Available online at www.sciencedirect.com
SCIENCE DIRECT®
BioSystems 78 (2004) 39–47
BioSystems
www.elsevier.com/locate/biosystems

Solving traveling salesman problems with DNA molecules encoding numerical values

Ji Youn Lee[a], Soo-Yong Shin[b], Tai Hyun Park[a,*], Byoung-Tak Zhang[b,1]

[a] *School of Chemical Engineering, Seoul National University, San 56-1 Shillim-Dong, Kwanak-Gu, Seoul 151-744, Republic of Korea*
[b] *School of Computer Science and Engineering, Seoul National University, San 56-1 Shillim-Dong, Kwanak-Gu, Seoul 151-744, Republic of Korea*

Received 4 August 2003; received in revised form 27 June 2004; accepted 30 June 2004

Abstract

We introduce a DNA encoding method to represent numerical values and a biased molecular algorithm based on the thermodynamic properties of DNA. DNA strands are designed to encode real values by variation of their melting temperatures. The thermodynamic properties of DNA are used for effective local search of optimal solutions using biochemical techniques, such as denaturation temperature gradient polymerase chain reaction and temperature gradient gel electrophoresis. The proposed method was successfully applied to the traveling salesman problem, an instance of optimization problems on weighted graphs. This work extends the capability of DNA computing to solving numerical optimization problems, which is contrasted with other DNA computing methods focusing on logical problem solving.

Keywords: DNA computing; Melting temperature control encoding method; Weight representation in DNA; Traveling salesman problem

1. Introduction

Deoxyribonucleic acid or DNA plays the role of memory in nature. DNA is the genetic material containing the whole information of an organism to be copied into the next generation of the species. DNA computing is a computational paradigm that uses synthetic (or natural) DNA molecules as information storage media. The techniques of molecular biology, such as polymerase chain reaction (PCR), gel electrophoresis,[2] and enzymatic reactions, are used as computational operators for copying, sorting, and splitting/concatenating

* Corresponding author. Tel.: +82 2 880 8020; fax: +82 2 875 9348.
E-mail addresses: thpark@plaza.snu.ac.kr (T.H. Park), btzhang@bi.snu.ac.kr (B.-T. Zhang).
[1] Corresponding Co-author. Tel.: +82 2 880 1833; fax: +82 2 875 2240.

[2] PCR is a technique that reproduces specific DNA strands by providing specific complementary sequences (primers) and enzymes (DNA polymerases). Gel electrophoresis is a technique to sort DNA molecules by their molecular weight.

0303-2647/$ – see front matter © 2004 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.biosystems.2004.06.005

Table 1
Comparison of DNA computers with conventional computers

	DNA computers	Conventional computers
Storage media	Nucleic acids	Semiconductors
Memory capacity	Ultra-high	High
Operators	Biochemical operations	Logical (and, or, not)
Operation	Simultaneous (parallel)	Bitwise (sequential)
Speed of each operation	Slow	Fast
Process	Stochastic	Deterministic

the information in the DNA molecules, respectively (Adleman, 1994).

Computing with DNA molecules has many advantages over conventional computing methods that utilize solid-state semiconductors. The properties of DNA computing compared with conventional computers are summarized in Table 1. Though DNA computing performs individual operations slowly, it can execute billions of operations simultaneously. This is contrasted with the electronic digital computers where individual operations are very fast; however, the operations are executed basically sequentially. The massive parallelism of DNA computing comes from the huge number of molecules which chemically interact in a small volume (Maley, 1998). DNA also provides a huge storage capacity since they encode information on the molecular scale.

Based on the massive parallelism, many researchers tried to solve a host of difficult problems, especially NP problems (Ouyang et al., 1997; Faulhammer et al., 2000; Sakamoto et al., 2000). NP (nondeterministic polynomial time) problems are a class of mathematical problems which have most likely exponential complexity, for which no efficient solution has been found yet (Garey and Johnson, 1979).

Most of the previous researches in DNA computing do not require the consideration of the representation of numerical data in DNA strands (Adleman, 1994; Ouyang et al., 1997; Faulhammer et al., 2000; Sakamoto et al., 2000). However, many practical applications in the real world involve edge-weighed graph problems. Examples include the shortest path problem, the traveling salesman problem, the minimum spanning tree problem, and the Steiner tree problem (Garey and Johnson, 1979). Therefore, representation of numerical data in DNA strands is an important issue toward expanding the capability of DNA computing to solve numerical optimization problems. There exists previous work to represent the numerical data with DNA, but the results are not satisfactory yet (Narayanan and Zorbalas, 1998; Shin et al., 1999; Yamamura et al., 2002).

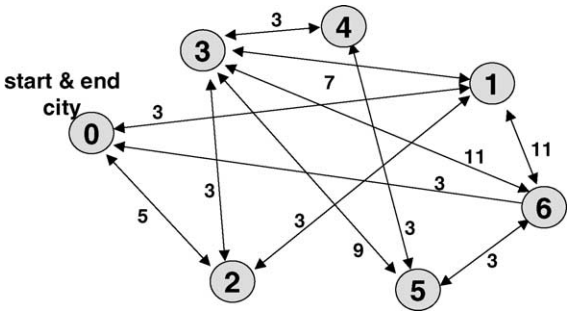

Fig. 1. The 7-city traveling salesman problem solved in this work. The paths start and end with city 0. The circles denote the cities and the arrows represent the roads. The number on each arrow represents the cost on the given road.

In this paper, we introduce a novel encoding method that utilizes a temperature gradient to overcome the drawbacks of previous work. The basic idea is to design the sequences so that the DNA strands for higher-cost values have higher melting temperatures than those for lower-cost values. We also suggest a molecular algorithm that includes a modified PCR protocol (more details in Section 3), which makes local search around the solution space more efficient and the DNA computing protocol more powerful. The traveling salesman problem (TSP) is used as a test problem for the proposed encoding method and molecular algorithm. TSP is to find a minimum cost (weight) path for a given set of cities (vertices) and roads (edges). The path must begin at a specified city and end there after passing through all the given cites (Garey and Johnson, 1979). TSP problems have also been used extensively for developing novel evolutionary algorithms (see, e.g., Fogel, 2000 and references therein). We solved a TSP instance consisting of 7 cities and 23 roads of five differing cost values as shown in Fig. 1.

The paper is organized as follows. In Section 2, we describe the DNA encoding scheme. Section 3 describes the molecular algorithm for solving the TSP problem using the encoding scheme. Section 4 de-

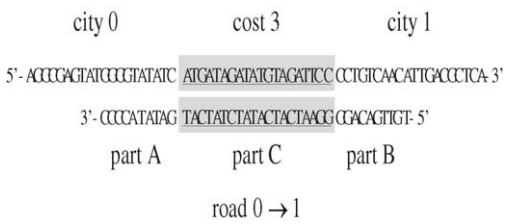

Fig. 2. Example of the encoding scheme for the road from city 0 to city 1.

scribes the experimental details for solving the problem using biochemical operations. Section 5 presents the experimental results and their analysis. Finally, in Section 6, we draw conclusions.

2. Melting temperature control encoding

Representation of weight information for weighted-graph problems is one of the most important but challenging problems. A couple of studies proposed methods for representing costs on DNA strands in graph problems. For example, Narayanan and Zorbalas (1998) presented a conceptual encoding method that represents costs with the lengths of DNA strands. However, this method requires very long DNA strands to solve problem instances containing a wide range of costs, making experimental implementation impossible. Yamamura et al. (2002) proposed a concentration control method. The concentrations of DNA strands were used as input and output data, i.e., the numerical data were encoded by the concentrations of DNA strands. This has the advantage that it facilitates local search among all candidate solutions rather than searching entire space exhaustively. However, this method could not ensure the optimal solution: the brightest band in the gel was not guaranteed to be the optimal solution. In addition, it was technically impossible to extract the single optimal solution from the brightest band which might contain several kinds of DNA strands. Though they showed the possibility of concentration control method, they failed in finding the final solution.

Shin et al. (1999) proposed a method for representing the real numbers in fixed-length DNA strands by varying the number of hydrogen bonds. Although the number of hydrogen bonds is an important factor in deciding the thermal stability of DNA strands, it is not a sufficient factor. The melting temperature (T_m) at which, under a given set of conditions, the half of double-stranded DNA is changed to single-stranded DNA is a more direct characteristic indicating the stability of a DNA duplex.

In this paper, we propose a melting temperature control encoding method. This method uses fixed-length DNA strands and represents costs by melting temperatures of the given DNA strands. Fig. 2 illustrates the encoding scheme for the path from city 0 to city 1 in the graph of Fig. 1. The road sequence consists of three parts: one for the departure city (part A), another for the cost sequence (part C), and the last for the arrival city (part B). The first part of the road sequence (3′-CCCCATATAG-5′) is a Watson-Crick (WC) complement of the second half of the departure city (5′-GGGGTATATC-3′) and the last part (3′-GGACAGTTGT-5′) is a WC complement of the first half of the arrival city (5′-CCTGTCAACA-3′). The middle part represents the cost information. Parts A and B act as linkers to bridge the cities.

To design the necessary sequences for weighted-graph problems using the proposed method, city sequences were designed first (Fig. 2). Each city sequence is designed to have a similar melting temperature, because city sequences should contribute equally to the thermal stability of paths. Then, cost sequences are designed to avoid unintended hybridizations with city sequences. Cost sequences are designed to have various melting temperatures according to the costs. A smaller cost is represented by a DNA sequence with a lower melting temperature, and therefore a more economical path has a lower melting temperature. Finally, road sequences that connect two cities are generated using the sequences of departure cities, arrival cities, and costs.

Melting temperature of a DNA strand is the most important factor of our encoding method. Several empirical methods are proposed to calculate the melting temperature. A classical method is a GC content method which uses the content of guanine (G) and cytosine (C) in the given strand as a main factor determining melting temperature. This method is generally acceptable to DNA strands longer than 50 nucleotides in pH 5–9 (Wetmur, 1991). Recently, a statistical method using thermodynamic parameters such as enthalpy (ΔH) and entropy (ΔS) has been proposed and widely used to calculate melting temperature of DNA oligonucleotide

(SantaLucia, 1998). The latter approach, known as the nearest-neighbor (NN) model, calculates the thermal stability of a DNA strand on the assumption that the stability of a given base pair depends on the identity and orientation of neighboring base pairs. It is more accurate than the GC content method and it is applicable to DNA strands up to 108 base pairs (bp).

Both the GC content method and the NN method were used in this study to calculate melting temperatures. If the sequences are short, the NN method is used; otherwise, the GC content method is applied. While the melting temperatures of a DNA strands up to 108 bp can be accurately determined by the NN method, the melting temperatures of lengthened DNA strands generated after hybridization and ligation can be appropriately determined by the GC content method rather than by the NN method.

3. Molecular algorithm for solving TSPs

As mentioned in Section 1, the massive parallelism is a great advantage of using DNA molecules and their biochemical reactions. However, simple exhaustive search algorithms that are based on straightforward parallelism do not scale up to hard problems. Therefore, a more intelligent molecular algorithm is necessary to maximize the ability of DNA computing.

To perform guided search while utilizing massive parallelism, it is useful to give biases in operators, i.e., experimental steps. The bias refers to additional assumptions which the algorithm will use in order to guide search for situations that have not been encountered so far (Mitchell, 1997). We carried out two approaches for those purposes: one is a biased generation of the initial pool and the other is a biased search of the solution set.

The first approach is to alter the concentration of DNA strands. The rate of biochemical reactions depends on the reaction rate constants and the reactant concentrations. Therefore, the concentration of DNA strands may be a useful factor for generating the optimal path more effectively. As the concentrations of DNA strands increase, the paths including them are generated more frequently. For this purpose, some constraints were given in the concentration of the initial DNA strands participating in the initial pool (Yamamura et al., 2002). Although this approach can amplify economical paths more intensively, it still has an inherent shortcoming because it is impossible to isolate the optimal path among the paths of the same length.

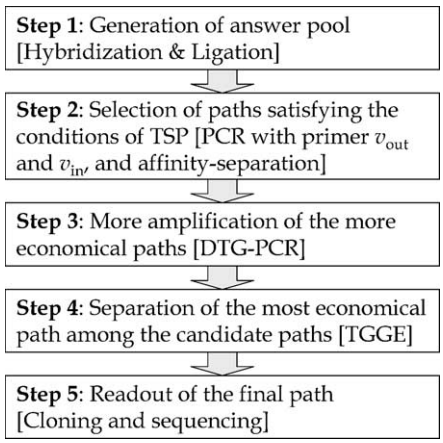

Fig. 3. Molecular algorithm for solving the TSP.

The second approach is to modify the PCR protocol so that it employs temperature gradient in the denaturation step. The denaturation temperature was low in the beginning of PCR and then it was increased gradually, cycle by cycle. With this denaturation temperature gradient PCR (DTG-PCR), the DNA strand of lower melting temperature (T_m) will be amplified more frequently. Consequently, more economical paths of lower T_m will be amplified more intensively.

Fig. 3 summarizes the molecular algorithm for solving the traveling salesman problem. It consists of five steps. In Step 1, the initial pool is generated by using DNA hybridization and ligation reactions with the oligonucleotides representing cities, roads, and costs. Step 2 selects only the paths that satisfy the necessary conditions to be the candidate solutions. The selection is achieved by a series of biochemical operations, such as polymerase chain reaction (PCR), gel electrophoresis, and affinity separations. The paths which start and end at a given city are selectively amplified by PCR. Therefore, TSP condition such as starting and ending vertex can be satisfied. Then, visiting all vertices once can be checked by gel electrophoresis and affinity separations. Gel electrophoresis checks the length of target paths, and affinity separation checks if all vertices are visited. In Step 3, more economical paths are more amplified by DTG-PCR. In Step 4, the most economical

Table 2
City and cost sequences for the seven-city traveling salesman problem

	Sequence (5′ → 3′)	T_m (°C)[a]	GC content (%)
City			
0	AGGCGAGTATGGGGTATATC	44.51	50
1	CCTGTCAACATTGACGCTCA	47.59	50
2	TTATGATTCCACTGGCGCTC	47.07	50
3	ATCGTACTCATGGTCCCTAC	44.77	50
4	CGCTCCATCCTTGATCGTTT	47.43	50
5	CTTCGCTGCTGATAACCTCA	46.77	50
6	GAGTTAGATGTCACGTCACG	45.46	50
Cost			
3	ATGATAGATATGTAGATTCC	35.03	30
5	GGATGTGATATCGTTCTTGT	41.97	40
7	GGATTAGCAGTGCCTCAGTT	46.51	50
9	TGGCCACGAAGCCTTCCGTT	53.57	60
11	GAGCTGGCTCCTCATCGCGC	55.07	70

[a] T_m was calculated using the NN method with 1 nM oligonucleotide concentration and 50 mM salt concentration.

path is separated from other candidate paths by temperature gradient gel electrophoresis (TGGE). Finally, the readout of the final solution is performed by cloning and sequencing the separated DNA strands in Step 5.

4. Materials and methods

4.1. Synthesis of DNA oligonucleotides

DNA oligonucleotides encoding cities and costs were designed, respectively, by 20 mer single-stranded DNA (ssDNA) as listed in Table 2. The sequences encoding roads were generated as described in Section 2. The 40 mer ssDNA coding a road consists of three parts and contains the cost information in the middle. All of the oligonucleotides were 5′-phosphorylated. Other 5′-biotinylated complementary vertexes were prepared for affinity separation.

4.2. Hybridization and ligation

The added amounts of oligonucleotides encoding roads were not equal but varied according to costs encoded by part C. As the cost increased, the added amount was decreased. For example, the concentration of the cost 3 was 25 μM and that of cost 11 was 5 μM. The oligonucleotide mixture was heated to 95 °C and slowly cooled to 20 °C at 1 °C/min for hybridization. The reaction mixture was then subjected to a ligation. For a ligation, 5 μl of the reaction mixtures, 350 units of *T4* DNA ligase (TaKaRa, Japan), ligase buffer (66 mM Tris–HCl, pH 7.6, 6.6 mM $MgCl_2$, 10 mM DTT, 0.1 mM ATP), and an appropriate volume of distilled water were mixed. The total reaction volume was 10 μl. The reaction mixture was incubated at 16 °C for 16 h.

4.3. PCR, gel electrophoresis, and affinity separation

For normal amplification, 0.5 μM of each primer and *AccuPower*® PCR PreMix (Bioneer, Korea) containing 1 unit of *Taq* DNA polymerase in 10 mM Tris–HCl, pH 9.0, 1.5 mM $MgCl_2$, 40 mM KCl, and 0.25 mM each dNTP were dissolved in distilled water. The total reaction volume was 20 μl. PCR was processed for 34 cycles at 95 °C for 30 s at (T_m − 5) °C for 30 s, and at 72 °C for 30 s. Initial denaturation and prolonged polymerization were executed for 5 min each. Agarose gel electrophoresis was performed with 2% Agarose-1000 (GibcoBRL, NY, USA) in 0.5× Tris–borate–EDTA buffer and gel was stained with ethidium bromide. As a marker, 50 bp DNA Ladder (GibcoBRL, NY, USA) was used. Polyacrylamide gel electrophoresis for oligonucleotides was performed with 20% native gel in 0.5× Tris–borate–EDTA buffer and gel was stained with ethidium bromide. Affinity separation was performed to produce ssDNA and to check whether all seven cities were included. The affinity separation protocol was basically the same as Adleman's experiment (Adleman, 1994). Details are not repeated here.

4.4. Denaturation temperature gradient PCR

A denaturation temperature gradient PCR is a modified PCR method that the denaturation temperature changes with cycle (Lee et al., 2003). The denaturation temperature started at low temperature (70 °C) in the beginning cycles of PCR. Then the denaturation temperature was gradually increased by 1 °C per cycle until it reached to 95 °C and maintained at 95 °C for the remaining cycles. Other PCR conditions were identical with those of the normal PCR described above.

4.5. Temperature gradient gel electrophoresis

For a temperature gradient gel electrophoresis, 8% denaturing polyacrylamide gel containing 20% for-

mamide and 4.2 M urea was used. The electrophoresis was performed at over 55–70 °C at 200 V for 3 h with the custom-built instrument. After the electrophoresis, the gel was silver-stained. The DNA strands in the main band were excised, eluted, and amplified by PCR. This PCR product was directly cloned into pBluescript SK+ and sequenced using T7 primer.

5. Results and discussion

5.1. Sequence design and verification

Proper design of oligonucleotides is important for a successful implementation in DNA computing (Brenneman and Condon, 2002). DNA is an information carrier, and hybridizations between DNA molecules basically carry out the computation. Therefore, unintended hybridizations such as base mismatches, shifted hybridizations, and hairpin formations can lead to fatal errors during the implementation steps. Although inborn errors still exist in the biochemical operations, possible unintended interactions among DNA strands can be minimized by a careful design (Shin et al., 2002). Table 2 shows the sequences designed by NACST/Seq which is a sequence design program using a multi-objective evolutionary algorithm (Kim et al., 2003). The orthogonalities of those oligonucleotides were confirmed by hybridization and gel electrophoresis (Fig. 4).

The oligonucleotides that correspond to the city sequences were hybridized and run in polyacrylamide gel. No upward shift resulted from the unintended hybridization was observed. Therefore, we can assure that the sequences are well designed and they will tend to hybridize with intended sequences favorably.

5.2. Experimental solution of a 7-city TSP

The molecular algorithm for solving the TSP starts from the generation of a path pool by the hybridization of DNA strands that represent the cities and roads. All possible paths of the 7-city traveling salesman problem can be simultaneously generated under

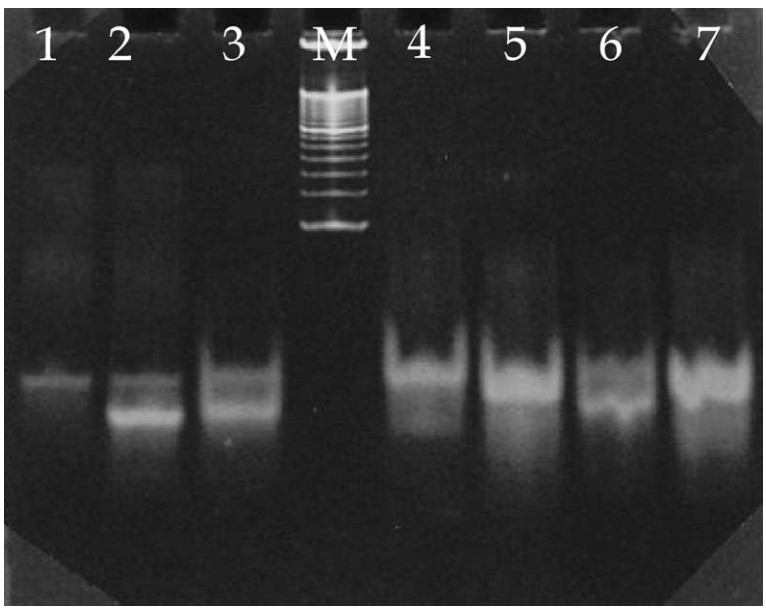

Fig. 4. Gel image that shows the orthogonalities of the sequences (20% polyacrylamide gel). Lane M denotes 50-bp ladder, lane 1 is for city 1 (200 pmol), lane 2 is for the mixture of cities 0 and 1 (each 200 pmol), and lane 3 is for the hybridized solution of cities 0 and 1. There was no upward shift by hybridization in the gel. Lane 4 is for the hybridized solution of cities 3 and 4, lane 5 is for the hybridized solution of cities 5 and 6, lane 6 is for the hybridized solution of cities 0, 1, and 2, and lane 7 is for the hybridized solution of cities 3, 4, 5, and 6. The gel image shows there are few interactions among DNA strands for the cities.

massive parallelism of DNA molecules. Assuming the paths are generated randomly by hybridization and ligation of short DNA oligonucleotides, the number of the possible paths is $\sum_{m=1}^{n} m!$, where n is the number of DNA oligonucleotides representing the cities. However, generations of longer paths are restricted due to various experimental conditions such as the limited reaction time and the limited reagents. In our experimental setting, we can assume that the generated paths are smaller than $\sum_{m=1}^{10} m = 4\,037\,913$. Though very restricted assumption, this is reasonable since the number of paths visiting 10 cities will be smaller than 10! In the test tube, there are approximately 8×10^{15} molecules which are subject to the hybridization and ligation. Therefore, it is very likely that a complete initial solution pool is generated. Our experimental results described below confirm this.

Fig. 5(A) shows the gel image of initial pool that generated by hybridization and ligation. Compared with the oligonucleotide mixture (lane 3), the ligated DNA strands were lengthened (lane 1). There were few copies around the 300 bp which is the length of path that contains eight cities. Therefore, the ligation reaction was repeated, and consequently upper-shifted ligation products were obtained (lane 2). Ligation is an essential step for generating random paths, and so the development of an efficient ligation method is necessary to produce DNA strands long enough to solve large problems.

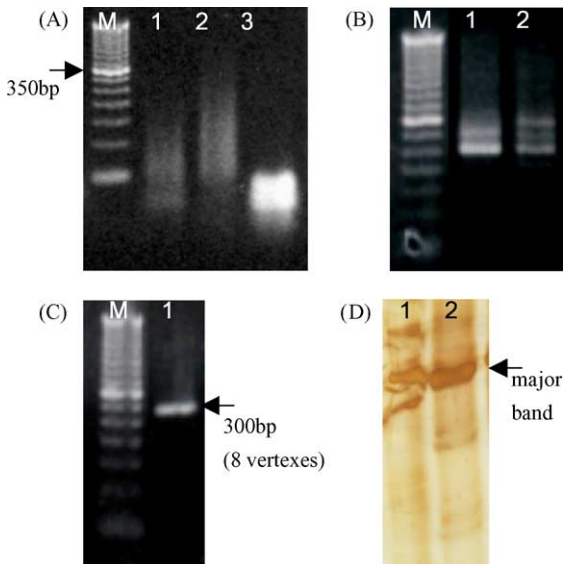

Fig. 5. Experimental results of gel electrophoresis on 2% agarose gel (A–C) and 8% denaturing polyacrylamide gel (D). M denotes the DNA size marker (50-bp ladder). (A) Lane 1 is the ligation product (set of generated random paths), lane 2 is the repetitive ligation product and lane 3 is the oligonucleotide mixture. More elongated DNA strands were observed after repetitive reaction. (B) DNA strands that start with city 0 and end with city 0. (C) The final DTG-PCR result (after step 3). (D) After the TGGE separation, the major band appeared and this was eluted and sequenced for readout of the path.

To satisfy the TSP conditions, first the path must begin at a specified city and end there (city 0 in Fig. 1). Second, the path must pass through all the given cities (cities 1–6 in Fig. 1). The first condition was implemented with two successive PCRs. The DNA strands corresponding to city 0 were used as primer in the first PCR and the DNA strands complementary to city 0 were used as primer in the second PCR. The PCRs were carried out separately since the two primers are complementary to each other. All PCR products were sieved by 2% agarose gel electrophoresis. After the first PCR, the DNA strands of unreasonable lengths were excluded because the candidate paths should be eight-cities long. The second PCR product appeared in several bands in the gel as shown in Fig. 5(B). The DNA strands of around 300 bp were excised, eluted, and amplified with 5′-biotinylated city 0 as a primer. The amplified product was used in the following affinity separation with streptavidin-coated paramagnetic particles to select the paths visiting every city, which is the second condition of TSP. Affinity separation was performed seven times with different biotinylated DNA strands. The biotinylated probe strands were complementary ones of the cities. The DNA strands that pass through all the given cities were separated by this procedure.

The DNA strands that survive the implementation procedures up to the affinity separation have the same length, start and end at city 0, and visit all the seven cities. Therefore, they have similar physical characteristics and cannot be easily separated. To solve the difficulty, we developed a DTG-PCR as a biased search operator. Because the base compositions are different, the thermodynamic characteristics are somewhat different to each other. The cost sequences were designed to have different melting temperature: low costs with low T_m and high costs with high T_m. So, the T_m of the most economical path is the lowest among the candidate paths.

DTG-PCR is a specified PCR protocol that modifies the denaturation temperature profile. If the denaturation temperature is decreased to a certain level in PCR, the DNA strands with denaturation temperatures lower than that temperature will be denatured and amplified. As the denaturation temperature is increased cycle by cycle in PCR, other DNA strands with higher denaturation temperatures will also be amplified. However, the economical paths will be more amplified and will occupy the major part of the solution. Therefore, they can be detected more easily. The effectiveness of DTG-PCR is presented in other work (Lee et al., 2003).

In this work, the denaturation temperature starts at 70 °C which is lower than the melting temperatures of template strands. After DTG-PCR, one main band was observed in the gel as shown in Fig. 5(C). This band might contain four different DNA strands of the possible paths '0 → 1 → 2 → 3 → 4 → 5 → 6 → 0 (the sum of weights: 21)', '0 → 1 → 6 → 5 → 4 → 3 → 2 → 0 (31)', '0 → 2 → 1 → 3 → 4 → 5 → 6 → 0 (27)', '0 → 2 → 3 → 4 → 5 → 6 → 1 → 0 (31)'. Those strands are of the same length and cannot be separated by the conventional gel electrophoresis. However, they have distinct melting behaviors, and so can be separated by temperature gradient gel electrophoresis (TGGE). TGGE is a gel electrophoresis method that is based on the correlation of the melting characteristic of a DNA strand to its electromigration. TGGE is an extremely sensitive method, and so it can detect even point mutations. We applied this electrophoresis method to the separation of the most economical path from other possible paths. The GC content of the four possible DNA strands varies from 40.67 to 44.00%. The melting temperatures of the DNA strands are 75.12, 76.84, 76.24, and 76.95 °C, respectively, by the NN model with 50 mM salt concentration and 1 nM DNA concentration and 70.84, 72.20, 71.66, and 72.20 °C, respectively, by the GC content method under the same condition. The PCR product containing those possible paths was separated by TGGE and one major band and other minor bands were observed (Fig. 5(D)). The DNA strands in the major band were cloned and sequenced, and the DNA sequence corresponded to the path '0 → 1 → 2 → 3 → 4 → 5 → 6 → 0' which has the smallest cost. The sequencing result of the final product confirmed that all of the experimental procedures were properly carried out in finding the optimal solution.

6. Conclusions

We presented an encoding method and the associated molecular algorithm that uses melting temperature of DNA strands to solve the traveling salesman problem efficiently. The experimental results show that a series of biochemical experimental operations successfully find the optimal solution of the 7-city traveling salesman problem.

One of the most important features of our method is that it can handle quantitative expression of real numbers using fixed-length DNA strands. The ability of representing and manipulating numerical data in DNA computing opens up the possibility of solving a wide range of numerical optimization problems. In terms of efficiency, the use of fixed-length DNA strands seems more likely to scale up than the variable representation where the DNA length grows with the problem size. In addition, our encoding method and molecular algorithm allow us to perform a guided search. In the solution generation step, for example, by varying the concentrations of the DNA strands we can drive the paths of smaller costs to be more frequently generated than the paths of larger costs. Also, denaturation temperature gradient polymerase chain reaction was used to guide the search by following temperature gradient, rather than filtering out the entire solution, and amplify better solutions (i.e., paths with smaller sums of costs) more intensively.

One limiting factor for the temperature gradient method is the precision of biochemical techniques available. For example, in our experiments we used 20 bases of nucleotides for representing a weight. The length of the DNA strands can be reduced as we improve the techniques for controlling experimental temperatures and the resolution of DNA detection technologies.

Acknowledgement

This research was supported by the MEC Project from the Ministry of Commerce, Industry and Energy, by the NRL Program from the Ministry of Science and Technology, and by the BK21-IT Program from the Ministry of Education and Human Resources Development. The ICT at Seoul National University provided research facilities for this study.

References

Adleman, L.M., 1994. Molecular computation of solutions to combinatorial problems. Science 266, 1021–1024.

Brenneman, A., Condon, A., 2002. Strand design for biomolecular computation. Theor. Comput. Sci. 287, 39–58.

Faulhammer, D., Cukras, A.R., Lipton, R.J., Landweber, L.F., 2000. Molecular computation: RNA solutions to chess problems. Proc. Natl. Acad. Sci. USA 97, 1385–1389.

Fogel, D.B., 2000. Evolutionary Computation: Toward a New Philosophy of Machine Intelligence, 2nd ed. IEEE Press.

Garey, M.R., Johnson, D.S., 1979. Computers and Intractability: A Guide to the Theory of NP-completeness. W. H. Freeman and Company.

Kim, D., Shin, S.-Y., Lee, I.-H., Zhang, B.-T., 2003. NACST/Seq: a sequence design system with multiobjective optimization. Lecture Notes Comput. Sci. 2568, 242–251.

Lee, J.Y., Zhang, B.-T., Park, T.H., 2003. Effectiveness of denaturation temperature gradient-polymerase chain reaction for biased DNA algorithms. In: Cheng, J., Reif, J. (Eds.), Preliminary Proceedings of the Ninth International Meeting on DNA Based Computers, p. 208.

Maley, C.C., 1998. DNA computation: theory, practice, and prospects. Evol. Comput. 6, 201–229.

Mitchell, T.M., 1997. Machine Learning. McGraw-Hill.

Narayanan, A., Zorbalas, S., et al., 1998. DNA algorithms for computing shortest paths. In: Koza, J.R. (Ed.), Proceedings of the Genetic Programming 1998. Morgan Kaufmann, pp. 718–723.

Ouyang, Q., Kaplan, P.D., Liu, S., Libchaber, A., 1997. DNA solution of the maximal clique problem. Science 278, 446–449.

Sakamoto, K., Gouzu, H., Komiya, K., Kiga, D., Yokoyama, S., Yokomori, T., Hagiya, M., 2000. Molecular computation by DNA hairpin formation. Science 288, 1223–1226.

SantaLucia Jr., J., 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc. Natl. Acad. Sci. USA 95, 1460–1465.

Shin, S.-Y., Kim, D., Lee, I.-H., Zhang, B.-T., et al., 2002. Evolutionary sequence generation for reliable DNA computing. In: Fogel, D.B. (Ed.), Proceedings of the Congress on Evolutionary Computation 2002. IEEE Press, pp. 79–84.

Shin, S.-Y., Zhang, B.-T., Jun, S.-S., et al., 1999. Solving traveling salesman problems using molecular programming. In: Angeline, P.J. (Ed.), Proceedings of the Congress on Evolutionary Computation 1999. IEEE Press, pp. 994–1000.

Wetmur, J.G., 1991. DNA probes: applications of the principles of nucleic acid hybridization. Crit. Rev. Biochem. Mol. Biol. 26, 227–259.

Yamamura, M., Hiroto, Y., Matoba, T., 2002. Solutions of shortest path problems by concentration control. Lecture Notes Comput. Sci. 2340, 231–240.