The 4th ICIBA 2015, International Conference on
Information Technology and Engineering Application
Palembang-Indonesia, 20-21 February 2015

New Pheromone Updating Mechanism for Ant Colony Optimization in Designing DNA Sequence

T.B.Kurniawan, Misinem, M.Z. Ahmad Nazri

Universitas Bina Darma
e-mail: tribasukikurniawan@yahoo.com
Universitas Bina Darma
e-email: misinem_kurniawan@yahoo.com
UnivesitiKebangsaan Malaysia
e-mail: zakree@ukm.edu.my

Abstract

DNA computing is relatively a new computing paradigm that uses bio-molecular as information storage media and biochemical tools as an information processing operators. Hybridization between a DNA sequence and its base-pairing complement is crucial to retrieve the information stored in DNA sequence and operate a computation operation in DNA computing. Since DNA reactions are probabilistic reactions, it can cause different results for the same situation, which can be regarded as error in computation. To overcome the drawbacks, this paper proposed an Ant Colony System (ACS) thatuses four nodes model to represent four DNA bases. In this model, the nearest-neighbor thermodynamic parameters Watson-Crick base pair $\Delta G\circ_{37}$ is used as the distance between nodes. Although the results are comparable to other approaches, the obtained DNA sequences tend to repeat the same pattern in some DNA bases. In this paper,hierarchical pheromone updating method for ACS is proposed. The obtained results from new approach were compared with the previous approaches and show a relative better pattern results.

Keywords : *Ant Colony Optimization (ACO), Ant Colony System (ACS), DNA Sequence Design, Hierarchical Pheromone Updating Strategy, and Thermodynamic*

1 INTRODUCTION

The main objective of DNA sequence design is to prevent mismatch hybridization amongsequences in the data set. Avoidance of mismatch hybridisation ensures that the generated DNA sequences are unique and cannot be hybridised with other sequences. Previousstudies have proposed a variety of DNA sequence design approaches (Brenneman and Condon, 2002; Shin et al., 2005) and applicationsof DNA sequence design can be seen in several areas (Hertemink et al, 1998; Ponchovsky and Ackermann, 2003; Frutos et al., 1997; Feldkamp et al., 2001; Tanaka et al., 2001; Marathe et al., 1999; Deaton et al., 2002; Zhang and Shin, 1998).

In Kurniawan et al. (2008), an 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 parameters Watson-Crick base pair $\Delta G\circ_{37}$ as the distance between nodes. Although the results are comparable 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 consider as a set of DNA Sequences, if there are one of DNA sequence in that set fail to fulfill the constraint, the set of DNA Sequences cannot be accept as a candidate of solution. In order 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 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, conclusion is drawn.

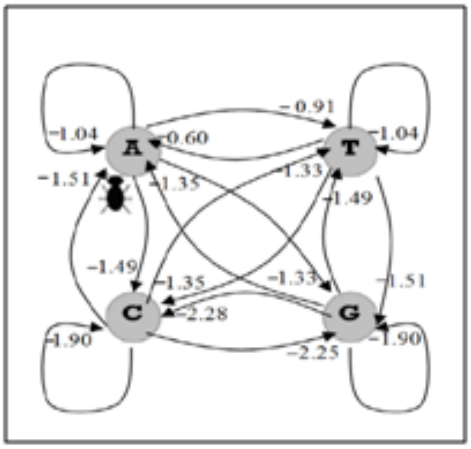

Figure 1: ACS modelling for DNA sequence design problem with thermodynamic value(Kurniawan et al., 2008)

Sequence	C	HA	HM	S
Ant Colony System				
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

Figure 2: The obtained result from the previous approach (Kurniawan et al., 2008)

2 RESEARCH METHODOLOGY

2.1 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 stableduplex with their complement are more needed. It is also important to ensure that two sequences are not complement 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 (Brenneman and Condon, 2002).

The objective of the DNA sequence optimizationproblem is basically to obtain a set of DNA sequences,where each sequence is unique or cannot be hybridizedwith other sequences in the set. In this paper, theobjective functions and constraints from (Kurniawan et al., 2008) are used.Two objective functions, namely hmeasure and similarity,are chosen to estimate the uniqueness of each DNAsequence. Moreover, two additional objective functions,which are hairpin and continuity, are used in order toprevent secondary structure of a DNA sequence.Furthermore, two constraints, which are GCcontentandmelting temperature, are used to keep uniform chemicalcharacteristics.

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

$$min\ f\ DNA = \Sigma_i w_i f_i \tag{1}$$

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

2.2 Ant Colony System

Ant colonyoptimization (ACO) is a population-based meta-heuristic for combinatorial optimization problems. The ACO was inspired by the ability of ants to find the shortest path between their nest and a source of food. Marco Dorigo introduced ACO in his PhD thesis (Dorigo, 1992) and applied it to the Traveling Salesman Problem (TSP). It has been applied to the quadratic assignment problem (Maniezzo et al., 1994),the vehicle routing problem (Bullnherimer et al., 1999), bin packing, stock cutting (Ducatelle and Levine, 2001) and RNA secondary structure prediction (McMella, 2006).

Ant Colony System (ACS) is an improved Ant System (AS) (Dorigo, 1992) in three main aspects: state transition rule, global updating rule, and local pheromone updating rule (Dorigo and Luca, 1997)The ACS algorithm for DNA Sequence Design Problem can be seen in (Kurniawan et al., 2008).

2.3 Hierarchical Updating Pheromone in Ant Colony System

Finite state machine as a model for constructing aDNA sequence was proposed by (Kurniawan et al., 2008) as shown in Figure3. Since there are only four nodes in the state as 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 tend to choose these path and the DNA sequence obtained from this tour will have pattern

Table 1: Comparison Results of Proposed Approach and result of previous work (Kurniawan et al., 2008)

Sequence	*continuity*	*hairpin*	*hmeasure*	*similarity*
[Previous work (Kurniawan et al., 2008)]				
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
[Proposed approach / Hierarchical pheromone updating]				
ACTAGCGACCGGATGATTAC	0	0	80	64
TCTACTACTCTACTCCGGCA	0	0	80	66
CGTAGCACCTTGCATTGATC	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

repeating in some DNA bases only as shown in Figure 2 and the DNA sequences willdifficult 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 having a single set node of DNA for a single run, this model has a set of node of a DNA for each level of construction process. For every level state transition in ACS mechanism, there is a set of DNA paths that can be chosen based on their own pheromone and not to choose node in same level.

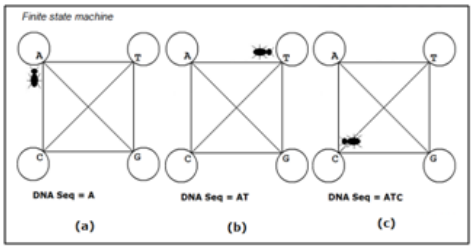

Figure 3: Finite state machine as a model for constructing a DNA Sequence(Kurniawan et al., 2008)

Figure 5 illustrates that in order to generate 20mer of DNA Sequence, there are 20 levels where each level has a set of node of DNAs. 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 node in same level is not connected by any paths for interconnection.Therefore, each node in the

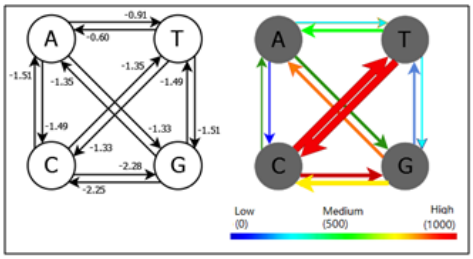

Figure 4: The paths of ACS model for DNA Sequence Problem and their pheromone for each path(Kurniawan et al., 2008

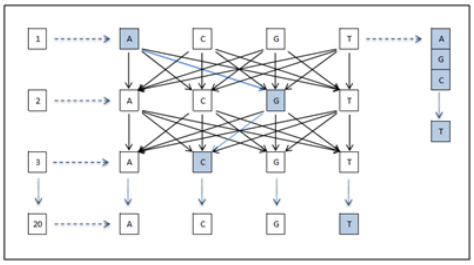

Figure 5: The Proposed Model, Hierarchical Updating Pheromone in ACS for DNA Sequence Design Problem

same level cannot be traversal to each other. This mechanism ensures that the pheromone will be updated with different path for each level of sequence.

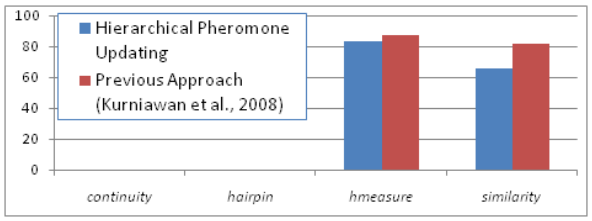

Figure 6: Average fitness comparison results between hierarchical pheromone updating and previous work (Kurniawan et al., 2008)

3 RESULTS AND DISCUSSION

The sequences obtained from the proposed approach arecompared with the sequences generated by different methods.For comparison purposes,each of the compared methods runs 300 times and the results obtained arereported based on a same parameters reported in (Kurniawan et al., 2008).

The comparison results between ourresults obtained from proposed approach with result obtainedfrom previous work (Kurniawan et al., 2008) are shown in Table 1 and Fig 6. Since

Table 2: Comparison Number of failed DNA Sequences

Approach	# Satisfy	# Not Satisfy	Total(300 x 7)
Hierarchical Pheromone Updating	1955(93.10%)	145 (6.90%)	2100
Pervious work (Kurniawan et al., 2008)	1034 (49.24%)	1066(50.76%)	2100

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 Fig 6 show that the new proposed approach obtained a much lower total objective values than the previous work(Kurniawan et al., 2008). The new proposed methodacquires lower similarityand hmeasure objective and the pattern shows that there are not repeating nodes in the solutions.

Further experiment has been conducted by collecting DNA sequencesthat were not able to satisfy the constraints. The collected sequences from two compared methods are shown in Table 2.The experiments runs for 300 iterationsand 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 shows that the proposed method has smaller value (6.90%) for not satisfying the DNA Sequenceconstraints compared to previous works (50.76%). The results shows that the new method has better opportunity to get a better solutions.

4 CONCLUSION

A new method called Hierarchical Pheromone Updating Mechanism for ACS in Designing DNA is proposed and implemented with four objective functions: *hmeasure, similarity, continuity*, and *hairpin* and two constraints: *GC-content* and *melting temperature*. The DNA sequences obtained from the proposed method were compared with the resultsreported in (Kurniawan et al., 2008). The results show that new method can generate better DNA sequences where there are no repeated pattern of nodes and get more chance to satisfy the constraints in order to obtain better results in term of *hmeasure* and *similarity* objective.

References

Brenneman.A, Condon. A, (2002), Strand design for biomolecular computation, *Theory: Comput. Sci*, Vol. *287*.

Bullnheimer B., Hartl R.F., Strauss C., (1999), An improved Ant System algorithm for the Vehicle Routing Problem, *Annals of Operations Research*, Volume *89*, Issue 0, 319 328.

Deaton R., Chen J., Bi H., Garzon M., Rubin H., and Wood D.H., (2002), A PCR based protocol for in vitro selection of noncrosshybridizingolgionucleotides, *Proceedings of the 8th International Workshop DNA Based Computers*, pp. 196204.

Dorigo M., (1992), Optimization, learning and natural algorithms, *PhD Thesis*, Dipartimento di Elettronica, Politechico di Milano, Italy.

Dorigo M., Luca. M.G, (1997), Ant Colony System: A cooperative learning approach to the Traveling Salesman Problem, *IEEE Trans. Evol, Comput*, vol. *1*.

Ducatelle F. and Levine J., (2001), Ant Colony Optimisation for Bin Packing and Cutting Stock Problems, *presented at UK Workshop on Computational Intelligence (UKCI-01)*, Edinburgh.

Feldkamp U., Saghafi S., Banzhaf W., and Rauhe H., (2001), DNA sequence generator - A program for the construction of DNA sequences, *Proceedings of the 7th International Workshop DNA Based Computers*, pp. 179188.

Frutos A.G., Thiel A.J., Condon A.E., Smith L.M., and Corn R.M., (1997), DNA computing at surfaces: Four base mismatch word design, *Proceedings of the 3rd DIMACS Workshop DNA Based Computers*, pp. 238.

Hartemink A,J., Gifford D.K, and Khodor J., (1998), Automated constraint based nucleotide sequence selection for DNA computation, *Proceedings of the 4th DIMACS Workshop DNA Based Computers*, pp. 227-235 36.

Kurniawan, T.B., Khalid, N.K., Ibrahim, Z., Khalid, M., Middendorf, M., (2008),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. 144149.

Maniezzo V., Colorni A., and Dorigo M., (1994), The ant system applied to the quadratic assignment problem,UniversiteLibre de Bruxelles, Belgium, *Tech. Rep.* IRIDIA/94-28.

Marathe A., Condon A.E, and Corn R.M, (1999), On combinatorial DNA word design, *Proceedings of the 5th DIMACS Workshop DNA Based Computers*, pp. 7589.

McMellan, N, (2006), RNA Secondary Structure Prediction using Ant Colony Optimisation, *Master Thesis*, School of Informatics, University of Edinburgh.

Penchovsky R. and Ackermann J., (2003), DNA library design for molecular computation, *Journal of Computational Biology*, Vol. *10*, No. 2, pp. 215229.

Shin S.Y., Lee I.H., Kim B, Zhang B.T. (2005), Multiobjective evolutionary optimization of DNA sequences for reliable DNA computing, *IEEE Transaction on Evolutionary Computation*, Vol. *9*, No. 2.

Tanaka F., Nakatsugawa M., Yamamoto M., Shiba T., and Ohuchi A., (2001), Developing support system for sequence design in DNA computing, *Proceedings of the 7th International Workshop DNA Based Computers*, pp. 340349.

Zhang B.T. and Shin S.Y., (1998), Molecular algorithms for efficient and reliable DNA computing, *Proceedings of Genetic Programming*, pp. 735742.