

SELF-ASSEMBLED DNA COMPUTING FOR DIRECTED GRAPH PROBLEM MODELLING

Nordiana Rajae and Osamu Ono

Institute of Applied DNA Computing,
Faculty of Electronics and Bioinformatics Engineering,
Meiji University
1-1-1 Higashi-mita, Tama-ku
Kawasaki-shi, Kanagawa-ken
214-8571 Japan
(*nrajae,ono*)@isc.meiji.ac.jp

Abstract

This paper presents modeling of a directed graph problem for Boolean matrix operation using self-assembled DNA computing. In this experiment, we show how a Boolean matrix product operation may be represented by a directed graph problem and solved using DNA computing. The directed graph problem which consists of vertices and paths are recreated using DNA oligonucleotides and it will be proven in this experiment that these yield the same results as the actual directed graph problem. The algorithm to model the self-assembled DNA into directed graph is based on Adleman-Lipton architecture. However, unlike the Adleman-Lipton's, restriction enzymes which are used to cut the DNA into unique sequences representing vertices and paths are replaced by Parallel Overlap Assembly (POA) method. Polymerase Chain Reaction (PCR) method is later used as test of reaction to identify the existence of paths to validate the accuracy of results which can be visualized through gel electrophoresis and UV process.

Keywords: DNA computing, self-assembled, POA, directed graphs, PCR.

Presenting Author's biography

Nordiana Rajae received her master degree from University of Newcastle upon Tyne, United Kingdom and is currently a postgraduate student in Institute of Applied DNA Computing, MEIJI University, Japan.



1 Introduction

Deoxyribonucleic or DNA is most commonly known as gene information carrying strands for all living matters. An interesting property regarding DNA is that it is made up from four nucleotide building blocks, A (adenine), T (thymine), C (cytosine) and G (guanine) where Watson-Crick complementary rule of A and T, C and G pairs hybridize to form double stranded molecules. Due to this unique property, the possibilities of exploiting this property into many applications have intrigued many researches in various fields, including molecular computing and bio engineering [1,2].

DNA computing was first introduced by Leonard M. Adleman in 1994 when he solved an ordinary problem with an extraordinary method – tackling the Hamiltonian Path Problem (HPP) using DNA. In HPP, the goal is to find a path through a directed graph that starts and ends at specified vertices and visits each vertex in the graph once, and only once. In his works, Adleman demonstrated an actual laboratory experiment to prove that such problem can be solved using DNA by conducting basic bio-molecular reaction fashion involving hybridization, denaturation, ligation, magnetic bead separation, polymerase chain reaction (PCR) [1,2]. By encoding DNA oligonucleotides to represent the vertices and paths, molecules representing the paths through the graph will be formed upon hybridization and ligation processes. This technique uses restriction enzyme to cut DNA into appropriate sequences to represent desired vertices and paths.

Due to his work, Adleman has proven that DNA computing can be used to solve computational problems. Since the basis for modern computers nowadays are based on Boolean matrix operations and logic functions, it is not impossible to use similar algorithm to recreate these using DNA as well.

However, restriction enzymes used in Adleman works cause liabilities to DNA computing as it is not only limited in number which constrains larger computations, but the experiments with restriction enzymes cannot guarantee fidelity in their results. As such, several other computing algorithms without using restriction enzymes were proposed. One of those is the Parallel Overlap Assembly (POA) method [3, 4]

Hence in this paper, instead of using restriction enzymes we will be using POA method to model a directed graph problem for Boolean matrix operations.

2 Modeling a Directed Graph Problem

2.1 Boolean Matrix Operation

In this experiment, we model a directed graph problem for Boolean matrix product operation in Figure 1.

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

Figure 1 Boolean matrix product operation

According to Kim (1982), the Boolean matrix operation in Figure 1 may be represented by a directed graph in Figure 2. [4]

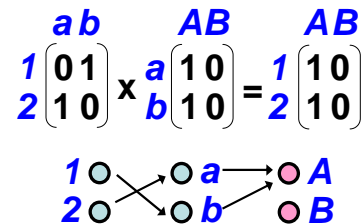


Figure 2 Directed graph representation for Boolean matrix product operation in Figure 1

A directed graph problem consists of vertices and paths to define a problem and the results, depending on the complexity of the problem, may be solved visually. As in Figure 2, the Boolean matrix product operation in Figure 1 may be represented by a directed graph problem with start points 1 and 2, intermediate points a and b, and end points A and B. These six points are called vertices. The contents of the Boolean matrix product operation represent the relationships between the said vertices. A path indicates value 1 and no path indicates value 0. For example, a Boolean relationship of value 1 is shown as a path between two vertices such as path 1b connecting vertices 1 and b.

If there is no path shown between two vertices, this indicates a Boolean relationship of value 0 between the vertices such as no path between 1 and a. The product matrix of the Boolean operation in Figure 1 is simplified as in Figure 3 where each result is allocated to different test tubes in the experiment.

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} T1 & T2 \\ T3 & T4 \end{pmatrix}$$

Figure 3 Allocation of each matrix product result into different test tubes.

2.2 Generating vertices and paths with DNA

Single stranded DNA sequences are used to represent the vertices in Figure 2. These single stranded DNA sequences are unique to one another and can be simulated using DNASequencesGenerator software. Table 1 shows the generated DNA sequence for the

vertices consisting of five 20-mer length sequences and one 30-mer length sequence for variation.

Table.1 DNA Sequences for Vertices

Vertices	DNA Sequence (5' – 3')
1	CAGCCACGTAGTAGAGCTAG
2	CTACCCGATGTAGTGCGTAG
a	CTTCCAGTTTATCGGCCATC
b	AAGTCACGCATCAAGTGTAC
A	AGCTGACAACGAACCAGACA
B	TTAGACCTGATTCAATGCCGAGTCTCACG

Paths are formed when two particular vertices are connected to each other. In this case, the path will contain combinations of the complementary DNA sequence for both vertices.

Table.2 Complementary DNA sequences for Vertices

Vertices	Complementary DNA Sequence (5' – 3')
1'	GTCGGTGCATCATCTCGATC
2'	GATGGGTACATCACGCATC
a'	GAAGGTCAAATAGCCGGTAG
b'	TTCAGTGCGTAGTTCACATG
A'	TCGACTGTTGCTTGGTCTGT
B'	AATCTGGACTAAGTTACGGCTCAG AAGTGC

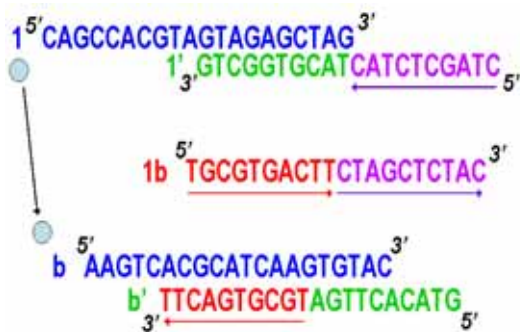


Figure 4 Relationship between DNA sequences for vertices and paths

Referring to Fig.4, a path 1b connects vertices 1 and b. By deriving the complements of both vertices, DNA sequence for a path can be derived from these as well. The DNA sequence for 1b is given as:

5' TCGTGACTTCTAGCTCTAC 3' (20-mer)

where which:

- The first 10-mer of the DNA sequence for a path is derived from the last 10-mer of the second vertex complementary b' (5' – 3')
- The latter 10-mer of the path DNA sequence is derived from the first 10-mer of the first vertex complementary 1'' (5' – 3')

More DNA sequences for paths corresponding to Fig. 2 are shown in Table 3.

Table.3 DNA Sequences for Paths

Paths	DNA Sequence (5' – 3')
1b	TGCGTGACTTCTAGCTCTAC
2a	AAACTGGAAGCTACGCACTA
aA	GTTGTCAGCTGATGGCCGAT
bA	GTTGTCAGCTGTACACTTGA

To stimulate the connection of paths to the vertices, primers are used. Primers are mostly chosen to represent the start points 1 and 2, and end points A and B. Primers in reality are not the start points, end points nor paths. Primers are actually shorter sequences (sub-sequences) of the original DNA sequence for vertices and paths and are used to signal a specific start and end sites of a template for replication. Therefore in this case, there are 4 primers used which are: 1 and 2 (representing start points heading towards end points) and A' and B' (representing end points heading backwards towards start points). Note that the primers for end points are complementary sequences.

According to Figure 2, there are paths from both start points 1 and 2 to end point A. However, there is no path from both start points 1 and 2 to end point B. Hence to check whether such paths exists or not in the modeled directed graph using DNA oligonucleotides, random combinations of start point primers and end point primers are chosen. Altogether we have four combinations of primers – 1A', 1B', 2A' and 2B' for any start points reaching any end points.

3 Recreating Directed Graph Problem for Boolean Matrix Product Operation using DNA

3.1 Parallel Overlap Assembly

Parallel Overlap Assembly (POA) is used as initial pool generation method instead of the more traditional hybridization-ligation method. There are a few advantages of using POA method, one of which is its capability of generating all possible paths from all vertices in massive parallel fashion.

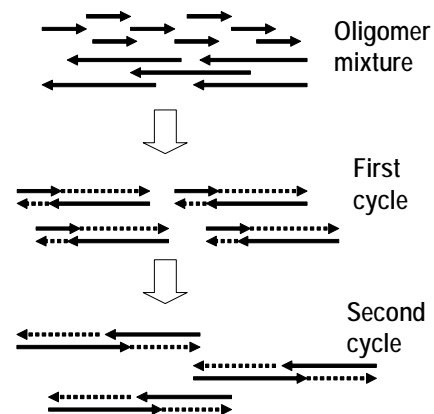


Figure 5 Parallel Overlap Assembly method

Figure 5 shows the operation of Parallel Overlap Assembly method. With each cycle, the DNA sequences will self-assemble to create new combinations of sequences. The POA procedure is usually run for 25 cycles. To recreate the exact same directed graph problem for Boolean matrix product operation in Figure 1, a master template containing all DNA sequences for vertices and paths are required.

In a single test tube, the following measurements were included:

Ddwater	67.5 ul
DNA	1 ul (for every DNA)
dNTP	10.0 ul
KOD	10 ul
KOD Dash	0.5 ul

The DNA template is then run in 25 cycles of 90°C (30sec), 55°C (30sec) and 74°C(10sec) per cycle.

3.2 Polymerase Chain Reaction method

Polymerase Chain Reaction (PCR) method is an effective copying machine for DNA and also used for DNA detection. Given a site-specific single molecule DNA, millions of copies of similar molecules can be created with PCR process. The PCR process makes use of primers to signal a specific start and end sites of template for replication. For the PCR process, the combination of primers to be put into different test tubes depend of the quantity of possible outcomes combinations [5,6].

Therefore in this experiment, the combinations of primers are determined by the matrix results prediction in Figure 3 and combinations of start points and end points as stated in section 2.2. Thus, by referring to matrix results prediction in Figure 3:

- ◆ T1 (test tube 1, combination of primers 1 and A')
- ◆ T2 (test tube 2, combination of primers 1 and B')
- ◆ T3 (test tube 3, combination of primers 2 and A')
- ◆ T4 (test tube 4, combination of primers 2 and B')

The combinations of primers for test tubes 1-4 are also shown in Figure 6.

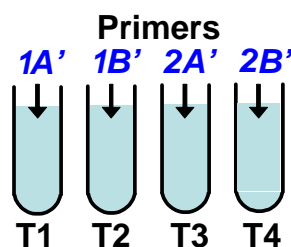


Figure 6 Primer combinations in test tubes

The paths created in POA process have limited amount of DNA sequences, therefore PCR is used to

amplify the existing paths for easier detection in the gel electrophoresis process.

Into each test tube, add the following:

each primer	2.5 ul
ddwater	13.875 ul
template DNA	1 ul
KOD	2.5 ul
KOD Dash	0.125 ul
dNTP	2.5 ul

Tubes are then spun for 13000 rpm in 25°C for 5 minutes before running them into 25 cycles of 90°C (30sec), 55°C (30sec) and 74°C(10sec) per cycle.

The result of this experiment is visualized by gel electrophoresis process to the test tubes T1-T4. Gel electrophoresis is a technique which can separate DNA strands in a solution based on their lengths or weights. In this experiment, we have delegated a length of 20-mer for vertices 1, 2, a, b and A; and 30-mer for B. This technique is based on the fact that DNA molecules are negatively charged. Hence, by putting them in an electric field, they will move towards the positive electrode at different speeds. The longer molecules will remain behind the shorter ones. Thus, by the variations in the lengths of path formations, we would be able to detect the existence of paths between vertices through this technique. Gel is stained with SYBR Gold (Molecular Probes) and the image is captured in UV.

4 Results and Discussions

The captured image of gel for the experiment is shown in Figure 7. The six lanes taken by UV light are (from the right), marker lane, T1 lane, T2 lane, T3 lane, T4 lane and master template DNA lane. The marker lane and the template DNA lanes are used as benchmark for the DNA lengths. From Figure 7, it is highly visible that the T1 and T3 lanes have highlighted bands but there are less visible (none) for lanes T2 and T4. The primer combinations for T1 and T3 are 1 and A', and 2 and A' respectively indicating there are paths from start point 1 to end point A, and from start point 2 to end point B.

These correspond with the predicted Boolean matrix product operation results in Figure 3. Therefore we can conclude that for T1 and T3, the Boolean matrix value is equal to 1 which indicates the existence of path. The non-existence of highlighted bands in T2 and T4 indicate non-existence of paths for primer combinations 1 and B', and 2 and B' with Boolean matrix value of 0.

- T1 = 60 bp highlighted band (value = 1)
- T2 = no highlighted band (value = 0)
- T3 = 60 bp highlighted band (value = 1)
- T4 = no highlighted band (value = 0)

In the highlighted bands of T1 and T3, the last band is at 60 bp ladder. The length of DNA oligonucleotides (in mer) is also counted in bp. Thus, a 60 bp ladder indicates a 60-mer length DNA sequence. This proves that the end DNA sequence for paths is correct as the routes from starting points to end point requires at least three vertices.

(Route 1)

Start point 1 b end point A
 (20 mer) + (20 mer) + (20 mer) = 60 mer

(Route 2)

Start point 2 a end point A
 (20 mer) + (20 mer) + (20 mer) = 60 mer

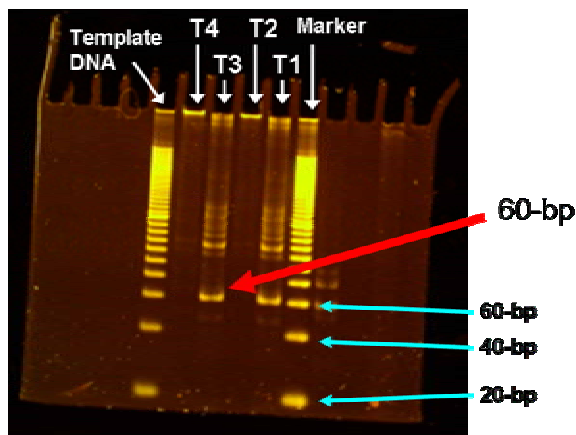


Figure7 Image of gel electrophoresis taken with UV

Conclusion

From the experimental results, we can conclude that the modeling of directed graph problems using DNA self-assembled DNA is proved to yield the same expected results from Boolean matrix product operation. Therefore, it is possible to model directed graph problems using self-assembled DNA oligonucleotides. However, there are several deciding factor to ensure the accuracy of the experimental results. Proper precaution to avoid contamination of DNA and primers has to be taken and the temperature of the wet process must be in suitable range to avoid creation of “faulty” results.

5 References

[1] L.M. Adleman. Molecular Computation of Solutions to Combinatorial Problems. *Science*, 266:1021, 1994.
 [2] L.M. Adleman. On Constructing a Molecular Computer. In *DNA based computers*, 1-22, 1996.
 [3] Yusei Tsuboi, Zuwairie Ibrahim, Osamu Ono: Demonstration of DNA-Based Semantic Model by Using Parallel Overlap Assembly. *ICIC 2005*: 562-570

[4] Kasai Nobuyuki, Zuwairie Ibrahim, Yusei Tsuboi and Osamu Ono. Matrix Multiplication by DNA Computing. 11th International Symposium on Artificial Life and Robotics 2006 (AROB2006), 476-479, 2006.
 [5] Zuwairie Ibrahim, Yusei Tsuboi and Osamu Ono. Direct Proportional Length Based DNA Computing for Shortest Path Problem. *International Journal of Computer Science & Applications*. 1:46-60,2004.
 [6] Ibrahim Z, Tsuboi Y, Ono O, Hybridization-ligation versus parallel overlap assembly: an experimental comparison of initial pool generation for direct-proportional length-based DNA computing, *IEEE Transactions on Nanobioscience*, Volume 2: Issue 2: 103-109, 2006
 [7] R.J. Lipton. DNA solutions of hard computational problems. *Science*, 268:542-545, 1995.
 [8] Russell Deaton, Max Garzon, John Rose, D.R. Franceschetti, S.E. Stevens, Jr. DNA Computing: A Review. *Fundamenta Informaticae*, 30:23-41, 1997.
 [9] robotics.stanford.edu/~serafim/Publications/2000_DNA_computing.pdf
 [10] Calude C.S. and Paun G. Computing with cells and atoms: an introduction to quantum, DNA and membrane computing. Taylor & Francis Inc. 2001.
 [11]Zuwairie Ibrahim, Osamu Ono, Yusei Tsuboi and Marzuki Khalid. A New DNA-Based Computing Approach for Single Source Shortest Path Problem. *International Conference on Intelligent Technologies 2003 (InTech2003)*, 99-107, 2003.
 [12]Peng Chen, Jing Li, Jian Zhao, Lin He and Zhizhou Zhang. Differential dependence on DNA ligase of type II restriction enzymes: A practical way toward ligase-free DNA automaton. *Biochemical and Biophysical Research Communications*, 353:733-737, 2007.
 [13]Shinpei Watanabe, Yusei Tsuboi, Zuwairie Ibrahim, Tsuneto Yamamoto and Osamu Ono. Adaptive DNA Computing Algorithm by Using PCR and Restriction Enzyme. *Proceedings of 2004 IEEE Conference on Cybernetics and Intelligent Systems*, 262-267, 2004.
 [14]M. Amos, DNA Computation, PhD Thesis, University of Warwick, UK, 1997.
 [15]<http://www.cs.rochester.edu/u/ogihara/research/DNA/sigact2.ps.gz>
 [16]Qi Ouyang, Peter D. Kaplan, Shumao Liu, Albert Libchaber, DNA Solution of the Maximal Clique Problem, *Science*, Vol 278:5337:446-449