Improved Algorithm for Solving the Maximal Clique Problem with DNA Computing

H. Aoyagi [#1], N. Rajaee[#1] and O. Ono [#1]

[#] *Department of Electronic and Bioinformatics, Meiji University,*
1-1-1 Higashimita, Tamaku, Kawasaki, Kanagawa, 214-8571 Japan
(Tel : 044-934-7377; Fax : 044-934-7289)
[1] *(ee77082, nrajaee, ono)@isc.meiji.ac.jp*

Abstract: In this paper, we improve an algorithm proposed by Q. Ouyang to solve a maximal clique problem. While we can determine the size of the largest clique by sorting the length of the DNA strands in Q. Ouyang proposal, we cannot determine the exact vertices involved in the maximal clique. Thus in our improved algorithm, the exact vertices for each clique can be derived straightforwardly from the read-out in the gel electrophoresis process.

Keywords: DNA computing, the maximal clique problem, the parallel overlap assembly

I. INTRODUCTION

DNA computing which uses bio-molecules as a basis for computation is thought one day to be able to outperform electronic computers in computing complex combinatorial problems such as NP hard. Since then, many proposals to solve NP hard problems with DNA computing have been published including an algorithm proposed by Q. Ouyang [1] for solving a maximal clique problem (MCP) in 1997. The MCP is encoded in DNA sequences by assigning a binary number to each possible clique in the graph as the data structure. A bit set to 1 represents a vertex in a clique and a bit set to 0 represents a vertex out of the clique. Therefore for a graph with N vertices, the complete set of possible cliques is ensemble into an N digit binary numbers. The data structure is then designed in the form of double stranded DNA (dsDNA) and restriction enzyme sites are embedded for each bit value = 1. These restriction enzyme sites provide a "cutting" reaction to the enzymes and were not amplified during PCR. Reading the size of the largest clique is then based on the shortest length of the elongated DNA strands.

In our work, we improve the algorithm proposed by Q. Ouyang for better qualitative results of the computation. While we can determine the size of the largest clique by sorting the length of the DNA strands in Q. Ouyang proposal, we cannot determine the exact vertices involved in the maximal clique. Thus in our improved algorithm, the exact vertices for each clique can be derived straightforwardly from the read-out in the gel electrophoresis process.

II. THE MAXIMAL CLIQUE PROBLEM

A complete graph is a simple graph in which every pair of distinct vertices is connected by an edge. The MCP is to find a maximum size complete sub-graph, a maximal clique, from a given graph. The MCP is proven to be an NP-hard problem, and is difficult to solve efficiently.

III. ALGORITHM

We discuss an algorithm to solve the MCP by showing the graph example in Fig.1. Consider a graph with eight vertices, sixteen edges and three maximal cliques. Fig.3 shows an adjacency matrix of the given graph. The complementary graph (Fig.4) contains all edges missing in the original graph.

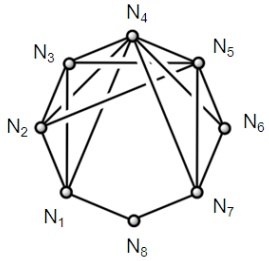

Fig.1. The graph of our example

Showing the connection of the node N_1

		N_1	N_2	N_3	N_4	N_5	N_6	N_7	N_8
Tube N1	N_1	1	1	1	1	0	0	0	1
Tube N2	N_2	1	1	1	1	1	0	0	0
Tube N3	N_3	1	1	1	1	1	0	0	0
Tube N4	N_4	1	1	1	1	1	1	1	0
Tube N5	N_5	0	1	1	1	1	1	1	0
Tube N6	N_6	0	0	0	1	1	1	1	0
Tube N7	N_7	0	0	0	1	1	1	1	1
Tube N8	N_8	1	0	0	0	0	0	1	1

Fig.2. Adjacency matrix for the graph

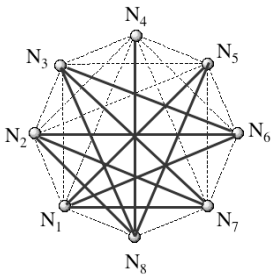

Fig.3. The complementary graph

The new algorithm consists of the following six steps.

Step 1 Cording (DNA sequence Generator)

First, we construct the sequence for value N_i and position E_i as described in Fig.1. There are eight value sections (N_1 - N_8) and its position (E_1 - E_9). The position sequences are used for connecting each node N_i (i=0,…,8). The eight value sections are set sandwiched between nine positions. Each combined oligonucleotide consists of two position motifs, $E_{i+1}N_iE_i$ for odd i, $\overline{E_{i+1}N_iE_i}$ for even i, where the bar shows the complementary sequence and the value of N_i can be either 1 or 0. In this experiment, we designed the DNA sequences in the form of ssDNA. We set E_i = 10 bp, N_i = 0 bp (if the value = 1) and N_i =15 bp (if the value = 0). The sequence of vertices are constructed from restriction enzyme (RE) sites for each bit value = 1.

Step 2 Merge (Parallel Overlap Assembly)

We produce a pool containing all possible solution with eight tubes T1-T8. Following Fig.2, we add combined ENE DNA sequences including nodes N_i^1 showing value 1 and all nodes N_i^0 showing value 0 into each tube. By performing POA, the DNA sequences for a position are annealed to the complementary sequence for the other DNA sequences and the combinations of $N_1N_2N_3N_4N_5N_6$, a pool containing possible solution, is generated.

Step 3 Cutting (Restriction Enzyme)

Contents of T1-T8 are digested with each RE (RE1-RE8) to eliminate illegally connected edges (Fig. 3) in each test tube.

Step 4 Amplify (Polymerase Chain Reaction)

Amplify only strands with E_0 and $\overline{E_9}$ for each tube by supplying primers E_0 and $\overline{E_9}$.

Step 5 Extract (Gel Electrophoresis)

The clique of the largest size is represented by the shortest DNA length. From gel electrophoresis process, extract only the shortest band 150bp (clique size = 4) for T1-T8 and separate into tubes T1'-T8'. Add RE1-RE8 into tubes T1'-T8' and observe the band lengths.

Step 6 Identify

The experiment result is observed by performing gel electrophoresis. Each vertex involved in individual cliques can be identified separately by the clustering of lengths for the vertices. From the result, you can find the vertices (1,2,3,4), (2,3,4,5), and (4,5,6,7) form the maximal cliques with size = 4.

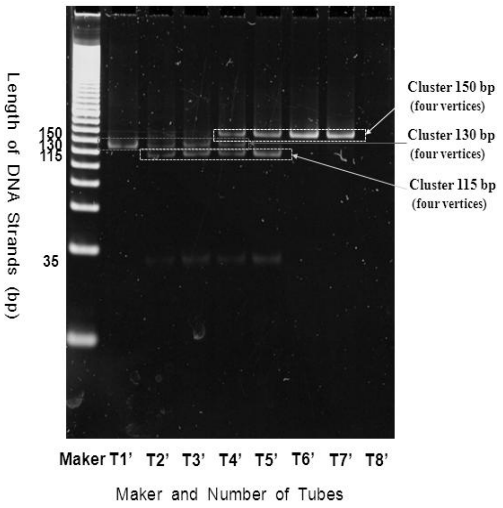

Fig.4. Gel Electrophoresis

VI. CONCLUSION

The experiment was carried out using basic bio-molecular tools as used such as parallel overlap assembly, polymerase chain reaction (PCR) and gel electrophoresis process. A major difference in our proposal is the inclusion of Step 6. In the original algorithm by Q. Ouyang, separate vertices involved in a clique were confirmed using molecular cloning technique which is time and resources inefficient. Using our algorithm the size of each clique can be determined by its DNA strand length and each vertex involved in individual cliques can be identified separately.

REFERENCES

[1] Q. Ouyang, P. D. Kaplan, S. Liu et al. (1997), DNA Solution of the Maximal Clique Problem, Science 278: 466-449

[2] I. Kim, D. J. Jeng, and J. Watada (2006), Redesigning Subgroups in a Personnel Network Based on DNA Computing, International Journal of Innovative Computing, Information and control 4:885-896