MACO.* See Algorithm 1.

#### 3.2. GSB

3.2.1. *MPSO.* MPSO was modified based on our previous studies IPSO [16] and PSOfold [18] which could predict the RNA secondary structure with excellent performance. The objective function is improved according to the size of stem and the number of pseudoknots. Consider the following:

$$f_i = aE_i - b \left[ \frac{\sum_{j=1}^N \delta(j)}{P} \right] - c \left[ \frac{\sum_{k=1}^M \varphi(k) / L + S}{W} \right], \quad (10)$$

where  $a$ ,  $b$ ,  $c$  is the weight;  $E_i$  is the free energy for the secondary structure in the  $i$ th particle;  $\delta$  is the number of pairs of the  $j$ th stem;  $P$  is the length of possible pairs;  $\varphi$  is the size of stem  $k$  which is higher than 4;  $M$  is the number of selected stems;  $L$  is the total number of stems which are higher than 4;  $S$  is the size of pseudoknots;  $W$  is the size of possible pairs.

3.2.2. *Algorithm of MPSO.* See Algorithm 2.

## 4. Results

The parameter details of ACRNA are number of ants = 100, number of iterations = 600,  $\rho = 0.2$ ,  $\alpha = 1$ , and  $\beta = 1$ . For IPSO, number of particles = 100, number of iterations = 600,  $\omega = 0.9$ ,  $c_1 = 2$ , and  $c_2 = 2$ . For PSOfold, number of particles = 100, number of iterations = 600,  $\omega = 0.9$ ,  $c_1 = 2$ , and  $c_2 = 2$ . For CSS, number of ants = 100, number of iterations = 600,  $\rho = 0.2$ ,  $\alpha = 1$ , and  $\beta = 1$ . For GSB, number of particles = 100, number of iterations = 600,  $\omega = 0.9$ ,  $c_1 = 2$ , and  $c_2 = 2$ . To generate the mfold results presented here, the mfold Web server version 3.1 was used with default settings. One noteworthy setting is the percentage of suboptimality. This percentage allows the user to control the number of suboptimal structures predicted by mfold. In this experiment, the value was set to return the 5 percent lowest energy structures.

The measures used for prediction accuracy on the majority of documents currently are sensitivity, specificity, and  $F$ -measure. In RNA secondary structure prediction, TP (true positive) indicates the number of base pairs predicted correctly; FN (false negative) denotes the number of base pairs which existed in real structure but were not predicted correctly; FP (false positive) represents the number of base pairs which existed in no real structure but was mistakenly

TABLE 1: RNA sequence details.

Organism	Accession number	RNA class	Length	Base pairs in known structure
<i>Drosophila virilis</i>	X05914	16S rRNA	784	233
<i>Xenopus laevis</i>	M27605	16S rRNA	945	251
<i>Haloarcula marismortui</i>	AF034620	5S rRNA	122	38
<i>Aureoumbra lagunensis</i>	U40258	Group I intron, 16S rRNA	468	113
<i>Metahizium anisopliae</i> var.	AF197120	Group I intron, 23S rRNA	394	120
<i>Hildenbrandia rubra</i>	L19345	Group I intron, 16S rRNA	543	138
<i>Acanthamoeba griffini</i>	U02540	Group I intron, 16S rRNA	556	131
<i>Homo sapiens</i>	J01415	16S rRNA	954	266
<i>Ailurus fulgens</i>	Y08511	16S rRNA	964	265
<i>Sulfolobus acidocaldarius</i>	D14876	16S rRNA	1494	468

TABLE 2: A comparison of the highest matching base pair structures from PAF and mfold in terms of sensitivity, specificity, and  $F$ -measure.

Sequence	Known bps	Predicted bps		TP		FP		FN		Sensitivity		Specificity		$F$ -measure	
		PAF	mfold	PAF	mfold	PAF	mfold	PAF	mfold	PAF	mfold	PAF	mfold	PAF	mfold
<i>D. virilis</i>	233	<b>247</b>	252	<b>103</b>	82	<b>144</b>	170	<b>130</b>	151	<b>44.2</b>	35.2	<b>41.7</b>	32.5	<b>42.9</b>	33.8
<i>X. laevis</i>	251	<b>235</b>	245	<b>116</b>	113	<b>119</b>	132	<b>135</b>	138	<b>46.2</b>	45.0	<b>49.4</b>	46.1	<b>47.7</b>	45.6
<i>H. marismortui</i>	38	<b>31</b>	34	28	<b>29</b>	<b>3</b>	5	10	<b>9</b>	73.7	<b>76.3</b>	<b>90.3</b>	85.3	<b>81.2</b>	80.6
<i>A. lagunensis</i>	113	<b>124</b>	133	<b>85</b>	74	<b>39</b>	59	<b>28</b>	39	<b>75.2</b>	65.5	<b>68.6</b>	55.6	<b>71.7</b>	60.2
<i>M. anisopliae</i> var.	120	<b>110</b>	116	91	<b>92</b>	<b>19</b>	24	29	<b>28</b>	75.8	<b>76.7</b>	<b>82.7</b>	79.3	<b>79.1</b>	78.0
<i>H. rubra</i>	138	<b>141</b>	167	82	<b>83</b>	<b>59</b>	84	56	<b>55</b>	59.4	<b>60.1</b>	<b>58.1</b>	49.7	<b>58.8</b>	54.4
<i>A. griffini</i>	131	<b>160</b>	174	<b>102</b>	95	<b>58</b>	79	<b>29</b>	36	<b>77.9</b>	72.5	<b>63.8</b>	54.6	<b>70.1</b>	62.3
<i>H. sapiens</i>	266	260	<b>258</b>	<b>124</b>	95	<b>136</b>	163	<b>142</b>	171	<b>46.6</b>	35.7	<b>47.7</b>	36.8	<b>47.2</b>	36.3
<i>A. fulgens</i>	265	249	<b>241</b>	<b>99</b>	74	<b>150</b>	167	<b>166</b>	191	<b>37.4</b>	27.9	<b>39.8</b>	30.7	<b>38.5</b>	29.2
<i>S. acidocaldarius</i>	468	<b>483</b>	496	<b>281</b>	271	<b>202</b>	225	<b>187</b>	197	<b>60</b>	57.9	<b>58.1</b>	54.6	<b>59.1</b>	56.2
Averages	202.3	<b>204</b>	211.6	<b>111.1</b>	100.8	<b>92.9</b>	110.8	<b>91.2</b>	101.5	<b>59.64</b>	55.28	<b>60.02</b>	52.52	<b>59.63</b>	53.66

TABLE 3: A comparison of the highest matching base pair structures from SARNA-predict, RnaPredict, ACRNA, PAF, PSOfold, and mfold in terms of sensitivity and specificity.

Sequence	Length	SARNA-Predict		RnaPredict		ACRNA		PAF		PSOfold		mfold	
		Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp
<i>D. virilis</i>	784	42.1	<b>42.2</b>	27.9	26.9	27.9	26.9	<b>44.2</b>	41.7	42.9	41.5	35.2	32.5
<i>X. laevis</i>	945	<b>46.2</b>	48.9	37.1	38.8	45.8	46.9	<b>46.2</b>	<b>49.4</b>	45.6	48.5	45.0	46.1
<i>H. marismortui</i>	122	71.1	90.0	71.1	90.0	<b>76.3</b>	78.4	73.7	<b>90.3</b>	71.1	87.1	<b>76.3</b>	85.3
<i>A. lagunensis</i>	468	74.3	64.6	60.2	51.9	63.7	54.1	75.2	<b>68.6</b>	<b>76.1</b>	64.7	65.5	55.6
<i>M. anisopliae</i> var.	394	63.3	65.0	62.5	62.0	63.3	65.0	75.8	<b>82.7</b>	75.0	74.4	<b>76.7</b>	79.3
<i>H. rubra</i>	543	<b>64.5</b>	56.7	57.2	49.1	58.0	47.9	59.4	<b>58.1</b>	58.0	53.0	60.1	49.7
<i>A. griffini</i>	556	74.0	58.8	61.8	50.3	73.3	55.2	<b>77.9</b>	<b>63.8</b>	76.3	59.2	72.5	54.6
<i>H. sapiens</i>	954	44.7	<b>48.8</b>	33.5	35.6	44.7	<b>48.8</b>	<b>46.6</b>	47.7	44.4	47.2	35.7	36.3
<i>A. fulgens</i>	964	35.1	38.9	30.9	33.9	34.3	38.7	<b>37.4</b>	<b>39.8</b>	35.1	39.6	27.9	30.7
<i>S. acidocaldarius</i>	1494	52.4	51.9	52.4	51.9	58.3	58.0	<b>60</b>	<b>58.1</b>	58.3	58.0	57.9	56.2
Averages	722.4	56.77	56.58	49.46	49.04	54.56	51.99	<b>59.64</b>	<b>60.02</b>	58.28	57.32	55.28	52.52

```

Initialize the parameters of MACO
Randomly initialize the solutions for all the ants
While current number of iterations < Max iteration
  For each ant in the population
    For each stem
      Decide whether to select current stem according to (7)
    End for
    Evaluate the solution according to the energy.
  End for
  For each stem in the set
    Update the pheromones according to (8) and (9)
  End for
End while
    
```

ALGORITHM 1

```

Initialize all the parameters of MPSO
While current number of iterations < Max iteration
  For each particle
    Update its velocity
    Update its position
    Restrict position and velocity
    Calculate fitness and Update local best
  End for
  Update the global best
  Turn the parameters of MPSO via fuzzy logic controllers
End while
    
```

ALGORITHM 2

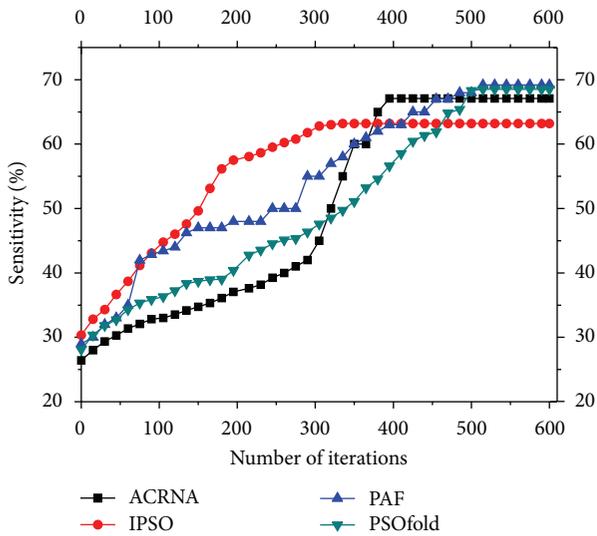


FIGURE 1: Average best sensitivity by running the algorithms 10 times on *A. griffini*.

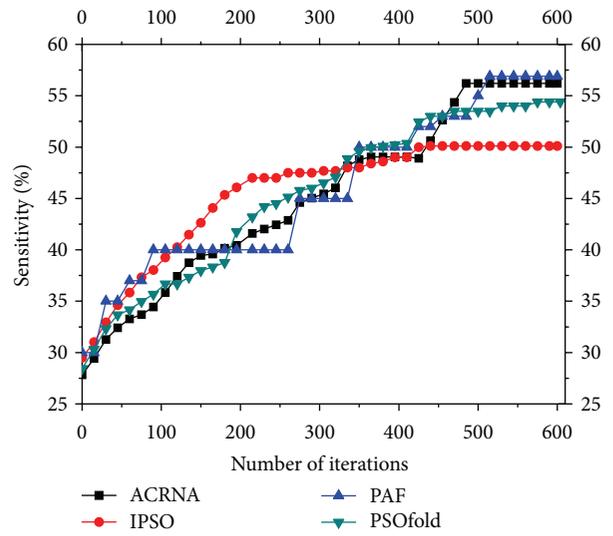


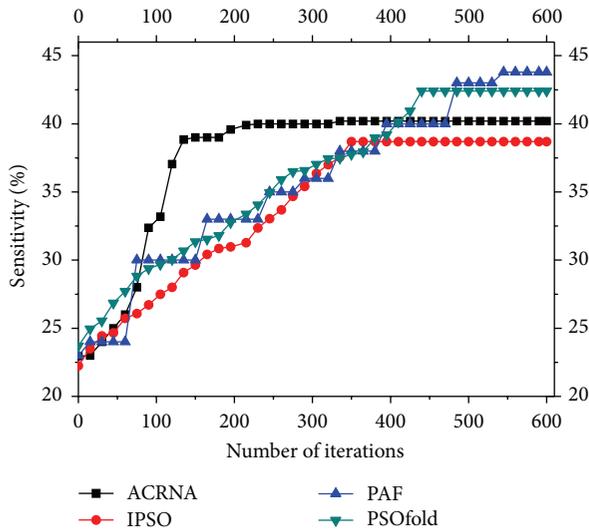
FIGURE 2: Average best sensitivity by running the algorithms 10 times on *H. rubra*.

predicted; TN (true negative) stands for the number of base pairs which were not matched and predicted correctly. The TN is rarely used in actual measurement because it is generally much larger than TP, FN, and FP. Sensitivity (Se)

means the percentage of all base pairs which was correctly predicted in the real structure; specific (Sp) refers to the percentage of all predicted base pairs which was correctly predicted. The general prediction is very difficult for both and

TABLE 4: Average sensitivity and specificity via ACRNA, IPSO, PAF, and PSOfold.

Sequence	Length	ACRNA		IPSO		PAF		PSOfold	
		se	sp	se	sp	se	sp	se	sp
<i>D. virilis</i>	784	25.1	24.5	27.8	27.1	<b>34.1</b>	<b>30.4</b>	32.6	29.9
<i>X. laevis</i>	945	40.2	38.8	38.7	36.7	<b>43.8</b>	<b>41.5</b>	42.4	40.1
<i>H. marismortui</i>	122	70.5	73.6	67.8	68.1	<b>70.6</b>	<b>76.0</b>	69.2	75.4
<i>A. lagunensis</i>	468	61.3	52.6	60.5	51.8	62.4	<b>53.8</b>	<b>63.9</b>	52.6
<i>M. anisopliae</i> var.	394	60.9	63.5	57.1	56.2	<b>66.5</b>	<b>64.1</b>	65.1	63.3
<i>H. rubra</i>	543	56.2	<b>48.8</b>	50.1	41.3	<b>56.9</b>	47.8	54.4	47.8
<i>A. griffini</i>	556	67.1	54.1	63.2	47.3	<b>69.2</b>	52.9	68.5	<b>54.5</b>
<i>H. sapiens</i>	954	34.1	32.1	31.9	33.5	<b>35.8</b>	<b>34.9</b>	32.4	33.3
<i>A. fulgens</i>	964	31.1	30.2	30.2	<b>32.5</b>	<b>34.2</b>	31.8	32.6	31.8
<i>S. acidocaldarius</i>	1494	56.0	55.4	52.5	51.4	54.5	<b>56.3</b>	<b>56.5</b>	55.9
Averages	722.4	50.25	47.36	47.98	44.59	<b>52.8</b>	<b>48.95</b>	51.76	48.46

FIGURE 3: Average best sensitivity by running the algorithms 10 times on *X. laevis*.

is always biased in favor of one side. A metric that combines both the specificity and sensitivity measures into one is  $F$ -measure; it can be used as a single performance measure for a predictor. The main result of the paper will be concluded by the sensitivity and specificity [18]. The specific formula is as follows:

$$\begin{aligned}
 Se &= \frac{TP}{TP + FN}, \\
 Sp &= \frac{TP}{TP + FP}, \\
 F &= \frac{2 \times \text{specificity} \times \text{sensitivity}}{\text{specificity} + \text{sensitivity}}.
 \end{aligned}
 \tag{11}$$

Ten sequences from the comparative RNA website are selected for evaluation of the proposed method, and the details of the sequence are described in Table 1. For these sequences the natural secondary structures are also available from the comparative RNA website. These sequences were chosen as they represent different sequence lengths and come from various genomes of organisms that are exposed to a range of physiological conditions. They represent four RNA classes: 5S rRNA, Group I intron 16S rRNA, 16S rRNA, and Group I intron 23S rRNA. Due to space constraints in some tables, we refer to these specific RNA sequences by an abbreviation of the name of the organism from which they originated [18].

Table 2 shows the comparative results of the highest matching base pair structures between PAF and mfold in regard to sensitivity, specificity, and  $F$ -measure. PAF predicts fewer base pairs on 8 sequences. PAF obtained higher TP in 7 cases and lower FP in all cases. The values FN predicted by PAF are also lower than mfold in 7 out of 10 cases. For sensitivity, specificity, and  $F$ -measure, PAF won in 7, 10, and 10 cases, respectively. Generally, PAF performs significantly better than mfold with respect to sensitivity, specificity, and  $F$ -measure.

Table 3 shows a more detailed analysis of comparing the highest matching base pair structures from SARNAPredict, RnaPredict, ACRNA, PAF, PSOfold, and mfold according to sensitivity and specificity. The results generated by SARNAPredict and RnaPredict were taken from the literature [8, 11]. From Table 3, it can be seen that PAF gets better sensitivity on 6 sequences and predicts higher specificity in 8 cases. SARNAPredict gets better results in 2 cases in terms of sensitivity and in 2 cases with regard to specificity. ACRNA wins on 1 sequence for sensitivity and on 1 sequence for specificity. PSOfold and mfold obtained better sensitivity on one sequence and two sequences, respectively. Among the six methods, PAF gets the best results in most cases. From another point of view, the average performance of PAF on sensitivity and specificity exceeds that of the other methods. It

is demonstrated that PAF is significantly superior to the other 5 methods.

In order to validate the stability of the proposed method, we ran ACRNA, IPSO, PAF, and PSOfold ten times and calculated the average highest matching base pair structures of each algorithm in terms of sensitivity and specificity. The detail results are shown in Table 4. From Table 4, it is easy to see that PAF obtains best sensitivity and specificity on 8 sequences and on 7 sequences, respectively. Thus, this proves that the proposed method performs stably on multiple sequences, considerably surpassing the other methods. The convergence progress is shown in Figures 1, 2, and 3. From the figures, we can clearly see that the proposed method could avoid trapping in the local optimum during the iterations. This is because the crucial stems lead the algorithm to the right direction.

## 5. Conclusion

In this paper, a framework, PAF, was proposed for RNA secondary structure prediction, which consists of CSS and GSB. In order to preserve crucial structures, MACO in CSS is proposed to find the important stems. MPSTO in GSB is developed to generate predicted structures in order to save searching spaces. The experimental results show that the performance of the proposed method is significantly better than those of the other metaheuristic methods in terms of sensitivity, specificity, and  $F$ -measure. We will try to enhance the performance of convergence and reduce time complexity in the future.

## Acknowledgment

This research was supported by The Chinese Government's Executive Program "Instrumentation development and field experimentation" (SinoProbe-09).

## References

- [1] D. H. Mathews and D. H. Turner, "Prediction of RNA secondary structure by free energy minimization," *Current Opinion in Structural Biology*, vol. 16, no. 3, pp. 270–278, 2006.
- [2] S. Ray and S. Pal, "RNA secondary structure prediction using soft computing," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 99, pp. 1–16, 2012.
- [3] E. Khosravani, G. Moradi, and S. Sajjadifar, "An accurate thermodynamic model to predict phase behavior of clathrate hydrates in the absence and presence of methanol based on the genetic algorithm," *Journal of Chemical Thermodynamics*, vol. 57, pp. 286–294, 2013.
- [4] Y. Liu, G. Wang, H. Chen, H. Dong, X. Zhu, and S. Wang, "An improved particle swarm optimization for feature selection," *Journal of Bionic Engineering*, vol. 8, no. 2, pp. 191–200, 2011.
- [5] Y. Li, G. Wang, and H. Chen, "An ant colony optimization based dimension reduction method for high-dimensional datasets," *Journal of Bionic Engineering*, vol. 10, no. 2, pp. 242–250, 2013.
- [6] D. F. Ouyang, "Investigating the molecular structures of solid dispersions by the simulated annealing method," *Chemical Physics Letters*, vol. 554, pp. 177–184, 2012.
- [7] B. A. Shapiro and J. C. Wu, "An annealing mutation operator in the genetic algorithms for RNA folding," *Computer Applications in the Biosciences*, vol. 12, no. 3, pp. 171–180, 1996.
- [8] 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, no. 4, pp. 727–740, 2010.
- [9] G. Benedetti and S. Morosetti, "A genetic algorithm to search for optimal and suboptimal RNA secondary structures," *Biophysical Chemistry*, vol. 55, no. 3, pp. 253–259, 1995.
- [10] B. A. Shapiro, D. Bengali, W. Kasprzak, and J. C. Wu, "RNA folding pathway functional intermediates: their prediction and analysis," *Journal of Molecular Biology*, vol. 312, no. 1, pp. 27–44, 2001.
- [11] K. C. Wiese, A. A. Deschenes, and A. G. 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, 2008.
- [12] K. C. Wiese and A. Hendriks, "Comparison of P-RnaPredict and mfold-algorithms for RNA secondary structure prediction," *Bioinformatics*, vol. 22, no. 8, pp. 934–942, 2006.
- [13] J. Yu, C. Zhang, Y. Liu, and X. Li, "Simulating the folding pathway of RNA secondary structure using the modified ant colony algorithm," *Journal of Bionic Engineering*, vol. 7, no. 4, pp. 382–389, 2010.
- [14] M. Geis and M. Middendorf, "A particle swarm optimizer for finding minimum free energy RNA secondary structures," in *Proceedings of the IEEE Swarm Intelligence Symposium (SIS '07)*, pp. 1–8, Honolulu, Hawaii, USA, April 2007.
- [15] M. Neethling and A. P. Engelbrecht, "Determining RNA secondary structure using set-based particle swarm optimization," in *Proceedings of the IEEE Congress on Evolutionary Computation (CEC '06)*, pp. 1670–1677, Vancouver, Canada, July 2006.
- [16] Y. Liu, H. Dong, H. Zhang, G. Wang, Z. Li, and H. Chen, "Prediction of RNA secondary structure based on particle swarm optimization," *Chemical Research in Chinese Universities*, vol. 27, no. 1, pp. 108–112, 2011.
- [17] C. Xing, G. Wang, Y. Wang, W. Shen, Y. Liang, and Z. Ji, "A novel method for RNA secondary structure prediction," in *Proceedings of the 7th International Conference on Natural Computation (ICNC '11)*, pp. 1136–1140, Shanghai, China, July 2011.
- [18] C. Xing, G. Wang, Y. Wang, Y. Zhou, K. Wang, and L. Fan, "PSOfold: a metaheuristic for RNA folding," *Journal of Computational Information Systems*, vol. 8, no. 2, pp. 915–923, 2012.
- [19] X. Li and M. Yin, "Application of differential evolution algorithm on self-potential data," *PLoS ONE*, vol. 7, no. 12, Article ID e51199, 2012.
- [20] X. Li and M. Yin, "An opposition-based differential evolution algorithm for permutation flow shop scheduling based on diversity measure," *Advances in Engineering Software*, vol. 55, pp. 10–31, 2013.
- [21] M. Dorigo, G. Di Caro, and L. M. Gambardella, "Ant algorithms for discrete optimization," *Artificial Life*, vol. 5, no. 2, pp. 137–172, 1999.
- [22] M. Dorigo, V. Maniezzo, and A. Coloni, "Ant system: optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 26, no. 1, pp. 29–41, 1996.

- [23] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1942–1948, December 1995.
- [24] R. Eberhart and J. Kennedy, "New optimizer using particle swarm theory," in *Proceedings of the 6th International Symposium on Micro Machine and Human Science*, pp. 39–43, October 1995.
- [25] J. Kennedy and R. C. Eberhart, "Discrete binary version of the particle swarm algorithm," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, pp. 4104–4108, October 1997.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

