New Generation Computing, 20(2002)251-261
Ohmsha, Ltd. and Springer-Verlag

A Separation Method for DNA Computing Based on Concentration Control

Masahito YAMAMOTO
Japan Science and Technology Corporation (JST)
Honmachi 4-1-8, Kawaguchi 332-0012, Japan
and Hokkaido University
North 13, West 8, Kita-ku, Sapporo 060-8628, JAPAN

Atsushi KAMEDA
Japan Science and Technology Corporation (JST)
Honmachi 4-1-8, Kawaguchi 332-0012, JAPAN

Nobuo MATSUURA, Toshikazu SHIBA, Yumi KAWAZOE and Azuma OHUCHI
Hokkaido University
North 13, West 8, Kita-ku, Sapporo 060-8628, JAPAN
masahito@dna-comp.org

Received 14 September 2001
Revised manuscript received 15 February 2002

Abstract A separation method for DNA computing based on concentration control is presented. The concentration control method was earlier developed and has enabled us to use DNA concentrations as input data and as filters to extract target DNA. We have also applied the method to the shortest path problems, and have shown the potential of concentration control to solve large-scale combinatorial optimization problems. However, it is still quite difficult to separate different DNA with the same length and to quantify individual DNA concentrations. To overcome these difficulties, we use DGGE and CDGE in this paper. We demonstrate that the proposed method enables us to separate different DNA with the same length efficiently, and we actually solve an instance of the shortest path problems.

Keywords: DNA Computing, Concentration Control, Shortest Path Problem, DGGE, CDGE.

§1 Introduction

DNA computing has been remarkable as a new computing paradigm, since

Adleman and Lipton first showed that an instance of NP-complete problems can be solved by a series of biochemical operations.[1,2] Based on the Adleman-Lipton paradigm, various DNA computing algorithms have been proposed for solving combinatorial problems.[3,4] In order to extend the applicable class of DNA computing, we have already proposed a concentration control method. This method enables us to use DNA concentrations as input data and as filters to extract target DNA. We have also presented an algorithm for solving the shortest path problems.[5] The algorithm encodes each vertex and edge into DNA similar to Adleman's experiment. Then, the input DNA with different concentrations, calculated according to the costs on the edges, is added into the pool, and a hybridization process is performed. After the hybridization process, we separate each DNA path in order to quantify the DNA. In the separation, we are unable to use the polyacrylamide gel electrophoresis (PAGE) that is commonly used for DNA computing, because there are generally a lot of paths with the same length. As previously reported, it is difficult to separate each DNA path from the resultant pool of the PCR process because doing so might contaminate other paths.[5]

In this paper, we present a new algorithm for solving the shortest path problems based on concentration control by using DGGE (Denaturing Gradient Gel Electrophoresis) and CDGE (Constant Denaturant Gel Electrophoresis) as a separation method. The separation method enables us to not only separate all DNA paths and quantify each DNA path, but also to reduce the total computational time. To verify this, we show that an instance of the shortest path problems that is actually solved.

§2 Algorithm for Shortest Path Problems

In this section, we describe an algorithm for solving the shortest path problems based on concentration control. Figure 1 shows the shortest path problem used in this paper. The shortest path problem for the graph is the problem to find the shortest path (minimizing the total costs including the path) from vertex 0 (start) to vertex 5 (goal). In this graph, the shortest path is the path 0-2-3-5, and the cost of the path is 18.

The shortest path problems belong to the class P, i.e., it is not hard to solve them. However it is worth solving them by DNA computing, because numerical evaluations are required in order to solve them. The proposed concentration control enables us to implement the numerical evaluations.

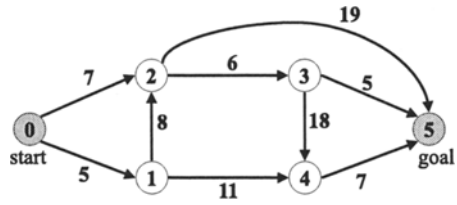


Fig. 1 A Directed Graph with Costs on Edges.

The algorithm for solving the shortest path problems is briefly described in the following.

Step 1: Encode the problem into DNA. Each vertex i in the graph is associated with a designed 20-mer sequence of DNA denoted O_i. For each edge $i \rightarrow j$ in the graph, an oligonucleotide $O_{i \rightarrow j}$ that is 3′ 10-mer of O_i followed by 5' 10-mer of O_j is created.

Step 2: Construct a lot of paths. To construct many paths in the graph shown in Fig. 1, a mixture containing each complementary oligonucleotide encoding vertices and each oligonucleotide encoding edges is made. The concentrations of the complementary oligonucleotides encoding vertices are set to the same value, and the relative concentration D_{ij} of the oligonucleotides encoding an edge $i \rightarrow j$ at cost C_{ij} is calculated by the following formula:

$$D_{ij} = (Min/C_{ij})^{\alpha}, \tag{1}$$

where *Min* represents the minimum value among the costs of all edges in the graph, and α is a parameter value. In this paper, we set the value of α to 2 on the basis of results of preliminary experiments. The hybridization process is performed for each mixture obtained in this way.

Step 3: Amplify the DNA paths by PCR. In order to select the DNA paths that begin with O_0 (start) and end with O_5 (goal), DNA amplification is performed by polymerase chain reaction (PCR). Two primers GC+O_0 and $\overline{O_5}$ are used. GC+O_0 is added 30 bp of GC clamp sequence at the 5′ end side of the O_0 sequence.

Step 4: Determine the oligonucleotide encoding the shortest path. The amplified DNA paths are separated by using DGGE and CDGE. DGGE is an electrophoretic method to identify a single change or some base changes in a segment of DNA.[6] As a result of the separation, the most intensive band is extracted and analyzed by sequencing. In a denaturing gradient acrylamide gel, double-stranded DNA is subjected to an increasing denaturant environment and melt in discrete segments called "melting domains." The melting temperature (T_m) of these domains is sequence-specific. When the T_m of the lowest melting domain is reached, the DNA becomes partially melted, creating branched structures. The partial melting domain of the DNA reduces its mobility in a polyacrylamide gel. Since the T_m of a particular melting domain is sequence-specific, a difference of base sequences alters the melting profile of that specific DNA when compared to the others. CDGE is a modification of DGGE. The optimal concentration of a denaturant used for CDGE is determined from the maximum split between specific DNA. CDGE can be used to rapidly screen samples for the quantification of each specific DNA.

§3 Simulation Model

We have developed a simulation model of the hybridization process.[7] By using this simulation model, we can estimate the concentrations of the resultant DNA paths from the concentration of each input DNA oligonucleotide. The following differential equation represents the rate of change of the concentration

of DNA complex x_i at time t.

$$\frac{d[x_i]}{dt} = \sum_{(j,k)\in M_i} (Va_{jk}[x_j][x_k] - Vd_i[x_i]) - \sum_{j\in B_i} (Va_{ij}[x_i][x_j] - Vd_m[x_i * x_j]) \quad (2)$$
$$(x_i * x_j \rightarrow x_m),$$

where $[x_i]$ represents the concentration of DNA complex x_i at time t and $x_i * x_j \rightarrow x_m$ means that DNA complex x_m can be generated by DNA complexes x_i and x_j. Va_{ij} is the reaction rate constant of annealing between two DNA complexes x_i and x_j, and Vd_i is that of denaturing DNA complex x_i. M_i is a set of pairs of indices of DNA complexes that can generate x_i by annealing with only one DNA complex, and B_i is a set of indices of DNA complexes that can generate x_i by denaturing into only two DNA complexes.

Our hybridization simulator model takes the initial concentration of each DNA oligonucleotide and computes the concentrations of the resultant DNA paths, but the influence of the lengths of the DNA complexes is not considered. The computation of the above form is performed by the 4-dimensional Runge-Kutta method. In this paper, for simplicity, we set the reaction rate constants all equal, i.e., $Va_{ij} = 0.001$ (constant) and $Vd_i = 0.0$ (constant).

§4 Experimental Results and Discussions

4.1 Experimental Conditions

In this subsection, we describe the experimental conditions in detail. The sequences of oligonucleotides encoded in step 1 are presented as shown in Fig. 1. The methods of step 2 are omitted because these are almost the same as those of the previous report.[5)]

Step 3: Amplify the DNA paths by PCR. In order to select the DNA paths that begin with O_0 (start) and end with O_5 (goal), DNA amplification is performed by using PCR. PCR is performed in a $50\mu l$ solution containing a $1\mu l$ ligation mixture, 2 mM dNTPs, two primers, and 1.25 units of KOD Dash DNA polymerase (TOYOBO Co., Japan). The appropriate PCR cycles are as follows:

- initial incubation at 94°C for 20 sec.
- 94°C for 30 sec.
- 50°C for 5 sec.
- 74°C for 30 sec.

The sequences of the two primers GC+O_0 and $\overline{O_5}$ are GCCGGGGCGGG-GCCGGGCGGGCGGCCCGGGTGGTTCCGATGTTTAAGCAA and TCTG-GCTGAGTCAAGTGATC, respectively. The GC clamp is effective in the next separation process.

Step 4: Determine the oligonucleotide encoding the shortest path. The concentrations of DNA paths are generated by performing perpendicular DGGE (20-55% denaturant, 12% acrylamide gel) and CDGE (43% denaturant, 12% acrylamide gel) and using an image analyzing system, and Chemi Doc and the

Table 1 Designed Sequences of Vertices and Edges for the Graph Shown in Fig. 1.

Name	Sequences ($5' \rightarrow 3'$)
$\overline{O_0}$	TGGTTCCGATGTTTAAGCAA
$\overline{O_1}$	AGTCCTATTCATAACGACGC
$\overline{O_2}$	CCACAAGAGGGTATGGGCGG
$\overline{O_3}$	TACGAGCACACGAAGGTAGT
$\overline{O_4}$	GTTCTAGCAGTCTCCAAAAG
$\overline{O_5}$	GATCACTTGACTCAGCCAGA
O_{01}	GAATAGGACTTTGCTTAAAC
O_{02}	CCTCTTGTGGTTGCTTAAAC
O_{12}	CCTCTTGTGGGCGTCGTTAT
O_{14}	CTGCTAGAACGCGTCGTTAT
O_{23}	TGTGCTCGTACCGCCCATAC
O_{25}	TCAAGTGATCCCGCCCATAC
O_{34}	CTGCTAGAACACTACCTTCG
O_{35}	TCAAGTGATCACTACCTTCG
O_{45}	TCAAGTGATCCTTTTGGAGA

software of Quantity One. The 100% denaturant consists of 7 M urea and 40% formamide. After electrophoresis, the gel is stained by SYBR Gold (Molecular Probes, Inc., USA)

4.2 Results and Discussions

Figure 2 shows visualized DNA paths amplified by PCR, captured by Chemi Doc. All of these DNA paths were added 30 bp GC clamp at the $5'$ end side because GC+O_0 primer was used for amplification by PCR. This GC clamp was necessary in DGGE to attain the optimal resolution. DNA bands of 90, 110, and 130 bp were detected from lane 1. These bands were 60, 80, and 100 bp DNA paths.

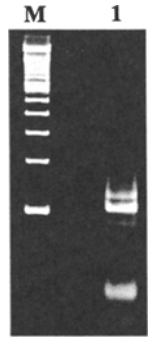


Fig. 2 Confirmation of Formation DNA Paths with 5% PAGE. Lane M, DNA Size Marker (100 bp ladder); Lane 1, Amplified DNA Paths after 14 Cycles of PCR.

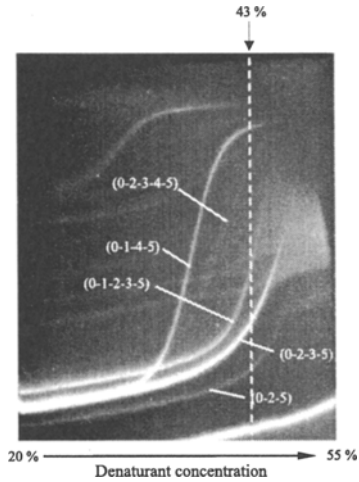


Fig. 3 Amplified DNA Paths after 14 Cycles of PCR Separated by Perpendicular DGGE (20–55% Denaturant) Run at 130 V for 2 Hours at 60°C in 1x TAE Buffer.

In order to determine the optimal denaturant concentration to separate all DNA paths, we performed perpendicular DGGE. Figure 3 shows visualized amplified DNA paths separated by perpendicular DGGE (20-55% denaturant), captured by Chemi Doc. Each DNA path was visualized specifically. To confirm which line in the visualized gel corresponded to which DNA path, we compared this result with the result of each specific DNA path used as the applied sample in the same experimental condition (data not shown). The employed specific DNA paths were amplified using specific primers. The specific primers for these amplifications are shown in Table 2. From this result, we could confirm that DNA paths of the same size can be separated by using DGGE. In DGGE, the migration of DNA in acrylamide gel depends on its sequences and sizes. The optimal denaturant concentration for the separation of each DNA path was 43% from the result of Fig. 3. However, it is not always possible to find the optimal denaturant such that all DNA paths can be seperated. Further gel analysis may be necessary.

Next, in order to quantify each DNA path, we performed CDGE on a 12% acrylamide gel in 43% denaturant. Figure 4 shows amplified DNA paths separated by CDGE, captured by Chemi Doc. To confirm which band in the amplified DNA path lanes corresponded to which DNA path, specific DNA paths were compared (lanes 2-7). The employed specific DNA paths were generated using specific primers (Table 2) and a ligation mixture as a template. The amplified DNA paths were separated into five bands (Fig. 4, lane 1), where each band was identified by comparing specific DNA path bands (Fig. 4, lanes 2-7). Five bands were identified with the DNA paths (0-1-4-5), (0-2-3-4-5), (0-1-2-3-5), (0-1-2-5)+(0-2-3-5), and (0-2-5), respectively, in order from the upper band. The DNA paths (0-1-2-5) and (0-2-3-5) were close migrations in 43% denaturant gel,

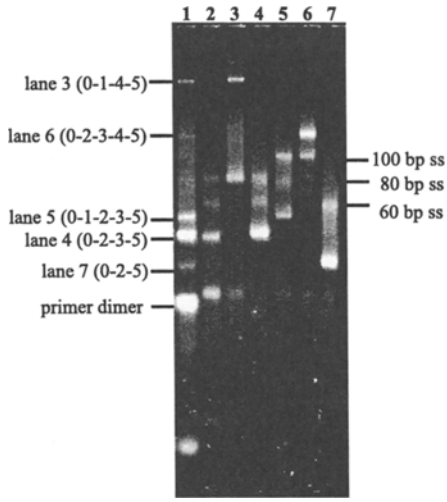


Fig. 4 Separation of DNA Paths by CDGE Run at 130 V for 3 Hours with a 12% Acrylamide Gel in a 43% Denaturant at 60°C. Lane 1, Amplified DNA Paths; Lane 2, 0-1-2-5; Lane 3, 0-1-4-5; Lane 4, 0-2-3-5; Lane 5, 0-1-2-3-5; Lane 6, 0-2-3-4-5; Lane 7, 0-2-5.

Table 2 Nucleotide sequences of primers. In GC+P_{0-1} and GC+P_{0-2}, 30 bp sequences from the 5′ end were GC clamp sequences. In $GC + P_{0-1}$ and $GC + P_{0-2}$, the nucleotide sequences that could anneal with a complementary strand of O_0 are underlined and those that could anneal with the 3′ end of a complementary strand of O_1 or O_2 are indicated by bold letters. In P_{5-2}, P_{5-3} and P_{5-4}, the nucleotides that could anneal with O_5 are underlined and those that could anneal with the 3′ end of O_2, O_3, or O_4 are indicated by bold letters. Specific DNA paths having the GC clamp at the 5′ end side were amplified by using two primer pairs among the primers in PCR. The employed primer pairs for each specific DNA path were described as follows: 0-2-5 (GC+P_{0-2}, P_{5-2}), 0-1-2-5 (GC+P_{0-1}, P_{5-2}), 0-1-4-5 (GC+P_{0-1}, P_{5-4}), 0-2-3-5 (GC+P_{0-2}, P_{5-3}), 0-1-2-3-5 (GC+P_{0-1}, P_{5-3}), 0-2-3-4-5 (GC+P_{0-2}, P_{5-4}).

Primers	Sequences (5' 3')
GC+P_{0-1}	GCCGGGGCGGGGCCGGGCGGGCGGCCCGGG
	TGGTTCCGATGTTTAAGCAA**AG**
GC+P_{0-2}	GCCGGGGCGGGGCCGGGCGGGCGGCCCGGG
	TGGTTCCGATGTTTAAGCAA**CC**
P_{5-2}	TCTGGCTGAGTCAAGTGATC**CCGC**
P_{5-3}	TCTGGCTGAGTCAAGTGATC**ACTA**
P_{5-4}	TCTGGCTGAGTCAAGTGATC**CTTT**

so it was difficult to distinguish these two DNA paths. In this case, we regarded the DNA path (0-1-2-5) as not detected, because the DNA path (0-1-2-5) was not seen in Fig. 3. Accordingly, the fourth band was determined to represent the DNA path (0-2-3-5). If these two DNA paths must be distinguished, we need to try CDGE at another denaturant concentration, which is able to separate these two DNA paths. In specific DNA path lanes (Fig. 4, lanes 2-7), we detected other bands except the main band. These bands appeared as single strand DNA.

In order to quantify the DNA path representing the shortest path, we analyzed the intensity of each DNA band by using Quantity One (software). The results of the analysis are shown in Table 3. The most intensive band was certainly identical to the shortest path. If N.D. is interpreted as zero, the correlation coefficient of the DNA amount in the laboratory and simulation experiments is 0.88. Therefore, the concentration control seems to work well, but further experiments and analyses on the simulation model are necessary. A theoretical analysis is also needed.

The advantage of DGGE separation was found to be direct quantification from amplified DNA paths. We could determine which DNA path was the shortest by a CDGE experiment, upon determining the optimal denaturant concentration capable of separating all DNA paths in advance. For the example graph, we could solve the shortest path problem. However, it can not be guaranteed that the shortest path is found by extracting only the most intense band. Actually, some other bands may have to be extracted and undergo sequencing, particularly for larger and more complicated graphs, including the case that there are multiple shortest paths in the given graph. Therefore, we have to regard the concentration control method as a kind of heuristic method for solving shortest path problems.

Table 3 Relative DNA amount of each DNA path. Column Lab. represents relative DNA amounts obtained by laboratory experiments, and column Sim. represents relative DNA amounts obtained by simulation experiments. N. D. represents amounts that could not be determined.

DNA paths	Lab.	Sim.
0-2-5	0.07	0.11
0-1-2-5	N. D.	0.08
0-1-4-5	0.13	0.32
0-2-3-5	1.00	1.00
0-1-2-3-5	0.25	0.76
0-2-3-4-5	0.04	0.04
0-1-2-3-4-5	N. D.	0.03

§5 Conclusions

In this paper, we demonstrate that an instance of the shortest path problems can be solved by a series of biochemical operations based on concentration

control. In particular, we showed that the proposed selection method is very effective for separating a lot of generated DNA paths with the same length, and for quantifying their concentrations. By the concentration control, irrelevant DNA complexes can be reduced, and by the separation method using DGGE and CDGE, the target DNA can be extracted efficiently. We believe that the combined use of these two methods will be useful for future DNA computing.

References

1) Adleman, L., "Molecular Computation of Solutions to Combitonatrial Problems," *Science, 266*, pp. 1021–1024, 1994.

2) Lipton, R. J., "DNA Solution of Hard Computational Problems," *Science, 268*, pp. 542–545, 1995.

3) Ouyang, Q., Kaplan, P., Lir, S. and Libchaber, A., "DNA Solution of the Maximal Clique Problem," *Science, 278*, pp. 446–449, 1997.

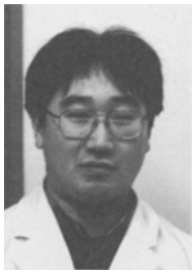
4) Sakamoto, K., Gouzu, H., Komiya, K., Kiga, D., Yokoyama, S., Yokomori, T. and Hagiya, M., "Molecular Computation by DNA Hairpin Formation," *Science, 288*, pp. 1223–1226, 2000.

5) Matsuura, N., Yamamoto, M., Shiba, T., Kawazoe, Y. and Ohuchi, A., "Solutions of Shortest Path Problems by Concentration Control," *in Preliminary Proc. of 7th International Meeting on DNA Computing*, pp. 231–240, 2001.

6) Wood, D. H., Chen, J., Antipov, E., Lemieux, B. and Cedeño, W., "*In Vitro* Selection for a OneMax DNA Evolutionary Computation," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 54*, pp. 23–37, 2000.

7) Yamamoto, M., Yamashita, J., Shiba, T., Hirayama, T., Takiya, S., Suzuki, K., Munekata, M. and Ohuchi, A., "A Study on the Hybridization Process in DNA Computing," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 54*, pp. 101–110, 2000.

Masahito Yamamoto, Ph.D.: He is associate professor of information engineering at Hokkaido University. He received Ph.D. from the Graduate School of Engineering, Hokkaido University in 1996. His current research interests include DNA computing based the laboratory experiments. He is a member of Operations Research Society of Japan, Japanese Society for Artificial Intelligence, Information Processing Society of Japan etc.



Atsushi Kameda, Ph.D.: He is the research staff of Japan Science and Technology Corporation, and has participated in research of DNA computing in Hokkaido University. He received his Ph.D. from Hokkaido University in 2001. For each degree he majored in molecular biology. His research theme is about the role of polyphosphate in the living body. As one of the researches relevant to it, he constructed the ATP regeneration system using two enzyme which makes polyphosphate the phosphagen.

Nobuo Matsuura: He is a master course student of Division of Systems and Information Engineering of Hokkaido University. His research interests relate to DNA computing with concentration control for shortest path problems, as a means of solution of optimization problems with bimolecular.

Toshikazu Shiba, Ph.D.: He is associate professor of biochemical engineering at Hokkaido University. He received his Ph.D. from Osaka University in 1991. He majored in molecular genetics and biochemistry. His research has progressed from bacterial molecular biology (regulation of gene expression of bacterial cells) to tissue engineering (bone regeneration). Recently, he is very interested in molecular computation and trying to apply his biochemical idea to information technology.

Yumi Kawazoe: She is a master course student of Division of Molecular Chemistry of Hokkaido University. Although her major is molecular biology, she is very interested in molecular computation and bioinformatics.

Azuma Ohuchi, Ph.D.: He is professor of Information Engineering at the University of Hokkaido, Sapporo, Japan. He has been developing a new field of complex systems engineering, i.e., Harmonious Systems Engineering since 1995. He has published numerous papers on systems engineering, operations research, and computer science. In addition, he is currently supervising projects on DNA computing, multi-agents based artificial market systems, medical informatics, and autonomous flying objects. He was awarded "The 30th Anniversary Award for Excellent Papers " by the Information Processing Society of Japan. He is a member of Operations Research Society of Japan, Japanese Society for Artificial Intelligence, Information Processing Society of Japan, Japan Association for Medical Informatics, IEEE Computer Society, IEEE System, Man and Cybernetics Society etc. He received PhD from Hokkaido University in 1976.