

---

# Acceleration based Particle Swarm Optimization (APSO) for RNA Secondary Structure Prediction

Jitendra Agrawal and Shikha Agrawal

---

## 1 Introduction

A ribonucleic acid (RNA) molecule consists of a chain of ribonucleotides connected together by covalent chemical bonds. Each ribonucleotide consists of one of the four bases such as adenine (A), cytosine (C), guanine (G) or uracil (U), and the exact sequence of bases along the chain, the primary structure of the molecule, determines the kind of RNA it is. The secondary structure is the outcome of hydrogen bonds between nucleotides that are situated far away in the chain. Characteristically these hydrogen bonds exist only between G and C, or A and U, or G and U (or vice versa). The nucleotides so linked are called base pairs and are labeled as GC, CG, AU, UA, or GU, or UG, the first base being the one with the smaller index in the chain.

Though RNA secondary structure forecast is fundamentally based on Nuclear Magnetic Resonance and X-ray crystallography, these methods are very complex, time-consuming and highly costly. This has given birth to the development of various mathematical and computational models which have now become highly indispensable in bioinformatics for RNA secondary structure forecast. Several investigational efforts have been carried out in the area of RNA secondary structure prediction with optimization techniques. The cardinal issue here is to find out, from among all the probable secondary structures, the one which has the minimum energy, because it is the most likely one to have a robust stable secondary structure. But from the

computer perspective, it emerges into an optimization issue spread over a search space and has an objective function [11]. No wonder, a number of diverse optimization techniques and strategies have been introduced as an aftermath. With a view to eliminate this vexed dilemma, the contribution of diverse meta heuristics, intimately linked with the GA in the iterative search procedure like simulated annealing (SA), particle swarm optimization (PSO), ant colony optimization (ACO), and tabu search (TS), have also become subject matter of hot debate and discussion [21, 13]. Particle Swarm Optimization algorithm is employed to optimize the structure of an RNA molecule, by means of a sophisticated thermodynamic model [9]. Herbert et al. [22] have skillfully explained the advantages of SARNAPredict, an RNA secondary structure prediction algorithm based on Simulated Annealing (SA). Estimation for the execution of SARNAPredict in terms of forecast precision is established with native structures. Hao wu et al. [23] cleverly put forward a fuzzy adaptive particle swarm optimization (FPSO) blending particle swarm optimization (PSO) and fuzzy logic control (FLC) to forecast RNA secondary structure with the least energy. This technique is targeted at forecasting pseudoknots in the huge search spaces. The arithmetical outcomes and statistical investigations have proved beyond doubt that the innovative approach is competent to locate an optimal feature subset from a bulky noisy data set.

The forecast of RNA secondary structure with minimum free energy is a very critical function. In the perfect RNA structure prediction the calculation of base pairs plays a very effective role. With an eye on forecasting highly perfect RNA secondary structure several optimization techniques have been introduced in the literature. But it is unfortunate that each and every technique suffers from certain lacuna in the area of precise structure prediction. The drawbacks in the existing methods are solved by constructing a new optimization technique named as Acceleration base Particle Swarm Optimization (APSO) algorithm.

The outline structure of the paper is organized as follows:- In section 2, a brief discussion on the standard

---

J. Agrawal (✉)

School of Information Technology, UTD, Rajiv Gandhi Pradyogiki Vishwavidyalaya, Bhopal, M.P, India  
e-mail: [jitendra@rgtu.net](mailto:jitendra@rgtu.net)

S. Agrawal

Department of Computer Science & Engineering, UIT, Rajiv Gandhi Pradyogiki Vishwavidyalaya, Bhopal, M.P, India  
e-mail: [shikha@rgtu.net](mailto:shikha@rgtu.net)

PSO and the proposed APSO algorithm is presented. The experimental results on the multimodal benchmark functions and RNA structure prediction are given in Section 3 and conclusion of the paper is given in Section 4.

## 2 Proposed Acceleration based Particle Swarm Optimization (APSO)

### 2.1 Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) [26] is triggered by the community actions of organic creatures, like fishes and birds equipped with supreme qualities of clustering jointly to function hand in hand in recognizing advantageous locations in a particular zone, as evidenced by the fishes looking out for a food source.

PSO emulates the swarm behavior and individuals represent potential solutions in a D-dimensional search space. Particle  $i$  is often composed of four vectors:

$X_i = (x_i^1, x_i^2, \dots, x_i^D)$  with  $x_i^d$  being its position in the  $d^{\text{th}}$  dimension,  $pbest_i = (pbest_i^1, pbest_i^2, \dots, pbest_i^D)$  with  $pbest_i^d$  being the best position in the  $d^{\text{th}}$  dimension that particle  $i$  has found by itself,  $V_i = (v_i^1, v_i^2, \dots, v_i^D)$  with  $v_i^d$  being the velocity in the  $d^{\text{th}}$  dimension, and  $gbest = (gbest^1, gbest^2, \dots, gbest^D)$  with  $gbest^d$  being the global best position in the  $d^{\text{th}}$  dimension that all particles have found. Particles in a swarm move through the search space by

$$V_i^d = V_i^d + c_1.r_1.(pbest_i^d - x_i^d) + c_2.r_2.(gbest^d - x_i^d) \quad (1)$$

$$x_i^d = x_i^d + \delta V_i^d \quad (2)$$

Where  $c_1$  and  $c_2$  are two constants often with the value of 2.0, and  $r_1$  and  $r_2$  are two independent random numbers uniformly generated in the range [0, 1] at each updating iteration from  $d = 1$  to  $D$ , respectively.

The standard PSO technique has been extensively applied in various investigational efforts to achieve an optimal key for the dilemma. With a view to achieve a highly accurate optimal outcome, the defects inherent in the PSO technique such as premature convergence and loss of diversity have to be properly tackled and surmounted by initiating effective alteration or augmentation in the procedure of PSO. The cardinal defect of the PSO technique is a random value choice during the new particles generation procedure where during the course of the velocity calculation procedure the acceleration coefficients are engendered arbitrarily. The arbitrary value choice in the velocity procedure makes it essential that the created particles are in random. The unfortunate fact is that the random populations fail miserably in generating further precise outcomes. Therefore, to attain a

further precise outcome and to steer clear of the PSO defects, rather than fixing the value of acceleration coefficient we introduce an Acceleration base Particle swarm optimization (APSO) technique in which acceleration coefficient values are updated on the basis of evaluation function. Next section briefly explain the proposed APSO.

### 2.2 Acceleration based Particle Swarm Optimization (APSO)

In our proposed technique, an Acceleration based Particle Swarm Optimization (APSO) algorithm is developed to solve premature convergence problem of the standard PSO. In APSO, the acceleration coefficients are selected based on the fitness value which will increase the accuracy of the results. The selection of acceleration coefficient values in APSO algorithm is described as follows,

$$nc_1 = c_1 \times (1 - \lambda) \quad (3)$$

$$nc_2 = c_2 \times (1 - \lambda) \quad (4)$$

In Equ. (3) and (4), the  $\lambda$  value is computed based on the fitness values is calculated as,

$$\lambda = \frac{\chi \left( 1 + \phi (f_{\max} - f_{\min})^\omega - (f_{\text{avg}}^\omega) \right)}{\delta (f_{\max} - f_{\min})^\omega - f_{\text{avg}}^\omega} \quad (5)$$

$$\delta = \left( \frac{f_{\max} - f_{\min}}{f_{\text{avg}}} \right)^\omega \quad (6)$$

Where,

$\chi$  - Alteration probability

$\omega, \phi$  - are the coefficient factors

$F_{\max}, F_{\min}, F_{\text{avg}}$  - denotes the maximum, minimum and average fitness of the particles

By exploiting Equ. (3) and (4), the acceleration coefficients values, the velocity formula which is given in Equ. (1) is updated by,

$$V_i^d = V_i^d + nc_1.r_1.(a_1) + nc_2.r_2.(a_2) \quad (7)$$

In Equ. (7), the  $a_1$  and  $a_2$  values are computed by comparing the particles which have the minimum fitness values with the pbest particles values. If the both particles values are same the value of  $a_1$  and  $a_2$  is set as zero otherwise the values are 1.

In the APSO, the particles are generated by using the RNA structures base pairs. To generate the particles, different length RNA structures are stored in the database  $D$ . The process of APSO algorithm in RNA secondary structure prediction is described as follows,

**Step 1:** Initially, the particles are initiated by generate RNA sequence sets which contain the RNA base pairs (A, U, C, and G) in an encoded format. Then, the parameters of each particle, including its position and velocity are initiated.

**Step 2:** Every generated particles fitness value is calculated by the RNA fold algorithm. The particles which have the minimum free energy are selected as the best particles.

**Step 3:** Update  $pbest_i$  of each particle and  $gbest_i$ . Based on these values, the particles velocity and positions are updated by exploiting the Eqn. (7) and (2). The particles are updated by using the both Equ. (7) and thus we get the new particles. From this new set of particles best particles are selected by comparing the particles values and these best particles are given to the fitness computation.

**Step 4:** Stop if the current optimization solution is good enough or some stopping criterion is satisfied. Other-wise, go to Step 2.

### 3.1 Results of APSO on Benchmark Functions

The proposed APSO method is first tested on eight benchmark functions which are described as follows,

To accomplish the performance analysis process, we have performed 30 independent runs of APSO, GA and PSO methods. Result of convergence in terms of average fitness value for number of iterations, of all the three algorithms, for all the above mentioned functions is shown in Figure 1.

As can be seen from Figure. 1, our proposed APSO method has obtained global optima in less number of iterations as compared to PSO and GA methods. Among three optimization methods, GA has given poorer performance when compared to PSO and APSO methods whereas PSO performance degrades when compared to APSO. Hence, our proposed APSO method has better convergence rate for all the eight standard functions than the PSO and GA methods.

## 3 Experimental Results

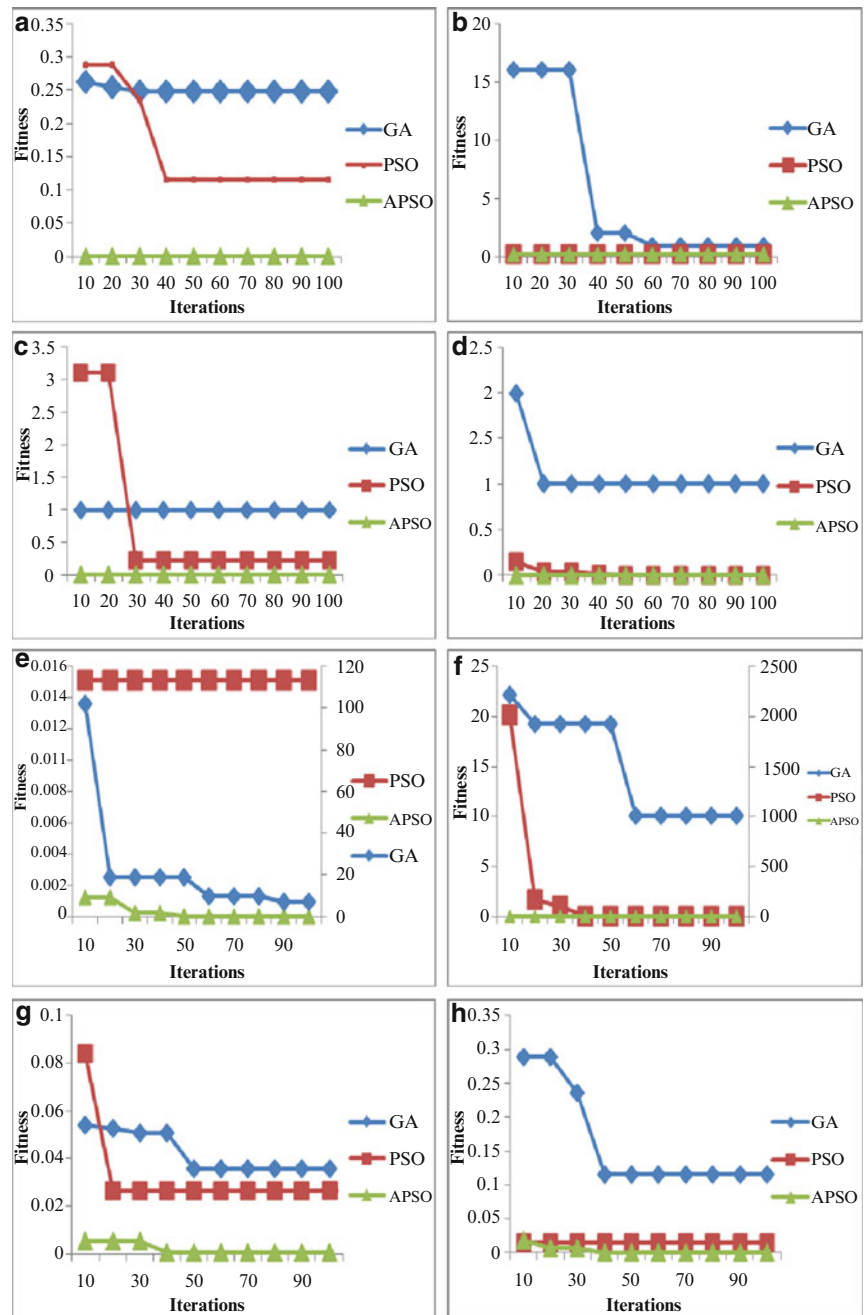
The proposed Acceleration based PSO (APSO) is implemented in the working platform of MATLAB and results are compared with GA and standard PSO. The parameters which are utilized in the APSO are listed in Table 1.

**Table 1** Parameters Values in APSO

Parameters	Values
Population Size	20
$r_1$ and $r_2$	[0,1]
$c_1$ and $c_2$	0.02
$nc_1$ and $nc_2$	0.04

Name	Equation
Extended Powell	$f_0(x) = \sum_{i=1}^{n/4} (x_{4i-3} + 10x_{4i-2})^2 + 5(x_{4i-1} - x_{4i})^2 + (x_{4i-2} - 2x_{4i-1})^4 + 10(x_{4i-3} - x_{4i})^4$
Dixon	$f_1(x) = (1 - x_1)^2 + (1 - x_{10})^2 + \sum_{i=1}^9 (x_i - x_{i+1})^2$
Sphere	$f_2(x) = \sum_{i=1}^n x_i^2$
Rastrigin	$f_3(x) = \sum_{i=1}^n (x_i^2 - 10 \cos(2\pi x_i) + 10)$
Rosenbrock	$f_5(x) = \sum_{i=1}^{n/2} 100(x_{2i} - x_{2i-1}^2)^2 + (1 - x_{2i-1})^2$
Schwefel	$f_6(x) = -\sum_{i=1}^n x_i \sin(\sqrt{ x_i })$
Griewank	$f_7(x) = \sum_{i=1}^n \frac{x_i^2}{4000} - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$
Bohachevsky	$f_8(x) = \sum_{i=1}^n (x_i^2 + 2x_{i+1}^2 - 0.3 \cos(3\pi x_i) - 0.4 \cos(4\pi x_{i+1}) + 0.7)$

**Fig. 1** Performance of APSO, PSO and GA methods



### 3.2 Results of APSO on RNA Secondary Structure Prediction

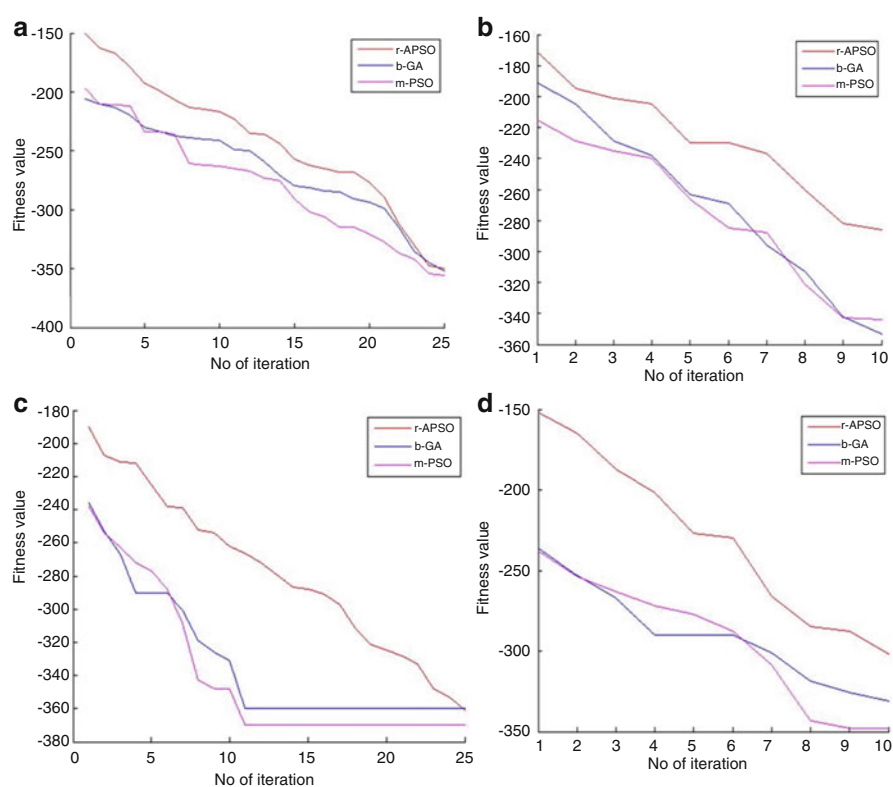
The proposed APSO, PSO & GA methods' performances are analyzed by conducting experiments on four different RNA sequences of varying sizes. The results of the 4 different experiments with 4 RNA sequences are given in Table 2.

Table 2 shows the performance of APSO, PSO & GA methods' performance in the RNA pair's prediction. In *S. cerevisiae* sequence 37 pairs are predicted in total, among which 31 pairs are correctly predicted with standard

deviation 4.9 in APSO and in PSO & GA methods 29 & 28 pairs are correctly predicted respectively. In *H.marismortui* RNA sequence, APSO correctly predicts 201 base pairs among the 233 predicted base pairs and GA & PSO methods have correctly predicted 192 and 190 base pairs respectively. In other two RNA sequences *X.laevis* and *D.virilis* our proposed APSO correctly predicts 222 base pairs in 254 predicted base pairs and 30 base pairs in 38 predicted base pairs. When compared to the GA and PSO methods our proposed APSO method has given superior base pair prediction accuracy.

**Table 2** Performance of (i) GA (ii) PSO (iii) APSO results on 4 RNA sequences

Sequences	Free Energy ( $\Delta G$ kcal/mol)	Pairs Predicted	Pairs Correct	Pairs Known Structure	Percentage correct (%)
<b>GA</b>					
<i>S.cerevisiae</i>	-43.4	$37 \pm 0$	$29 \pm 4.2$	37	78.3
<i>H.marismortui</i>	-53.6	$233 \pm 0$	$192 \pm 10.5$	233	82.4
<i>X.laevis</i>	-205.9	$254 \pm 0$	$212 \pm 11.6$	254	83.6
<i>D.virilis</i>	-101.8	$38 \pm 0$	$27 \pm 3.9$	38	71.05
<b>PSO</b>					
<i>S.cerevisiae</i>	-44.4	$37 \pm 0$	$28 \pm 4.4$	37	75.6
<i>H.marismortui</i>	-52.5	$233 \pm 0$	$190 \pm 10.3$	233	81.5
<i>X.laevis</i>	-201	$254 \pm 0$	$215 \pm 11.4$	254	84.6
<i>D.virilis</i>	-103.5	$38 \pm 0$	$26 \pm 3.8$	38	68.3
<b>APSO</b>					
<i>S.cerevisiae</i>	-40.4	$37 \pm 0$	$31 \pm 4.9$	37	83.7
<i>H.marismortui</i>	-45.3	$233 \pm 0$	$201 \pm 11.6$	233	86.2
<i>X.laevis</i>	-193	$254 \pm 0$	$222 \pm 12.8$	254	87.4
<i>D.virilis</i>	-97.8	$38 \pm 0$	$30 \pm 4.4$	38	78.9

**Figure 2** Performance of APSO, GA, PSO techniques in terms of free energy in RNA sequences

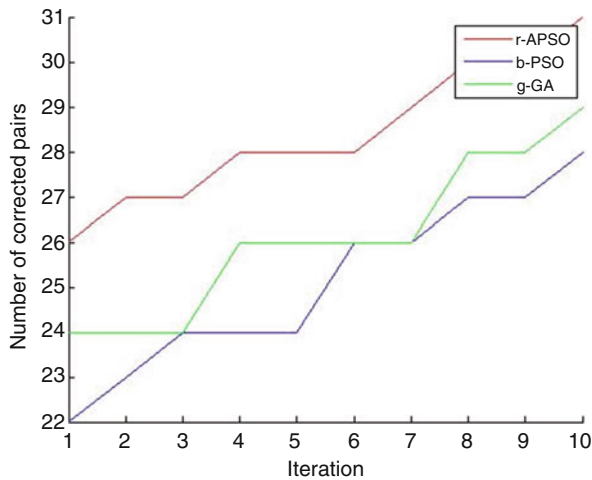
Performance of APSO is further evaluated on the basis of convergence characteristic on the 4 RNA sequences and is presented in Figure 2. APSO shows superior performance than GA and PSO in all the cases.

Prediction accuracy is shown in Figure 3. As can be seen from Fig. 4, our proposed APSO method has given a greater number of correctly predicted pairs than the GA and PSO methods in all 10 iterations. While comparing GA & PSO methods, GA has given superior prediction accuracy but this is lower than that of the APSO method.

Hence, our proposed APSO method has given high performance in RNA structure prediction than the GA and PSO methods.

## 4 Conclusion

In this paper we have proposed a new Acceleration base Particle Swarm Optimization (APSO) algorithm for finding minimum energy RNA secondary structures. The proposed



**Figure 3** Comparison graph of our proposed APSO and GA, PSO methods performance in terms correctly predicted base pairs of sequence X.laevis

APSO is first tested on eight benchmark functions and then the algorithm is extended to solve RNA secondary structure prediction problems. The experimental result on benchmark functions shows that APSO acquires a better performance than the GA and standard PSO algorithms. Simulation to predict RNA secondary structure is performed on 4 different RNA sequences of varying lengths. Computational result shows that APSO algorithm not only converges well but also has high prediction accuracy as compared to GA and standard PAO algorithms. Hence, our proposed APSO algorithm is able to predict more corrected structures and also find better secondary structures in terms of free energy.

## Reference

1. Hao Wu, Yan-feng Shi, Xing Jin, Gang Wang and Hao Dong, "A Fuzzy Adaptive Particle Swarm Optimization for RNA Secondary Structure Prediction", In Proceedings of International Conference on Information Science and Technology, Nanjing, pp. 1390-1393, 2011.
2. Arturo Díaz Perez Mario A. Garcia-Martinez, "FPGA Accelerator for RNA Secondary Structure Prediction", In Proceedings of 12th Euromicro Conference on Digital System Design / Architectures, Methods and Tools, Patras, pp. 667-671, 2009.
3. Eberhart R.C. and Kennedy J., "Particle swarm optimization," IEEE International Conference on Neural Networks, Volume 4(27), pp 1942-1948, 1995.
4. Esquivel S., Coello C. and Coello A., "On the use of particle swarm optimization with multimodal functions", IEEE Congress on Evolutionary Computation (CEC), pp1130-1136, 2003.
5. Herbert H. Tsang and Kay 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. 727-738, 2010.
6. Kangtai Wang and Ning Wang, "A novel RNA Genetic Algorithm for Parameter Estimation of Dynamic Systems", Journal of Chemical Engineering Research and Design, Vol. 88, No. 11, pp. 1485-1493, 2010.
7. Marais Neethling and A.P. Engelbrecht, "Determining RNA Secondary Structure using Set based Particle Swarm Optimization", In Proceedings of IEEE Congress on Evolutionary Computation, Vancouver, pp. 1670-1677, 2006.
8. Poli R., "An Analysis of the publications on the applications of Particle Swarm Optimization", Journal of Artificial Evolution & Applications, 2007.
9. S.S. Ray, M. Bachhar, and S.K. Pal, "RNA Secondary Structure Prediction in Soft Computing Framework: A Review," Proc. IEEE Third Int'l Conf. Computer Science and Information Technology, vol. 5, pp. 430-435, 2010.
10. Shubhra Sankar Ray and Sankar K. Pal, "RNA Secondary Structure Prediction Using Soft Computing", IEEE/ACM Transactions on Computational Biology and Bioinformatics, Vol. 10, No. 1, pp. 2-17, 2013.