

# Population-Based Ant Colony Optimization with New Hierarchical Pheromone Updating Mechanism for DNA Sequence Design Problem

MISINEM<sup>1</sup>, ERMATITA<sup>2\*</sup>, Dian Palupi RINI<sup>2</sup>, Reza Firsandaya MALIK<sup>2</sup>  
and Tri Basuki KURNIAWAN<sup>1</sup>

<sup>1</sup>Faculty of Computer Science, Universitas Bina Darma, Palembang, Indonesia

<sup>2</sup>Faculty of Computer Science, Universitas Sriwijaya, Palembang, Indonesia

\*Corresponding author: ermatita@unsri.ac.id

## Abstract

DNA computing research in the last 20 years has attracted the attention of many researchers from various scientific disciplines. Many researchers have succeeded in demonstrating the ability of bio-molecules DNA computing to be used as storage media and biochemical tools as information processing operators. The reaction process between chemical compounds in DNA is probabilistic, which means that one experiment with others can get different results, even in the same conditions. To overcome this problem, the design of the DNA sequence should be received much attention. The DNA sequence design is a process to get a set of DNA sequences that will be unique to each other. The method of DNA sequence formation can be part of an optimization process, in which each DNA sequence must be designed in such a way as not to react with other DNA sequences in the same set. This process can increase the success of a DNA computing experiment. In this paper, an enhanced method from previous research is proposed. Instead of using Ant Colony Optimization (ACO) algorithm, the population-based version of ACO (P-ACO) is used, to take advantage of the population-like optimization algorithm to handle the multi-objective optimization problem. A new hierarchical pheromone updating mechanism is proposed as well to obtain a highly accurate and efficient tool to feed the information for the ants to follows the best path. In this paper, we use the same model for the ants to solve the problem, the Watson-Crick  $\Delta G^{o37}$  thermodynamic temperature parameter pairs used as distance among nodes in the model. The result from the proposed method and previous research compared then it shows the proposed method get the relatively better result in terms of the pattern of DNA sequence and the objectives function.

**Keywords:** population-based ant colony optimization, DNA sequence design, new pheromone

## Introduction

Although DNA computing is the research area of biomolecular science, other fields of science have also been passionately following developments in this field, included in the soft computing research is which has demonstrated very advanced technology using artificial intelligence algorithms.

The main objective of DNA sequence design is to prevent mismatch hybridization among sequences in the data set. Avoidance of mismatch hybridization ensures that the generated DNA sequences are unique and cannot be hybridized with other sequences. Previous studies have proposed a variety of DNA sequence design approaches [1-2] and applications of DNA sequence design can be seen in several areas [3-11].

In [12], the Ant Colony System (ACS) as a derivative from Ant Colony Optimization (ACO) which uses four nodes model to represent four DNA bases was proposed as shown in Figure 1. The ACS uses the nearest-neighbor thermodynamic parameter's Watson-Crick base pair  $\Delta G^{o37}$  as the distance between nodes. Although the results are shown relatively better performance to other approaches, the obtained DNA sequences tend to be

repeating only in some DNA bases as shown in Figure 2. This pattern makes a generated DNA Sequence difficult to fulfill the constraints in DNA Sequence Design Problem. Since the solution is considered as a set of DNA Sequences if there is one DNA sequence in

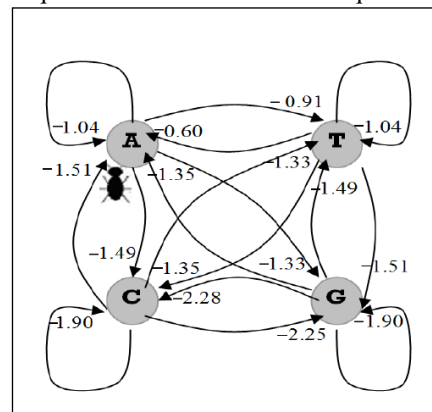


Figure 1. ACS modeling for DNA sequence design problem with thermodynamic value [12]

Sequence	C	HA	HM	S
Ant Colony System				
GC GAAGATGGTCTATATCGG	0	0	102	71
ATCCTCTCTCTAACGCAAG	0	0	82	86
CTCTCTCTCTTAATCCACA	0	0	69	91
GCATAACTCTCCGGACCTAT	0	0	85	80
GGCTTGCTATCGATCTGATG	0	0	95	78
TTCGACTCTGAATACGCC	0	0	98	79
ACTACACCTCTCTTATGCC	0	0	80	88

Figure 2. The obtained result from the previous approach [12]

that set fails to fulfill the constraint, the set of DNA Sequences cannot be accepted as a candidate of solution. To overcome this problem, this paper proposed a new pheromone updating strategy.

This paper is organized as follows. In section 2, the DNA sequence design problem is defined and the proposed approach is discussed in detail, Section 3 the sequence generation results are shown and compared with those of other existing methods and analysis is discussed. In Section 4, the conclusion is drawn.

## RESEARCH METHODOLOGY

### DNA Sequence Design Problem

In DNA computing, hybridization between a DNA sequence and its base-pairing complement is the most important factor to retrieve the information stored in DNA sequences and operate the computation processes. For this reason, the desired set of good DNA sequences, which have a stable duplex with their complement are more needed. It is also important to ensure that two sequences are not complemented to each other. Non-interacting sequences should be prohibitive or relatively unstable, compared with any perfectly matched duplex formed from a DNA sequence and its complement [13].

The objective of the DNA sequence optimization problem is basically to generate a set of DNA sequences, where each sequence is must unique or cannot be hybridized with other sequences in their set. In this paper, the four objective functions and two constraints from [12] are used. Two of the objective functions, namely *h-measure* and *similarity*, are chosen to estimate the uniqueness of each DNA sequence. Moreover, two other objective functions, which are *hairpin* and *continuity*, are used in order to prevent the secondary structure of a DNA sequence. Furthermore, two constraints, which are *GC-content* and melting temperature (*Tm*), are used to keep uniform chemical characteristics.

DNA sequence optimization is actually a multi-objective optimization problem. However, the problem is converted into a single-objective problem, which can be formulated as follows:

$$\min f_{DNA} = \sum_i w_i f_i \quad (1)$$

where  $f_i$  is the objective function for each  $i \in \{h\text{-measure}, similarity, hairpin, continuity\}$  and  $w_i$  is the weight for each  $f_i$ . In this study, weights are set to one.

### Population-based Ant Colony Optimization

The Population-based Ant Colony Optimization (P-ACO) algorithm was proposed by Guntch and Middendorf [14] based on the ACS algorithm, where a population of solutions is directly transferred to the next iteration. In order to construct solutions, similar to Ant Colony System (ACS), P-ACO uses Eq. (2) for state transition rule.

$$P_k(r, s)_{ACS} = \begin{cases} arg \max_{u \in J_k(r)} \{ [\tau(r, u)] \cdot [\eta(r, u)]^\beta \} & \text{if } q \leq q_0 \\ S & \text{otherwise} \end{cases} \quad (2)$$

These solutions are then used to compute the pheromone information for the ants in the next iteration. In Eq. (2),  $q$  is a uniformly distributed random variable  $[0, 1]$ ,  $q_0$  is between 0 and 1, and  $S$  is a variable which is selected according to the probability distribution as defined in Eq. (3).

$$P_k(r, s)_{ACS} = \begin{cases} \frac{[\tau(r, s)]^\alpha \cdot [\eta(r, s)]^\beta}{\sum_{u \in J_k(r)} [\tau(r, s)]^\alpha \cdot [\eta(r, s)]^\beta} & \text{if } s \in J_k(r) \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where  $J_k(r)$  is the set of feasible nodes; that is, edges  $(r, s)$  where  $r$  is the current city and  $s$  is the next city which is not visited by the  $k^{th}$  ant.  $(r, u)$  is the other edges, where  $u$  is all cities not yet visited by the  $k^{th}$  ant. The parameters  $\alpha$  ( $\alpha \geq 0$ ) and  $\beta$  ( $\beta \geq 0$ ), control the relative importance of the pheromone versus the heuristic information  $\eta(r, u)$ , which is given by Eq. (4):

$$\eta(r, u) = \frac{1}{d_{rs}} \quad (4)$$

In TSP, for example,  $d_{rs}$  is the distance between cities  $r$  and  $s$ , and  $\tau(r, s)$  is pheromone trails, which refer to desirability of visiting city  $s$  directly after  $r$ . For implementation purposes, the pheromone trails are collected into a pheromone matrix whose elements are all of  $\tau(r, s)$ .

The P-ACO algorithm also performs a pheromone updating to the pheromone matrix but uses a different mechanism. The best solution at every iteration is chosen to be stored into the population, and then the positive pheromone updating rule, as shown in Eq. (5) is applied.

$$\forall_i \in [1, n]: \tau_{i\pi(i)} = \tau_{i\pi(i)} + \Delta \quad (5)$$

Next, the negative pheromone updating rule, as shown in Eq. (6) is performed when a solution leaves the population.

$$\forall_i \in [1, n]: \tau_{i\sigma(i)} = \tau_{i\sigma(i)} - \Delta \quad (6)$$

where  $\Delta$  is the amount of pheromone for the positive and negative updating process.

$$\Delta = \frac{(\tau_{max} - \tau_0)}{m} \quad (7)$$

where  $m$  is the number of solutions in a population,  $\tau_{max}$  and  $\tau_0$  are the maximum and initial pheromones, respectively.  $\tau_{max}$  and  $\tau_0$  values are defined by the user, where  $\tau_0 < \tau_{max}$ .

### New Hierarchical Updating Pheromone in Population-based Ant Colony Optimization

The finite state machine as a model for constructing a DNA sequence was proposed by [12] as shown in Figure 3. Since there are only four nodes in the state to represent a node in DNA, the tour will visit a node repeatedly. Furthermore, the pheromone in some path will become higher than others as shown in Figure 4. This will cause the tour to tend to choose these paths and the DNA sequence obtained from

this tour will have pattern repeating in some DNA bases only as shown in Figure 2 and the DNA sequences will difficult to satisfy the constraints *GC-content* and *Tm*.

In order to overcome these problems, a hierarchical pheromone updating technique is proposed as shown in Figure 5. Instead of having a single set node of DNA for a single run, this model has a set of the node of DNA for each level of the construction process. For every level state transition in the P-ACO mechanism, there is a set of DNA paths that can be chosen based on their own pheromone and not choose node at the same level.

Figure 5 illustrates that in order to generate a 20-mer of DNA Sequence. For that, there are 20 levels of pheromone matrix, which each pheromone matrix' level for each level of DNA in the sequence. For each level, there are four nodes, where each node has four different paths that link to the other node in the next level. Each

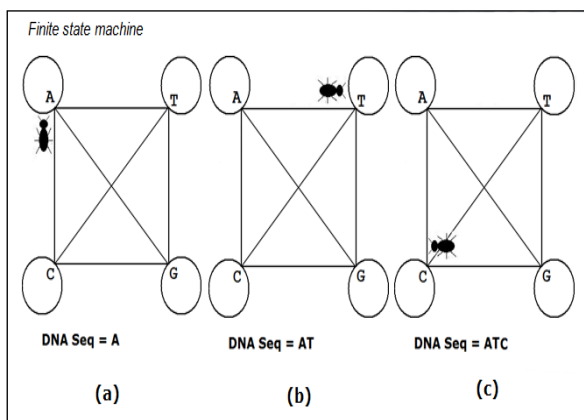


Figure 3. Finite state machine as a model for constructing a DNA Sequence [12]

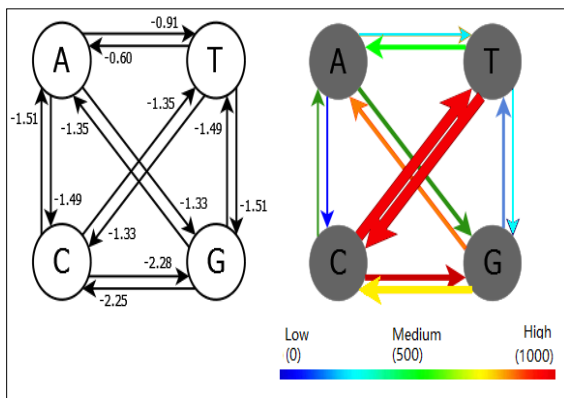


Figure 4. The paths of the ACS model for DNA Sequence Problem and their pheromone for each path [12].

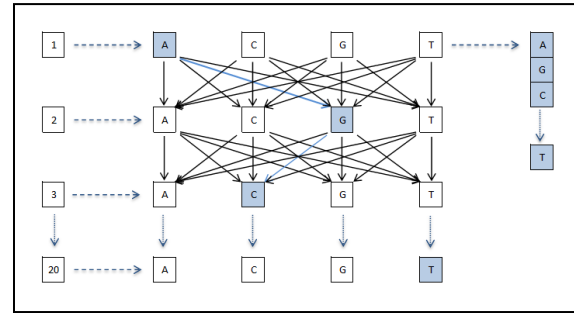


Figure 5. The Proposed Model, Hierarchical Updating Pheromone in P-ACO for DNA Sequence Design Problem

Node at the same level is not connected by any paths for interconnection. Therefore, each node at the same level cannot be traversal to another node. This mechanism ensures that the pheromone will be updated with different paths for each level of sequence.

One pheromone matrix will be built for one population to obtain the best performance for only one objective. That means there is four pheromone matrix for four objectives, such as *h-measure*, *similarity*, *continuity*, and *hairpin*. Each pheromone matrix has a mechanism to share their knowledge with another pheromone matrix, which used the best so far result from other populations.

## RESULTS AND DISCUSSION

The sequences obtained from the proposed approach are compared with the sequences generated by different methods. For comparison purposes, each of the compared methods runs 300 times and the results obtained are reported based on the same parameters reported in [12].

The comparison results between our results obtained from the proposed approach with the result obtained from previous work [12] are shown in Table 1 and Figure 6. Since this optimization process is about finding the minimum values for the objective function, the smallest value will be the best result. On the other hand, since the multi-objective problem is converted into a single objective problem, the overall results reported in this paper are by the total objective values. Table 1 and Figure 6 show that the new proposed approach obtained a much lower total objective values than the previous work [12]. The newly proposed method acquires lower *similarity* and *h-measure* objective and the pattern shows that there are not repeating nodes in the solutions.

Further experiment has been conducted by collecting DNA sequences that were not able to satisfy the constraints. The collected sequences from two compared methods are shown in Table 2. The experiments run for 300 iterations and for each iteration generates 7 sequences. Thus, 21,000 DNA sequences were generated for each method. Table 2 tabulates the results obtained from the experiments show that the proposed method has a smaller value (6.90%) for not satisfying the DNA Sequence constraints compared to previous works (50.76%). The results show that the new method has a better opportunity to get a better solution.

Table 1. Comparison Results of Proposed Approach and result of previous work [12].

Sequence	<i>continuity</i>	<i>hairpin</i>	<i>h-measure</i>	<i>similarity</i>
Proposed approach / New Hierarchical pheromone updating				
ACTAGCGACCGGATGATTAC	0	0	80	64
TCTACTACTCTACTCCGGCA	0	0	80	66
CGTAGCACCTTGCAATTGATC	0	0	81	67
CGACGTACGAGTATGATCTC	0	0	82	64
GATGCGATAGAGATCTACGC	0	0	88	68
CCGACTGAATTGACGACTGA	0	0	83	66
GTGGAGAACAGGCTTATGAG	0	0	86	66
Total	0	0	580	461
Average	0	0	82.86	65.86
Previous work [12]				
GCGAAGATGGTCTATATCGG	0	0	102	71
ATCCTCTCTCCTAACGCAAG	0	0	82	86
CTCTCTCCTCCTAATCCACA	0	0	69	91
GCATAACTCTCCGGACCTAT	0	0	85	80
GGCTTGCTATCGATCTGATG	0	0	95	78
TTCGACTCTTGAATACGCGC	0	0	98	79
ACTACACCTCCTCTTATGCC	0	0	80	88
Total	0	0	611	573
Average	0	0	87.29	81.86

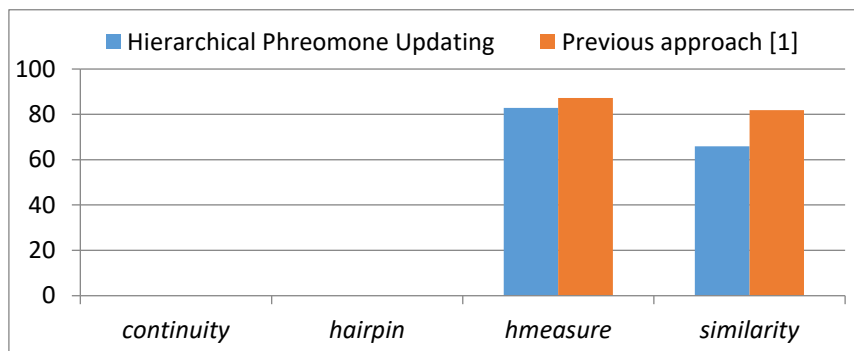


Figure 6. Average fitness comparison results between hierarchical pheromone updating and previous work [12].

Table 2. Comparison Number of failed DNA Sequences Generation

Approach	# Satisfy	# Not Satisfy	Total (300 x 7)
New Hierarchical Pheromone Updating	1955 (93.10%)	145 (6.90%)	2100
Previous Work [12]	1034 (49.24%)	1066 (50.75%)	2100

## CONCLUSION

A new method called Hierarchical Pheromone Updating Mechanism for P-ACO in Designing DNA is proposed and implemented with four objective functions: h-measure, similarity, continuity, and hairpin and two constraints: GC-content and Tm. The DNA sequences obtained from the proposed method were compared with the results reported in [12]. The results show that a new method can generate

better DNA sequences where there is no repeated pattern of nodes and get more chances to satisfy the constraints in order to obtain better results in terms of h-measure and similarity objective.

## ACKNOWLEDGMENT

This research was financially supported by Universitas Bina Darma, Palembang, Indonesia.

**REFERENCES**

- [1] J Xiao, X Zhang, and J Xu, A membrane evolutionary algorithm for DNA sequence design in DNA computing, *Chinese Science Bulletin*, Vol. 57 (6), pp.698-706, 2012.
- [2] M Caserta and S VoB, A hybrid algorithm for the DNA sequencing problem, *Discrete Applied Mathematics*, Vol. 163, Part I, pp. 87-99, 2014.
- [3] J Ren and Y Yao, DNA Computing Sequence Design Based on Bacterial Foraging Algorithm, *Proceedings of ICBRA, International Conference on Bioinformatics Research and Application*, pp. 1-7, 2018.
- [4] J.M C Gonzales and J M Gil, An efficient design for a multi-objective Evolutionary Algorithm to Generate DNA Libraries suitable for Computation, *Interdisciplinary Sciences: Computational Life Sciences*, Vol. 11 (3), pp. 542-558, 2019.
- [5] G Yang, B Wang, X Zheng, C Zhou and Q Zhang, IWO Algorithm Based on Niche Crowding for DNA Sequence Design, *Interdisciplinary Sciences: Computational Life Sciences*, Vol 9 (3), pp.341-349, 2017.
- [6] LJ Cheng and SJ Wei, Research progress in the construction of DNA Logic self-assembly, *Journal of Light Industry* Vol 6, 2016.
- [7] R Lopez, Y Jyue Chen, SD Ang, S Yekhanin, K Makarychev, MZ Racz, G Seelig, K Strauss, and L Ceze, DNA assembly for nanopore data storage readout, *Nature: Communications* Vol 10 (2933), 2019.
- [8] D Woods, D Doty, C Myhrvold, J Hui, F Zhou, P Yin, and E Winfree, Diverse and robust molecular algorithms using reprogrammable DNA Self Assembly, *Nature: International Journal Science*, Vol 567, pp. 366-372, 2019.
- [9] L Yuanyuan, Z Xuedong, W Bin, Z Shihua, Z Changjun, The Optimization of DNA Encoding Based on Chaotic Optimization Particle Swarm Algorithm, *Journal of Computational and Theoretical Nanoscience*, Vol 13 (1) pp.443-449, 2016.
- [10] X Peng, X Zheng, B Wang, C Zhou and Q Zhang, A micro-genetic algorithm for DNA encoding sequences design, *International Conference on Control Science and System Engineering (ICCSSE)*, 2016.
- [11] B Yimwadsana, P Artiwet, On Optimizing DNA Sequence Design for DNA Logic AND Circuit. *Tencon IEEE Region 10 Conference* 2018.
- [12] Kurniawan, T.B., Khalid, N.K., Ibrahim, Z., Khalid, M., Middendorf, M.: An Ant Colony System for DNA Sequence Design Based On Thermodynamics. In: *Proceedings of the Fourth IASTED International Conference Advances in Computer Science and Technology (ACST 2008)*, Langkawi, Malaysia, pp. 144–149, 2008.
- [13] Brenneman, A. Condon, Strand design for biomolecular computation, *Theory: Comput. Sci*, Vol. 287, 2002.
- [14] M. Guntsch and M. Middendorf: Applying Population-based ACO to Dynamic Optimization Problems, In *Proc Ant Algorithms, Third International Workshop, ANTS 2002*. Pp. 111-122 (2002)