Hybrid Concentration-Controlled Direct-Proportional Length-Based DNA Computing for Numerical Optimization of the Shortest Path Problem

Zuwairie Ibrahim[1,2], Yusei Tsuboi[2], Osamu Ono[2], and Marzuki Khalid[3]

[1] Department of Mechatronics and Robotics, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor Darul Takzim, Malaysia
zuwairie@fke.utm.my
[2] Institute of Applied DNA Computing (IADC), Meiji University, 1-1-1 Higashi-mita, Tama-ku, Kawasaki-shi, Kanagawa-ken 214-8571, Japan
{zuwairie, tsuboi, ono}@isc.meiji.ac.jp
http://www.isc.meiji.ac.jp/~i3erabc/IADC.html
[3] Center for Artificial Intelligence and Robotics (CAIRO), Universiti Teknologi Malaysia, City Campus, Jalan Semarak, Kuala Lumpur, Malaysia
marzuki@utmkl.utm.my

Abstract. DNA computing often makes use of hybridization, whether for vastly generating the initial candidate answers or amplification by using polymerase chain reaction (PCR). The main idea behind DNA computing approaches for solving weighted graph problems is that if the degree of hybridization can be controlled, then it is able to generate more double stranded DNAs (dsDNAs), which represent the answer of the problem during *in vitro* computation. Previously, length, concentration, and melting temperature, have been exploited for encoding of weights of a weighted graph problem. In this paper, we present a hybrid approach, which is called concentration-controlled direct-proportional length-based DNA computing (CCDPLB-DNAC), that combines two characteristics: length and concentration, for encoding and at the same time, effectively control the degree of hybridization of DNA. The encoding by length is realized whereby the cost of each path is encoded by the length of the oligonucleotides (oligos) in a proportional way. On the other hand, the hybridization control by concentration is done by varying the amount of oligos, as the input of computation, before the computation begins. The advantage is such that, after an initial pool generation and amplification, polyacrylamide gel electrophoresis (PAGE) can be performed to separate the survived dsDNAs according to their length, which directly decodes the results. The proposed approach shows significant improvement in term of materials used and scalability. The experimental results show the effectiveness of the proposed CCDPLB-DNAC for solving weighted graph problems, such as the shortest path problem.

1 Introduction

A new computing paradigm based on DNA molecules has appeared in 1994 when Leonard M. Adleman [1] launched a novel *in vitro* approach to solve the so-called

A.J. Ijspeert et al. (Eds.): BioADIT 2006, LNCS 3853, pp. 206 – 221, 2006.

Hamiltonian Path Problem (HPP) with seven vertices by DNA molecules. Based on Adleman's evolutionary approach, input is encoded by random DNA sequences. Computation is a series of bio-molecular reactions, which involves hybridization, denaturation, ligation, magnetic bead separation, and polymerase chain reaction (PCR). The output of the computation, also in the form of DNA molecules can be read out and visualized by electrophoretical fluorescence operation.

Four years later, in 1998, a length-based DNA computing which is called constant-proportional length-based DNA computing (CPLB-DNAC) for Traveling Salesman Problem (TSP) is proposed by Narayanan and Zorbalas [2]. A constant increase of length of DNA strands is used to encode the actual length of the distances. A drawback of this method is that, there is a possibility of an occurrence of concatenated DNA strands of two distances which could be longer than the DNA strand of the longest distance that has been encoded. This may lead to errors in computing the shortest path [3]. This scheme, however, has not been realized by any laboratory experiment.

In order to solve the shortcoming of CPLB-DNAC, an alternative approach called direct-proportional length-based DNA computing (DPLB-DNAC) is proposed by Ibrahim *et al.* [4] for solving the shortest path problem. In this approach, the cost of an edge is encoded as a direct-proportional length DNA. After an initial pool generation, numerous solution candidates can be generated. By using PCR, it is possible to amplify the optimal combination which represents the solution to the shortest path problem. The output of the computation can be visualized by applying PAGE, where the DNA duplex representing the solution appears as the shortest band of PAGE.

On the other hand, Yamamoto *et al.* [5] presented concentration-controlled DNA computing (CC-DNAC) for accomplishing a local search for the shortest path problem. Although DNA computing with concentration control method enables local search among all the candidate solutions, it cannot guarantee that the most intensive band is the DNA representing the shortest path in the given graph. In addition, it is technically difficult to extract a single optimal solution from the most intensive band [3]. This difficulty, however, has been solved using denaturating gradient gel electrophoresis (DGGE) and constant denaturant gel electrophoresis (CDGE) [5].

Lee *et al.* [6] proposed a DNA computing approach called temperature gradientbased DNA computing (TG-DNAC) for solving TSP. Denaturation temperature gradient polymerase chain reaction (DTG-PCR) has been introduced where DNA duplex of correct solutions will be denatured and amplified by the PCR operation. As the denaturation temperature increases, other DNA strands will be also subsequently amplified. However, the amount of correct solutions will also be exponentially increased, which does affect the final solution.

In this paper, we propose a combination of both schemes of CC-DNAC and DPLB-DNAC to born a hybrid approach, which is called as concentration-controlled direct-proportional length-based DNA computing (CCDPLB-DNAC). The protocol of CCDPLB-DNAC, in fact, is similar as DPLB-DNAC. But the difference is that the amount of poured DNA representing the edges varies closely to the weight of edges. As a result, the concentration of input during initial pool generation will be different as well, and does influence the degree of hybridization. It is found that the proposed

approach offers significant improvements in term of material usage and scalability than that of DPLB-DNAC without concentration-controlled.

In designing and developing the proposed CCDPLB-DNAC, the scopes of the research have been defined. Firstly, the shortest path problem is chosen as a benchmark for this research because the shortest path problem is a kind of problem that involves numerical optimization, even though this problem is not an NP-complete problem. Secondly, a small directed weighted graph, $G = (V, E, \omega)$, which consists of a set of vertices $V = \{ v_1, v_2, v_3, v_4, v_5\}$, a set of edges $E = \{ [v_1,v_2], [v_1,v_3], [v_3,v_4], [v_4,v_5], [v_2,v_5], [v_2,v_3], [v_2,v_4] \}$, and weight, ω, which is assigned to each edges, will be constructed and used an the input of computation as shown in Figure 1. It is clear that the number vertices $|V|$ and the number of edges $|E|$ are 5 and 7, respectively. Even though this graph is small enough for computation but it is big enough if the computation is to be realized by unconventional DNA computing approach, and thus, is a good example in order to show the feasibility of the proposed direct-proportional length-based DNA computing.

2 A Note on Concentration-Controlled Method

Yamamoto *et al*. carried out CC-DNAC for accomplishing a local search for the shortest path problem [7] by avoiding the generation of hopeless solutions. In this research, the vertices and edges are encoded into DNA sequences. During the encoding process, the vertex sequence is synthesized with the same concentration. The relatively different concentration, D_{ij} of the oligonucleotides encoding an edge $i \rightarrow j$ at cost C_{ij} is calculated by using the following formula:

$$D_{ij} = \left(\frac{Min}{C_{ij}} \right)^{\alpha} \tag{1}$$

where *Min* represent the minimum value among the costs of all edges in the graph, and α is set to 2. After all oligonucleotides for nodes with the same concentration, oligonucleotides for edges with different concentration, and complement oligonucleotides are synthesized, an initial pool generation is done in a test tube. During the initial pool generation, the rate of biochemical reactions depends heavily on the reaction rate constants and reactant concentrations. Thus, as the concentration of DNA strands increase, the paths including them can be generated more frequently and the hopeful DNA paths can be generated with high concentration.

Even though the shortest path problem is belonging to the class P, i.e., it is not hard to solve this problem, it is worth to be solved by DNA computing because numerical evaluations are required during the computation [5]. The input to the shortest path problem is a weighted directed graph $G = (V, E, \omega)$, a start node u and an end node v. The output of the shortest path problem is a (u,v) path with the smallest cost. In the case given in Figure 1, if u is V_1 and v is V_5, the cost for the shortest path will be 100 and the optimal path is clearly shown as $V_1 - V_3 - V_4 - V_5$. If the input graph is shown Figure 1 (a), by using equation (1), all the numerical weights are transformed into relative concentrations as shown in Figure 1 (b).

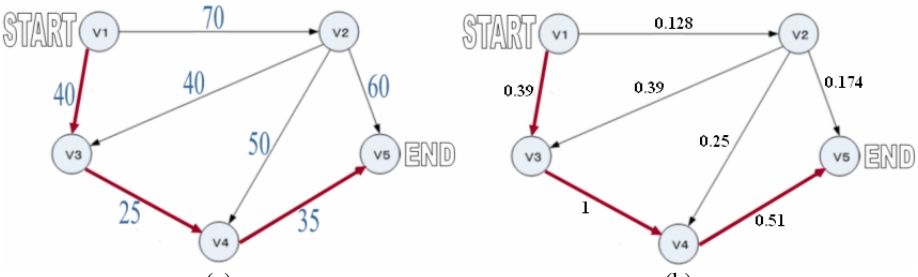

Fig. 1. (a) Example showing a weighted directed graph $G = (V, E)$ with the shortest path shown as V_1 - V_3 - V_4 - V_5 (b) Relative concentration as calculated using equation 1

3 DNA Sequence Design and Synthesis

Let n be the total number of nodes in the graph. The DNA sequences correspond to all nodes and its complements are designed. Let V_i (i= 1, 2, … , n) and $\overline{V_i}$ (i= 1, 2, … , n) be the 20-mer DNA sequences correspond to the ith node in the graph and its complement respectively. By using the available software for DNA sequence design, DNASequenceGenerator [8], the DNA sequences V_i is designed and listed in Table 1. Melting temperature, T_m is calculated based on Sugimoto nearest neighbor thermodynamic parameter [9]. The GC contents (GC%) and melting temperature (T_m) of each sequence are also shown. Table 2, on the other hand, shows the complement of the node sequences.

We introduce three rules to encode each edge in the graph as follows:

(i) If there is a connection between V_1 to V_j, where $j \neq n$, design the oligonucleotide (oligo) for that edge as

$$V_1\,(20) + W_{1j}\,(\omega\text{- }30) + V_j\,(20)$$

(ii) If there is a connection between V_i to V_j, where $i \neq 1$, $j \neq n$, design the oligo for that edge as

$$V_i\,(20) + W_{ij}\,(\omega\text{- }20) + V_j\,(20)$$

(iii) If there is a connection between V_i to V_n, where $i \neq 1$, design the oligo for that edge as

$$V_i\,(20) + W_{in}\,(\omega\text{- }30) + V_n\,(20)$$

where V, W, and '+' denote the DNA sequences for nodes, DNA sequences for weight, and 'join' respectively. The designed oligos consist of three segments; two node segments and an edge segment. 'ω' denotes the weight value for corresponding DNA sequences for weight W_{ij}, where W_{ij} denotes the DNA sequences representing a cost between node V_i and V_j. The value in parenthesis indicates the number of DNA bases or nucleotides for each segment. Table 3 lists all the oligos based on the proposed rules, where the node segments and edge segments are distinguished by capital and small letters respectively. Again, DNASequenceGenerator [8] is employed. At the end of this stage, the oligos of the complement of nodes and edges are synthesized.

Table 1. DNA sequences for nodes

Node, V_i	20-mer Sequences (5'-3')	GC%	Melting Temperature, T_m (ºC)
V_1	AAAGCTCGTCGTTTAGGAGC	50	60.9
V_2	GCACTAGGGATTTGGAGGTT	50	60.3
V_3	GCTATGCCGTAGTAGAGCGA	55	60.5
V_4	CGATACCGAACTGATAAGCG	50	60.6
V_5	CGTGGGTGGCTCTGTAATAG	55	60.5

Table 2. Complement of node

Complement Node, $\overline{V_i}$	20-mer Complement Sequences (3'-5')
$\overline{V_1}$	TTTCGAGCAGCAAATCCTCG
$\overline{V_2}$	CGTGATCCCTAAACCTCCAA
$\overline{V_3}$	CGATACGGCATCATCTCGCT
$\overline{V_4}$	GCTATGGCTTGACTATTCGC
$\overline{V_5}$	GCACCCACCGAGACATTATC

Table 3. DNA sequences for edges

Edge	DNA Sequences
V_4–W_{45}–V_5	5'-CGATACCGAACTGATAAGCG ccaagCGTGGGTGGCTCTGTAATAG-3'
V_3–W_{34}–V_4	5'-GCTATGCCGTAGTAGAGCGA ccgtcCGATACCGAACTGATAAGCG-3'
V_1–W_{13}–V_3	5'-AAAGCTCGTCGTTTAGGAGCacgtcggttc GCTATGCCGTAGTAGAGCGA-3'
V_2–W_{23}–V_3	5'-GCACTAGGGATTTGGAGGTT ccgtcttttacccaagtaatGCTATGCCGTAGTAGAGCGA-3'
V_2–W_{24}–V_4	5'-GCACTAGGGATTTGGAGGTT acgtgttttaaggaagtacggtaagctgcg CGATACCGAACTGATAAGCG-3'
V_2–W_{25}–V_5	5'-GCACTAGGGATTTGGAGGTT gcgtcgcgtaaggcagtaccggactctgcc CGTGGGTGGCTCTGTAATAG-3'
V_1–W_{12}–V_2	5'-AAAGCTCGTCGTTTAGGAGC cggtggtttaacgaagtcctgtactatgggttatttgcag GCACTAGGGATTTGGAGGTT-3'

4 Concentration-Controlled Direct-Proportional Length-Based DNA Computing for the Shortest Path Problem

Currently, there are two kinds of initial pool generation methods for solving weighted graph problem: hybridization/ligation and parallel overlap assembly (POA). The hybridization/ligation method has been firstly introduced by Adleman [1] to solve HPP. For hybridization/ligation method, during the operation, the link oligos hybridize through the hydrogen bonds by enzymatic reaction. The hybridization/ligation reaction is well shown in Figure 2 [10].

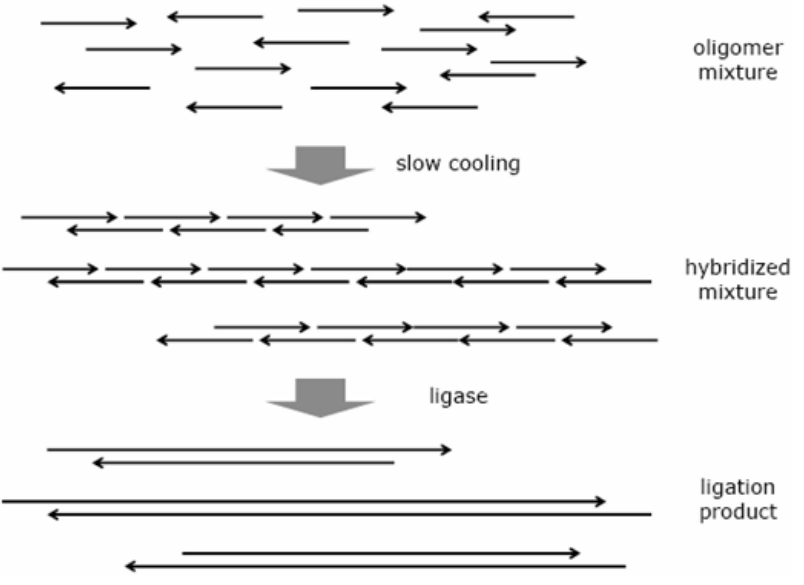

Fig. 2. Hybridization/ligation method for initial pool generation. The arrowhead indicates the 3' end.

POA has been used [11] and broadly applied in gene construction [12-14], gene reconstruction [15], and DNA shuffling [16]. POA involves thermal cycle and during the thermal cycle, the position strings in one oligo anneals to the complementary strings of the next oligo. The 3' end side of the oligo is extended in the presence of polymerase enzyme to form a longer dsDNA. One cycle of parallel overlap assembly is depicted in Figure 3 [10]. After a number of thermal cycles, a data pool with all combinations could be built.

Lee *et al.* [10] did a comparison between hybridization/ligation method and POA for initial pool generation of DNA computing. They came out with a conclusion that for the initial pool generation of weighted graph problems, POA method is more efficient than that of hybridization/ligation method. According to [10], the advantages of POA over hybridization/ligation method for initial pool generation are as follows:

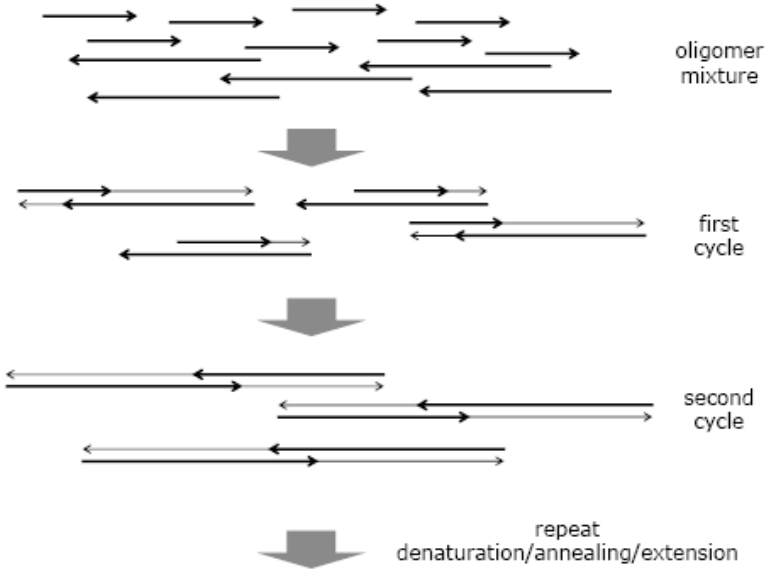

Fig. 3. Parallel overlap assembly for initial pool generation. The thick arrows represent the synthesized oligos which are the input to the computation. The thin arrows represent the elongated part during polymerization. The arrowhead indicates the 3' end.

(i) The initial pool size generated from the same amount of initial oligos is about twice larger than that of hybridization/ligation method. Though, if a larger problem is considered, the initial pool size is too small to contain the complete pool. POA, however, with more cycle and large experimental scale could include the practical pools.
(ii) Initially, two single-stranded DNA molecules partially hybridize in the annealing step and then they are extended by polymerase. The elongated DNA molecules are denatured to two single-stranded DNA in the next denaturation step, and they are subjected to the annealing reaction at the next cycle. Therefore, POA does maintain the population size and the population size can be decided by varying the initial number of oligos.
(iii) In hybridization/ligation method, the population size decreases as reaction progress. The population size decreased by a factor of the number of components composing it in hybridization/ligation method. As the problem size increases, the required initial pool size increases dramatically. Moreover, initial pool generation by POA requires fewer strands than hybridization/ligation method to obtain similar amount of initial pool DNA molecules because complementary strands are automatically extended by polymerase.
(iv) POA does not require phosphorylation of oligos which is prerequisite for the ligation of oligos.
(v) POA demands less time than hybridization/ligation method. Hybridization requires one and half hour while ligation required more than 12 hours. Hence, POA for 34 cycles requires only two hours. Therefore, POA is much more efficient and economic method for initial pool generation.

As stated in [3], "*In addition, the fact that larger weights are encoded as longer sequences is contrary to the biological fact that; the longer the sequences are, the more likely they hybridize with other DNA strands, though we have to find the shortest DNA strands*". From the biological point of view, this argument is definitely true. In order to overcome the limitation of general length-based DNA computing, the authors discovered that by utilizing POA for initial pool generation, a phase where numerous combinations of random routes of the graph are generated in the solution, a shortcoming, which is the biological influence contributed by the length of the oligos could be eliminated.

In order to generate the initial pool of the direct-proportional length-based DNA computing for the example problem by using POA method, the input to the computation are all the synthesized oligos as listed in Table 3 and the complement sequences for each nodes, which are listed in Table 2. These inputs are poured into a test tube and the cycles begin. In fact, the operation of POA is similar as polymerase chain reaction (PCR) but the difference is that POA operates without the use of primers. As PCR, one cycle consists of three steps: denaturation, hybridization, and extension.

At this stage, an initial pool of solution has been produced and it is time to filter out the optimal combinations among the vast alternative combinations of the problem. Unlike conventional filtering, this process is not merely throwing away the unwanted DNA duplex but rather copying the target dsDNA exponentially by using the incredibly sensitive PCR process. This can be done by amplifying the DNA duplex that contain the start node V_1 and end node V_5 using primers. After the PCR operation is accomplished, there should be numerous number of DNA strands representing the start node V_1 and end node V_5 traveling through a possible number of nodes. The output solution of the PCR then undergoes gel electrophoresis operation. During this operation, the dsDNA $V_1 - V_3 - V_4 - V_5$, which representing the shortest path starting from V_1 and ending at V_5 can be visualized by the shortest band of gel electrophoresis with higher intensity than the other dsDNA, if any.

At this moment, based on the shortest length DNA duplex, one only knows that the shortest path begins from V_1 and ends at V_5. However, the information does not contain the nodes that passed through the shortest path. The information regarding all the nodes in the shortest path as well as their sequence can be obtained by applying graduated PCR operation. For the sake of explanation, the DNA molecules representing the answer of the shortest path $V_1 - V_3 - V_4 - V_5$ is taken again for instance. After the shortest band DNA is extracted from the gel, graduated PCR is performed by running four different PCR operations to the solution containing DNA duplex $V_1 - V_3 - V_4 - V_5$ separately. The pair of primers used for every PCR reaction are $V_1/\overline{V_2}$, $V_1/\overline{V_3}$, $V_1/\overline{V_4}$, and $V_1/\overline{V_5}$. It is expected that for the final solution containing the strand $V_1 - V_3 - V_4 - V_5$, 100 base-pairs (bp), graduated PCR will produce bands of x, 50, 75, and 100 in successive lanes of a gel as depicted in Figure 4. The symbol x denotes the absence of a band corresponding to the omission of nodes V_2 along the DNA duplex. This means that there are intermediate nodes, V_3 and V_4 in between the start node V_1 and the end node V_5. Therefore, the shortest path of the graph can be readout as $V_1 \rightarrow V_3 \rightarrow V_4 \rightarrow V_5$.

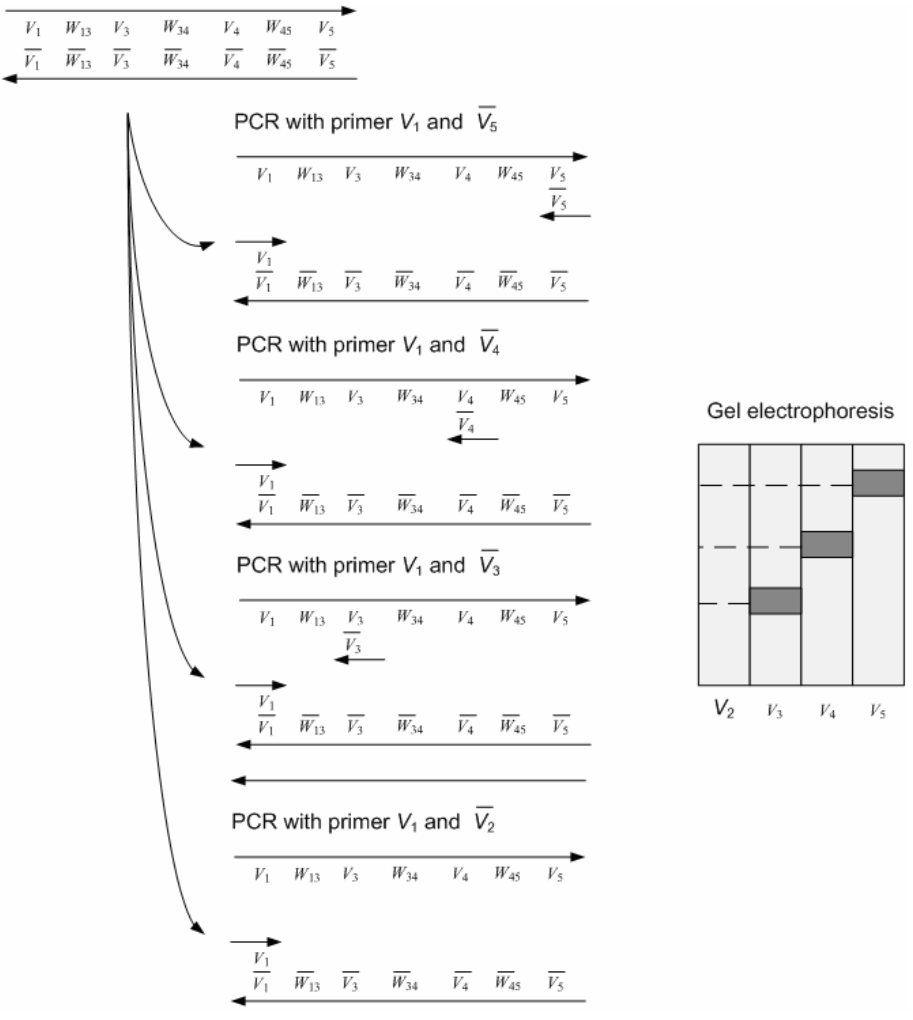

Fig. 4. Examples showing the results of the graduated PCR

6 Experimental Protocols, Results, and Discussions

Firstly, the relative concentration of edges oligos are translated into actually amount of DNA as listed in Table 4. Based on Table 4, the initial pool generation by POA is performed in a 25 µ*l* solution containing 7.842 µ*l* oligos (Proligo Primers & Probes, USA), 2.5 µ*l* dNTP (TOYOBO, Japan), 2.5 µ*l* 10x KOD dash buffer (TOYOBO, Japan), 0.125 µ*l* KOD dash (TOYOBO, Japan), and 12.033 µ*l* double distilled water (ddH_20) (Maxim Biotech). The reaction consists of 25 cycles and for each cycles, the appropriate temperatures and time are as follows:

- 94ºC for 30s
- 55ºC for 30s
- 74ºC for 10s

The product of parallel overlap assembly is shown in Figure 5. In order to select the paths that begin at V_1 and end at V_5, DNA amplification is done by PCR. PCR is performed in a 25 μl solution consists of 2.5 μl for each primers, 1 μl template, 2.5 μl dNTP (TOYOBO, Japan), 2.5 μl 10x KOD dash buffer (TOYOBO, Japan), 0.125 μl KOD dash (TOYOBO, Japan), and 13.875 μl ddH_20 (Maxim Biotech). The reaction consists of 25 cycles as follows:

- 94ºC for 30s
- 55ºC for 30s
- 74ºC for 10s

which is the same as POA. The sequences used as primers are AAAGCTCGTCGTTTAGGAGC (V_1) and GCACCCACCGAGACATTATC ($\overline{V_5}$).

In order to visualize the result of the computation, the product of PCR is subjected to PAGE for 90 minutes at 200V. After that, the gel is stained by SYBR Gold (Molecular Probes) and the gel image is captured. Figure 5 shows the output of the computation of DPLB-DNAC without concentration-controlled. Figure 6 on the other hand, shows the output of CCDPLB-DNAC.

Table 4. Actual amount of each edges oligos

Edge	Without Concentration-Controlled	With Concentration-Controlled
V_4–V_5	1 μl	0.51 μl
V_3–V_4	1 μl	1 μl
V_1–V_3	1 μl	0.39 μl
V_2–V_3	1 μl	0.39 μl
V_2–V_4	1 μl	0.25 μl
V_2–V_5	1 μl	0.174 μl
V_1–V_2	1 μl	0.128 μl
Total	7 μl	2.842 μl

According to the gel image of Figure 5, it is clear that without the concentration-controlled, DPLB-DNAC is able to produce several shortest paths during the computation. In this case, based on the output in lane 2, up to four shortest paths is generated *in vitro* and visualized by gel electrophoresis. In contrast, if concentration-controlled is applied, lane 2 of Figure 6 consists of only one band containing DNA duplex of the shortest path $V_1 - V_3 - V_4 - V_5$ (100bp), which survived after POA and PCR. As expected, it is likely that the concentration of dsDNAs other than dsDNAs representing the answer of the shortest path problem tends to be small. Even though only one band is shown, it is more than enough since the band exactly represent the answer of the shortest path problem.

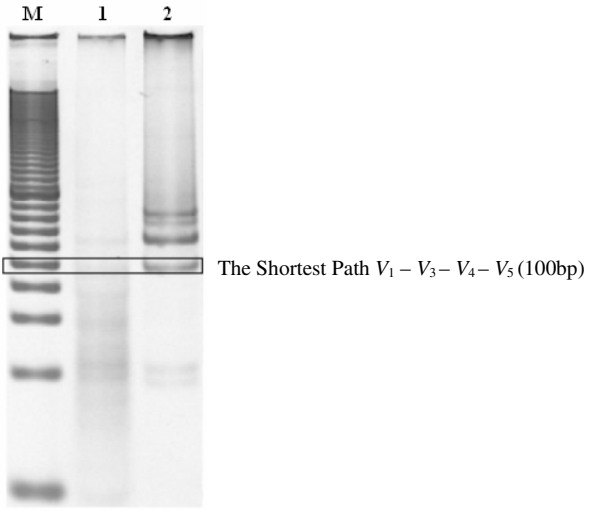

Fig. 5. Experimental results of gel electrophoresis on 10% PAGE in the case of DPLB-DNAC. Lane M denotes 20-bp ladder, lane 1 is the product of POA, and lane 2 is the product of PCR.

For graduated PCR, 4 identical DNA mixtures, which are the product of PCR of CCDPLB-DNAC are subjected to PAGE for 40 minutes at 200V. After that, the gel is stained by SYBR Gold (Molecular Probes, USA). Quantum Prep™ Freeze 'N Squeeze DNA Gel Extraction Spin Columns (Bio-Rad, Japan) is used during the DNA extraction from the polyacrylamide gel. By using a clean razor blade, the band of interest, which is the shortest band, is carefully excised from the gel. The gel slice is chopped and placed into the filter cup of the Quantum Prep Freeze 'N Squeeze DNA Extraction Spin Column. Then, the filter cup is placed into a dolphin tube. The Quantum Prep Freeze 'N Squeeze DNA Extraction Spin Column is placed in a -20ºC freezer for 5 minutes and the sample is spun at 13,000 x g for 3 minutes at room temperature. The purified DNA is collected from the collection tube and ready for PCR.

After the DNA extraction from the polyacrylamide gel, four different PCR, namely PCR1, PCR2, PCR3, and PCR4, is run to the purified solutions. The pair of primers used for every PCR is listed in Table 5.

Each PCR is performed in a 25 µ*l* solution consists of 2.5 µ*l* for each primers, 1 µ*l* template, 2.5 µ*l* dNTP (TOYOBO, Japan), 2.5 µ*l* 10x KOD dash buffer (TOYOBO, Japan), 0.125 µ*l* KOD dash (TOYOBO, Japan), and 13.875 µ*l* double-distilled water (ddH_20) (Maxim Biotech, Inc, Japan). The reaction consists of 25 cycles and for each cycles, the appropriate temperature are as follow:

- 94ºC for 30s
- 55ºC for 30s
- 74ºC for 10s

Again, the product of graduated PCR is subjected to PAGE for 40 minutes at 200V and the gel is stained by SYBR Gold (Molecular Probes, USA). Finally, the gel image is captured. Figure 7 shows the gel image of the product of graduated PCR. Four

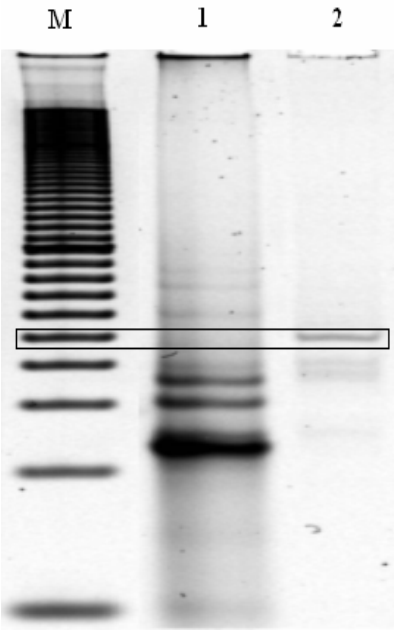

Fig. 6. Experimental results of gel electrophoresis on 10% PAGE in the case of CCDPLB-DNAC. Lane M denotes 20-bp ladder, lane 1 is the product of POA, and lane 2 is the product of PCR.

Table 5. Four set of primers used for the graduated PCR

Name	Forward Primers	Reverse Primers
PCR1	AAAGCTCGTCGTTTAGGAGC	CGTGATCCCTAAACCTCCAA
PCR2	AAAGCTCGTCGTTTAGGAGC	CGATACGGCATCATCTCGCT
PCR3	AAAGCTCGTCGTTTAGGAGC	GCTATGGCTTGACTATTCGC
PCR4	AAAGCTCGTCGTTTAGGAGC	GCACCCACCGAGACATTATC

bands of x, 50, 75, and 100 base pairs (bp) in successive lanes of the gel are successfully produced and therefore, as expected, the shortest path of the graph can be read-out as $V_1 \rightarrow V_3 \rightarrow V_4 \rightarrow V_5$.

Two significant benefits of graduated PCR for DPLB-DNAC have been identified. The first is due to its capability to show and visualize the detail output of the shortest path computation based on DPLB-DNAC. The other benefit is that at the same time, the correctness of DPLB-DNAC for the computation of the shortest path problem can be proved. Hence, the authors found that graduated PCR should be essentially incorporated in the DPLB-DNAC in order to improve the overall performance of DPLB-DNAC.

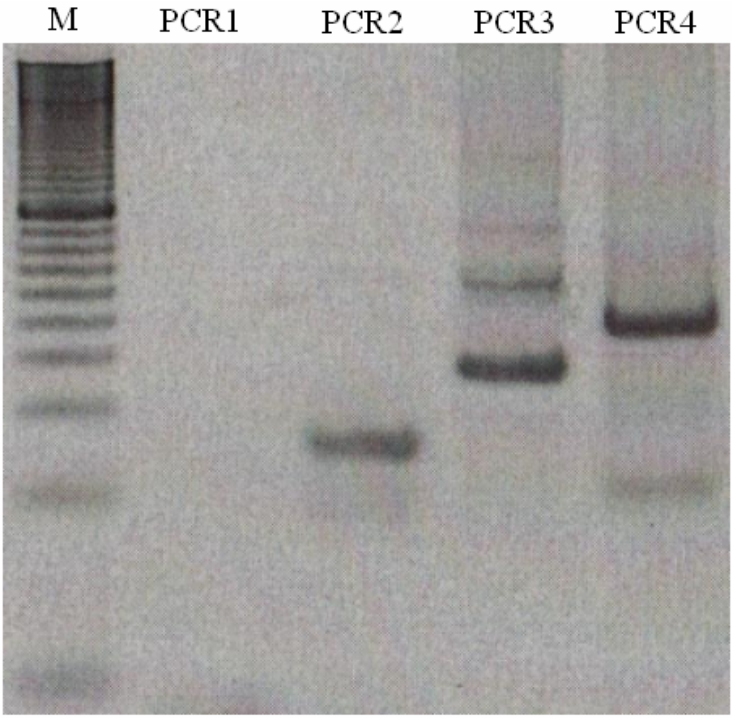

Fig. 7. Experimental results of gel electrophoresis on 10% polyacrylamide gel. Lane M denotes 20-bp ladder.

Scaling is certainly the main problem of DNA computing, especially for generate-and-test DNA computing, as the proposed approach. In order to extend the proposed approach to a larger problem, two issues should be considered: molecular's weight and the capability to select the final solution. As an example, if Adleman's work for solving HPP is further examined, a 70-node problem requires 10^{25} kg of nucleotides, and this is quite a lot for a small test tube [17]. Hence, an advanced high reaction facility, such as microreactor [18], is highly important. In this research, we showed how the scalability of our approach can be improved linearly, in two steps. The first step is during initial pool generation, where POA is employed instead of hybridization/ligation. As previously discussed, POA is able to generate two times bigger initial pool, in term of size, than that of hybridization/ligation. The second step is during the computation, in term of the amount of DNA used for the computation. In our case, we used 2.842 μl edge oligos in the case of concentration-controlled. It is about half of the amount of edge oligos if the computation is done without concentration-controlled, where 7 μl should be used.

In DNA computing for weighted graph problems, after the *in vitro* computation, a subsequent reactions or bio-molecular operations should be employed in order to detect the final solution. As an example, for the CC-DNAC method, separation as DGGE and CDGE should be used, whereas in TG-DNAC method, DTG-PCR

separation should be utilized. This operations are relatively complicated than normal gel electrophoresis. In our approach, the adopted protocol for detecting the final solution is simple, where PAGE is more already enough to visualize the result of the computation.

However, one limitation of the proposed approach is that, the minimum weight of edges that can be encoded is limited and the weight falls in a very narrow range. This is mainly because the length of the solution is not only proportional to the length of the path it encodes but also the number of vertices in the path. Hence, the lower bound, in term of minimum weight that can be encoded by the proposed approach is achieved when:

$$\omega - \frac{3}{2}\beta = 0 \tag{2}$$

Hence, the minimum weight, which can be encoded by oligos, $\omega_{\min}$ is attained as:

$$\omega_{\min} = \frac{3}{2}\beta \tag{3}$$

where β is the number of DNA bases used to represent the node sequences [19].

7 Conclusions

Based on massive parallelism inherent in DNA computing, many researchers have tried to solve various NP-complete problems. These are mathematical problems which have exponential complexity and no efficient solution has been found yet. Even though the shortest path problem is not belonging to the class of NP problems, it is important to solve them since this kind of problems occur frequently in many real world problems. Thus, in this paper, we have presented an improve hybrid approach called 'concentration-controlled direct-proportional length-based DNA computing' to solve weighted graph problems using molecular computing. Based on this approach, both length and concentration are used as input and output data is recognized by length only. For the sake of initial pool generation, two kinds of methods are reviewed: hybridization/ligation and POA. Since POA offers several advantages in term of materials usage and reaction time, for a successful demonstration of CCDPLB-DNAC, we found that POA for initial pool generation is critically important. Further, by varying the amount of input DNA, less DNA is used for computation, which further offers the advantage in term of material usage. Since less amount of DNA can be used to generate the combination representing the answer of the problem, indirectly, this will advances the performance of the proposed approach from scalability point of view. Also, in this paper, we have presented graduated PCR, as an extended operation of CCDPLB-DNAC. Based on the proposed approach, the product of PCR of CCDPLB-DNAC is subjected to DNA extraction from polyacrylamide gel, PCR, and PAGE. As supported by the experimental results, graduated PCR is able to visualize detail additional information of the shortest path,

such as the intermediate vertices and the order of these vertices in the shortest path. Finally, it is expected that the proposed approach, would extend the applicability of DNA computing for solving intractable weighted graph problems.

Acknowledgements

This research was supported partly by the IEEE Computational Intelligence Society (CIS) Walter J Karplus Student Summer Research Grant 2004 for a research visit in September 2004 at the DNA Computing Laboratory, Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Hokkaido, Japan. The first author would like to thank Masahito Yamamoto for discussions that led to improvements in this work and also the permission to practice various kinds of biochemical experiments in the laboratory. Also, the first author is sincerely grateful to Atsushi Kameda, Satoshi Kashiwamura, and members of DNA Computing Laboratory of Hokkaido University for fruitful explanations and kind assistance during the practice of biochemical experiments. Lastly, the first author is very thankful to Universiti Teknologi Malaysia (UTM) for granting a study leave in Meiji University under SLAB-JPA scholarship.

References

1. Adleman, L.: Molecular Computation of Solutions to Combinatorial Problems," Science, Vol. 266 (1994) 1021-1024
2. Narayanan, A., Zorbalas, S.: DNA Algorithms for Computing Shortest Paths," in Proceedings of Genetic Programming (1998) pp. 718-723
3. Lee, J.Y., Shin, S.Y., Augh, S.J., Park, T.H., Zhang, B.T.: Temperature Gradient-Based DNA Computing for Graph Problems with Weighted Edges, Lecture Notes in Computer Science, Vol. 2568 (2003) 73-84
4. Ibrahim, Z., Tsuboi, Y., Ono, O., Khalid, M.: Direct-Proportional Length-Based DNA Computing for Shortest Path Problem", International Journal of Computer Science and Applications (IJCSA), Technomathematics Research Foundation, Vol. 1 (2004) 46-60
5. Yamamoto, M., Kameda, A., Matsuura, N., Shiba, T., Kawazoe, Y., Ahochi, A.: A Separation Method for DNA Computing Based on Concentration Control, New Generation Computing, Vol. 20 (2002) 251-262
6. Lee, J.Y., Shin, S.Y., Augh, S.J., Park, T.H., Zhang, B.T.: Temperature Gradient-Based DNA Computing for Graph Problems with Weighted Edges," in Preliminary Proceedings of the Eighth International Meeting on DNA Based Computers (2002) 41-50
7. Yamamoto, M., Matsuura, N., Shiba, T., Ohuchi, A.: DNA Solution of the Shortest Path Problem by Concentration Control. Genome Informatics (2000) 466-467
8. Udo, F., Sam, S., Wolfgang, B., Hilmar, R.: DNA Sequence Generator: A Program for the Construction of DNA Sequences," In Proceedings of the Seventh International Workshop on DNA Based Computers (2001) 23-32
9. Sugimoto, N., Nakano, S., Yoneyama, M., Honda, K.: Improved Thermodynamic Parameters and Helix Initiation Factor to Predict Stability of DNA Duplexes, Nucleic Acid Research, Vol. 24 (1996) 4501-4505

10. Lee, J.Y., Lim, H.W., Yoo, S.I., Zhang, B.T., Park, T.H.: Efficient Initial Pool Generation for Weighted Graph Problems using Parallel Overlap Assembly, in Preliminary Proceedings of the Tenth International Meeting on DNA Based Computers (2004) 357-364
11. Kaplan, P.D., Ouyang, Q., Thaler, D.S., Libchaber, A.: Parallel Overlap Assembly for the Construction of Computational DNA Libraries, Journal of Theoretical Biology, Vol. 188, Issue 3 (1997) 333-341
12. Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., Pease, L.R.: Site-Directed Mutagenesis by Overlap Extension using the Polymerase Chain Reaction, Gene, Vol. 77 (1989) 51-59
13. Jayaraman, K., Fingar, S.A., Fyles, J.: Polymerase Chain Reaction-Mediated Gene Synthesis: Synthesis of a Gene Coding for Isozymec of Horseradish Peroxidase, Proc. Natl. Acad. Sci. U.S.A., Vol. 88 (1991) 4084-4088
14. Stemmer, W.P., Crameri, A., Ha, K.D., Brennan, T.M., Heyneker, H.L.: Single-Step Assembly of a Gene and Entire Plasmid from Large Numbers of Oligodeoxyribonucleotides, Gene, Vol. 164 (1995) 49-53
15. DeSalle, R., Barcia, M., Wray, C.: PCR Jumping in Clones of 30-million-year-old DNA Fragments from Amber Preserved Termites, Experientia, Vol. 49 (1993) 906-909
16. Stemmer, W.P.: DNA Shuffling by Random Fragmentation and Reassembly: In Vitro Recombination for Molecular Evolution," Proc. Natl. Acad. Sci. U.S.A., Vol. 91 (1994) 10747
17. Zucca, M.: DNA Based Computational Models: Ph.D. Thesis, Politecnico di Torino, Italy (2000)
18. Noort, D., Gast, F.U., McCaskill, J.S.: DNA Computing in Microreactors, Lecture Notes in Computer Science, Vol. 2340 (2001) 33-45
19. Ibrahim, Z., Tsuboi, T., Ono, O., and Khalid, M.: A Study on Lower Bound of Direct-Proportional Length-Based DNA Computing for Shortest Path Problem, Lecture Notes in Computer Science, Vol. 3314 (2004) 71-76