International Journal of Advanced Research in Engineering and Technology (IJARET)
Volume 8, Issue 3, May - June 2017, pp. 12–18, Article ID: IJARET_08_03_002
Available online at http://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=8&IType=3
ISSN Print: 0976-6480 and ISSN Online: 0976-6499

FAST PARALLEL DNA SOLUTION TO ORIENTED COLORING PROBLEM

Kavitha J
Department of Basic Sciences and Humanities
New Horizon College of Engineering
Bangalore, India.

ABSTRACT

In this paper, a DNA based computing model for solving the oriented coloring problem is proposed. This model shows how to use DNA strands to construct solution space of molecules for the oriented coloring problem and how to apply the DNA algorithm to solve the oriented coloring problem using biological operations. The algorithm is highly parallel and has satisfactory fidelity. The time complexity of the algorithm is O (n2+m2), where n is the number of vertices of the graph and m is the number of edges of the graph.

Key words: NP-complete problem, oriented coloring problem, DNA based parallel algorithm, parallel computing, Polynomial time algorithm, Time complexity.

Cite this Article: Kavitha J Fast Parallel DNA Solution to Oriented Coloring Problem**.** *International Journal of Advanced Research in Engineering and Technology,* 8(3), 2017, pp 12–18.
http://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=8&IType=3

1. INTRODUCTION

Through advances in molecular biology [1, 2], it is now possible to produce 10^{18} or more DNA strands in a tube. Those 10^{18} or more DNA strands can also be applied for representing 10^{18} or more bits of information. Biological operations can be used to simultaneously operate 10^{18} or more bits of information. Or we can say that 10^{18} or more data processors can be executed in parallel. Hence, it becomes obvious that biological computing can provide a very huge parallelism for dealing with problems in the real world. Especially, the problems from the NP-complete class are well known to be exponentially more difficult than evaluating determinants whose entries are merely numerical. It is very difficult to solve these kinds of problems even if very massive supercomputers are used when the problem size becomes large.

On the other hand, DNA computers have full potential of high performance computing Technology [12,13,14,15]. One test tube can be viewed as a processing unit like standard computer architecture. Furthermore, DNA algorithms using biological operations have natural parallelism because DNA strands are separated melted, annealed in test tubes in parallel

Feynman [3] first proposed molecular computation in 1961, but his idea was not implemented by experiment for a few decades. In 1994 Adleman [1] succeeded to solve an instance of the Hamiltonian path problem in a test tube, just by handling DNA strands. Lipton [4] demonstrated that the Adleman techniques could be used to solve the satisfiability problem. Adleman et al. [5] proposed sticker for enhancing the Adleman- Lipton model. In recent years methods for solving several well known NP- Complete problems [16,17,18,19,20] have been proposed.

In this paper, we develop a DNA-based algorithm to solve the oriented coloring problem, which is a well known NP-complete problem, based on Adleman-Lipton model. we use DNA sequence to construct a solution space for the oriented coloring problem. Furthermore, this work presents clear evidence of the ability of DNA based computing to solve NP complete problems. The rest of the paper is organized as follows. In section 2, the Adleman-Lipton model is introduced in detail. In section 3, the oriented coloring problem is defined and the construction of a solution space for the oriented coloring problem is introduced. In section 4, a DNA algorithm is proposed to solve the oriented coloring problem of any directed graph with n vertices for a given three colors. The time complexity of the proposed algorithm is described and the correctness of the algorithm is discussed.

2. THE ADLEMAN-LIPTON MODEL

DNA is the major information storage molecule in living cells. A DNA (deoxyribonucleic acid) is a polymer, which is strung together from monomers called deoxyribonucleotide. Distinct nucleotides are detected only with their bases. Those bases are respectively, abbreviated as A (adenine), G (guanine), C (cytosine) and T (thymine). Two single strands of DNA can form a double strand, if the respective bases are the Watson- Crick complements of each other - A matches T and C matches G; The length of the single stranded DNA is the number of nucleotides comprising the single strand. The length of the double stranded DNA is counted in the number of base pairs.

The Adleman-Lipton model: The DNA operations in the Adleman-Lipton model [1, 4, 6, 7, and 8] are described below. These operations will be used for figuring out solution of the oriented coloring problem. A test tube is a set of molecules of DNA (that is a multi-set of finite strings over the alphabet A, C, G, T). Given a tube, one can perform the following operations:

- **Denaturation:** Given a test tube T, Denaturation (T) dissociates each double strand in T into two single strands.
- **Annealing:** Given a test tube T, Annealing (T) produces all feasible double strands in T. (The produced double strands are still stored in T after annealing).
- **Synthesis:** Synthesis (to produce) a DNA of a desired strand.
- **Amplification:** To make copies of the given DNA strands.
- **Cutting:** Cut a DNA at a particular place in the strand.
- **Ligation:** Ligate DNA strands with complementary sticky ends.
- **Extract:** Given a tube T and a short single strand of DNA, S, the operation extract produces two tubes + (T, S) and − (T, S). + (T, S) is all of the molecules of DNA in T which contain the strand S as a sub-strand and − (T, S) is all of the molecules of DNA in T which do not contain the short strand S.
- **Detect:** Given a tube T, the answer is ‘yes’ if T includes at least one DNA molecule, and the answer is ‘no’ if it contains none.

- **Discard:** Given a tube T, the operation will discard the tube T.
- **Read:** Given a tube T, the operation is used to describe a single molecule, which is contained in the tube T. Even if T contains many different molecules each encoding a different set of bases, the operation can give an explicit description of exactly one of them.
- **Union (T_i, T):** This operation in parallel creates the set T which is the set union of the sets T_i.

3. THE ORIENTED COLORING PROBLEM

Given a graph G = (V, E), where V is the set of vertices and E is the set of edges with |V | = n and |E| = m. In graph theory, oriented graph coloring is a special type of graph coloring. Oriented coloring is an assignment of colors to vertices of an oriented graph such that no two adjacent vertices get the same color, and if (x, y) and (u, v) are arcs of the graph then it is not possible that colors of x and v and of y and u are the same. An oriented chromatic number of a graph G is the least number of colors needed in an oriented coloring. For a given graph G determining any assignment of 3 colors to G is a oriented coloring of G or not is an NP- complete problem [9].

3.1. Construction of Solution of DNA Sequence for the Oriented Coloring Problem

In the Adleman-Lipton model, their main idea is to first generate solution space of DNA sequences for those problems resolved. Then, basic biological operations are used to select legal solutions from the solution space. Therefore, the first step of resolving the oriented coloring problem is to produce a test tube which contains all possible assignment of colors to the vertices of the graph. The input is a directed graph G = (V, E), where V is the set of vertices and E is the set of edges. |V | represents the number of vertices in G and |E| represents the number of edges in G. Let |V | = n, |E| = m and the three colors be c_1, c_2, c_3.

The initial set which contains an assignment of colors to the vertices of the graph of the form $GGGN_ip_1c_sp_1N_iCCCGGGN_ip_2c_sp_2N_iCCC...GGGN_ip_nc_sp_nN_iCCC$, where s = 1,2,3, p_i represents the position of the vertex v_i which is a 4-mer DNA sequence, N_i is a 5-mer DNA sequence represents the ith assignment of colors to the given graph where $1 \leq i \leq 3^n$. For different coding of N_i,

$1 \leq i \leq 3^n$, the DNA strand $GGGN_ip_1c_sp_1N_iCCCGGGN_ip_2c_sp_2N_iCCC...GGGN_ip_nc_sp_nN_iCCC$ represents different assignment of colors to the given graph G. The numbers of vertices are more, use different length of oligonucleotide. An edge from a vertex v_i to v_j is encoded as $P_i^C N_i^C GGGCCCN_i^C P_j^C$ for all i = 1 to 3^n. The initial set of DNA molecules encoding all candidate solution to the oriented coloring problem is synthesized using ABI3948 nucleic acid synthesis and purification system [10, 11].

Illustration:

Consider the following graph.

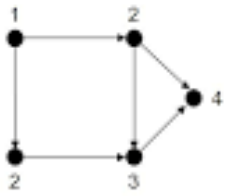

Figure 1

The given graph has 5 vertices. The position of the five vertices of the graph and colors c_1, c_2, c_3 is encoded as follows:

P_1: ATGC P_2: AATG P_3: GCTA P_4: CGAA P_5: TTCG c_1: gtat c_2: gatt c_3: tatt

Our DNA model involves a long single strand which made of number of sub strands, and each sub strand represents the position of a vertex with an assigned color. The algorithm 1 uses three colors c_1, c_2, and c_3 to color the given graph. Single strand in the form $GGGN_ip_1c_sp_1N_iCCC$ are used to encode one possible coloring of each vertex in the graph. The sequence N_i is used to find the ith assignment of color to the given graph. An assignment of colors to the given graph is of the form $GGGN_ip_1c_sp_1N_iCCCGGGN_ip_2c_sp_2N_iCCC...GGGN_ip_nc_sp_nN_iCCC$. For different coding of N_i, i = 1 to 3^n, this strand represents all possible encoding of the given graph. The middle sequence CCCGGG is recognizable by restriction endonuclease SmaI which can split it at the middle site. For the above given graph, we can generate all possible assignment of colors to the given graph using the DNA codes given above. Thus, all the 3^5 DNA strands which encode the assignments of colors to the graph can be synthesized using ABI 3948 nucleic acid synthesis and purification system.

4. THE DNA ALGORITHM FOR SOLVING THE ORIENTED COLORING PROBLEM

4.1. Algorithm1

The proposed DNA-based algorithm to solve the oriented coloring problem is described in this section. It can be applied to solve the oriented coloring problem of any directed graph with n vertices for a given three colors.

Input (T), where the tube T, includes solution space of DNA sequences which are encoding of all possible assignments of three colors to the vertices of the given graph G.

For j = 1 to n

For s = 1, 2, 3 and all k such that (j, k) $\in$ E

$T_1 \leftarrow + (T, p_jc_sp_j)$

$T_2 \leftarrow - (T, p_jc_sp_j)$

$T_3 \leftarrow + (T_1, p_kc_sp_k)$

$T_4 \leftarrow - (T_1, p_kc_sp_k)$

discard (T_3)

$T = T_2 \cup T_4$

EndFor

EndFor

If Detect (T) = yes then

For j = 1 to n-1

For s = 1, 2, 3 and all k such that (j, k) $\in$ E

$T_1 \leftarrow + (T, p_jc_sp_j)$

For l = j+1 to n

For s = 1, 2, 3 and all m such that (l, m) $\in$ E

$T_2 \leftarrow + (T_1, p_m c_s p_m)$

Discard(T_2)

$T_3 \leftarrow - (T_1, p_m c_s p_m)$

EndFor

EndFor

EndFor

EndFor

If Detect (T_3) = yes then

Proper colorings in T_3 are oriented coloring of G.

else

oriented coloring is not possible

else

No coloring is proper coloring.

4.2. Implementation of the Algorithm

This section describes the implementation of the algorithm.

- Start with all the DNA sequences that represent the assignments of colors to the given graph G.
- Eliminate all the DNA strands that do not represent the proper coloring of the graph. For each edge (j, k) ∈ E, remove the strands from a test tube T, that contain subsequences $p_j c_s p_j$, $p_k c_s p_k$, where s = 1, 2, 3.
- If the test tube T does not contain DNA strands, then conclude no coloring is a proper coloring. If not for every pair of edges (j, k) and (l, m) ∈ E, the algorithm checks the colors of the vertices of j and m and of k and l. If the colors of j and m and of k and l are not same then the proper coloring of G is the oriented coloring of G.

4.3. The Complexity of the DNA Algorithm

The oriented coloring problem with three colors for any directed n-vertex graph G can be solved with O (n^2+m^2) biological operations in the Adleman-Lipton model. The algorithm described in section 4 can be applied for solving the oriented coloring problem for any directed n-vertex graph G with given three colors. This Algorithm includes three main steps. The steps 2-11 are mainly used to determine the proper coloring of G and to remove illegal coloring of G from all of the 3^n possible assignment of colors to the given graph G. For each vertex $v_i \in V$,$1 \leq i \leq n$ and the three colors c_1, c_2, c_3, the steps 4 and 5 take 3^n extraction operations. Since every vertex $v_i \in V$, has at most n−1 adjacent vertices, the steps 6 and 7 take 3n×3(n−1) extraction operations. The step 8 takes 3n discard operations. The steps 13-23 are used to find whether in proper coloring of G, for any two edges

(j, k) and (l, m) the vertices j and m and k and l have the same color. The steps 13-23 needs mC_2 extraction operations. Hence, the time complexity of Algorithm is O (n^2+m^2) biological operations in the Adleman-Lipton model.

4.4. Correctness of the Algorithm

The oriented coloring problem with three colors for any n-vertex directed graph can be resolved using the algorithm proposed in section 4. The input of the algorithm is a test tube,

which contains all 3^n possible assignment of colors to the given n- vertex directed graph G. First the algorithm detects all possible proper coloring of G by repeating the steps 2 to 11 for n number of times. The first test tube T_1 contains all the DNA strands in which the first vertex v_1 has the color c_1, the second test tube T_2 contains all the DNA strands in which the first vertex v_1 has the color c_2 or c_3. The third test tube T_3 contains all the DNA strands in which the vertex v_1 and its adjacent vertices have the color c_1. The fourth test tube T_4 contains all the strands in which the vertices adjacent to the vertex v_1 have the colors c2 or c3. Therefore the test tube T_3 collects all the strands in which the vertices adjacent to v_1 and the vertex v_1 have the same color and the test tube T_4 collects all the strands in which the vertices adjacent to v_1 and the vertex v_1 have different colors. Step 8 uses the "discard operation" to remove all the illegal coloring to the vertex v_1 and its adjacent vertices for the color c_1. Step 9 merges the content of the tubes T_2 and T_4. Now the test tube T contains the DNA strands in which the first vertex v_1 has the color c_1 and its adjacent vertices have different colors c_2 or c_3. By repeating the same procedure the test tube T contains all the DNA strands in which the vertex v_1 and its adjacent vertices have different colors. The steps 2-11 are repeated for all the remaining n−1 vertices. If the test tube T does not contain a DNA strand, no coloring is a proper coloring.

If not for every pair of edges (j, k) and (l, m) ∈ E, the algorithm checks the colors of the vertices of j and m and of k and l. If the colors of j and m and of k and l are not same then the proper coloring of G is the oriented coloring of G. Otherwise oriented coloring is not possible.

5. CONCLUSION

The major advantage of DNA computing lies in its high parallelism. In this paper, a DNA based algorithm for solving oriented coloring problem based on biological operations in Adleman-Lipton model has been proposed. This algorithm can determine not only the oriented coloring but also all the oriented coloring of the given graph in polynomial time. The efficiency of this method can be seen from the time complexity of our algorithm O (n^2+m^2).

REFERENCES

[1] Adleman , L. Molecular solutions to combinatorial problems. Science. 1994, 266, 1021-1024.

[2] Sinden.. R.R.: DNA structure and Function. Academic Press. New York. 1994I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[3] Feynman, R.P. in Gilbert,D.H.(Ed.), Minaturization. Reinhold Publishing Corporation. New York. 1961,282- 296

[4] Lipton, R.J.: DNA solution of hard computational problems. Science. 1995, 268, 542-545

[5] Roweis, S., Winfree, E., Burgoyne, R., Chelyapov, N.V., Goodman,M.F., Rothemund, P.W.K., Adleman,L.M. . A Sticker based model for DNA computation, in: Landweber,L., Baum, E.(Eds.), 2nd Annual Workshop on DNA Computing, DIMACS : Series in Discrete Mathematics and theoretical Computer science, American Mathematical Society, Princeton University. 1999, 1-29

[6] Boneh,D., Dunworth,C., Lipton,R.J., Sgall,J. : On the computational power of DNA, in. Discrete Applied Mathematics. Special Issue on Computational Molecular Biology. 1996, 71, 79-94

[7] Paun,G., Rozenberg,G., Salomaa,A. DNA Computing: New Computing Paradigm. Springer-Verlag. New York. 1998, ISBN : 3-540-64196-3

[8] Adleman,L.M. . On constructing a molecular computer, DNA Based Computers, in : Lipton,R., Baum,E.(Eds.), DIMACS series in Discrete Mathematics and Theoretical Computer Science. American Mathematical Society. 1996, 1-21

[9] Garey,M.R., Johnson,D.S. Computer and intractability. Freeman, San Fransico, CA,1979

[10] Qinghua Liu. et al. DNA computing on surfaces, Nature. 2000, 403, 175-179

[11] Chee,M. et al. Accessing genetic information with high-density DNA arrays. Science 1996, 276 , 610-614 .

[12] Yu-Xing Yang, Ai-Min Wang, Ji-Ian Ma: Multiseparation based DNA algorithm of graph vertex coloring problem. 2008, ICNC 7, 547-550

[13] Ya Chun Liu et al. DNA solution of graph coloring problem. Journal of chemical information and modeling 2002, 42(3) , 524-528

[14] YMiniyi Guo et al. Is optimal solution of every NPcomplete or NP-hard problem determined from its characteristic for DNA-based computing. Bio systems. 2005, 80(1), 71-82

[15] Jiang Xingpeng et al. A new DNA algorithm to solve graph coloring problem. Progress in natural science 2007, 17(6), 733-738

[16] Wang.S, Yuan. J. DNA computing of directed linegraphs. Communication in mathematical and in computer chemistry 2006, 56(3), 479-484

[17] Israel Marck, K. H. Zimmermann. Parallel bioinspired algorithms for NP-complete graph problems. Journal of parallel and distributed computing 2009, 69(3), 221-229

[18] Guillermo M. Mallen_ Fullerton et al. Modified classical graph algorithms for the DNA fragment assembly problem, Algorithms, 2015, 8, 754-773.

[19] Li et al, Solving the maximum weight vertex independent problem with DNA molecules in Adleman-Lipton model, Journal of computational and theoretical nanoscience, 12, 8, 2015, 1940 – 1943.

[20] Zhiquan Frank Qiu, Mi Lu . A new approach to advance the DNA computing. Applied soft computing 2003, 3(2), 177-189.

[21] Dr. Kavitha Joseph. Fast Parallel DNA Algorithm based on Adleman-Lipton Model: The Independent Dominating Set Problem. International Journal of Computer Engineering and Technology, 6(11), 2015, pp. 01-10.