DIMACS Series in Discrete Mathematics
and Theoretical Computer Science

Solid Phase DNA Solution to the Hamiltonian Path Problem

Nobuhiko Morimoto, Masanori Arita, and Akira Suyama†

Abstract. A solid-phase method for solving the Hamiltonian path problem (HPP) is described. The method employs only fast and simple DNA operations amenable to full automation. Single-stranded DNA molecules representing paths with no city visited twice are synthesized city-by-city from the start city on the surface of a solid support. The solution can thus be found in the computation time proportional to the number of cities. As well as the stepwise path synthesis, a pruning technique developed for the removal of looping paths helps the reduction of DNA molecules necessary for the computation; thus definitely increasing the size of problems solvable on a DNA-based computer. Experiments using Adleman's seven-city instance of the HPP showed that the path extension cycle was very accurate and took only about 45 min. Our solid-phase method has originally been developed for solving the HPP, but it could also be applied to other problems requiring a massive parallelism in computation.

1. Introduction

DNA-based computation utilizes the massive parallelism of DNA molecular reactions to solve the NP-complete problem in the computation time proportional to the size of a problem. In 1994, Adleman first demonstrated DNA molecules can be used to find the solution to a tiny instance of the directed Hamiltonian path problem (HPP) [**1**]. He generates all possible paths in a graph with DNA molecules at once. Then he finds the solution out of them by simultaneously subjecting all DNA molecules to a sequence of selection methods commonly used in molecular biology experiments. After that, Lipton and other researchers proposed DNA-based algorithms for solving other combinatorial NP-complete problems [**2**] [**3**]. These studies have blazed a trail in the field of DNA-based computation.

The massive parallelism of DNA-based computation proposes a new architecture of high-speed parallel computers. There are about 6×10^{15} DNA molecules in a DNA oligonucleotide solution with a volume ($100\,\mu$l) and a concentration (0.1 mM) usually used in biological experiments. So, if letting every DNA molecule represent

---

1991 *Mathematics Subject Classification.* Primary 68-06; Secondary 68M99, 68R99.

The second author was supported by Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists. This work was supported in part by the Japan Society for the Promotion of Science "Research for the Future" Program (JSPS-RFTF96100101).

†To whom correspondence should be addressed.

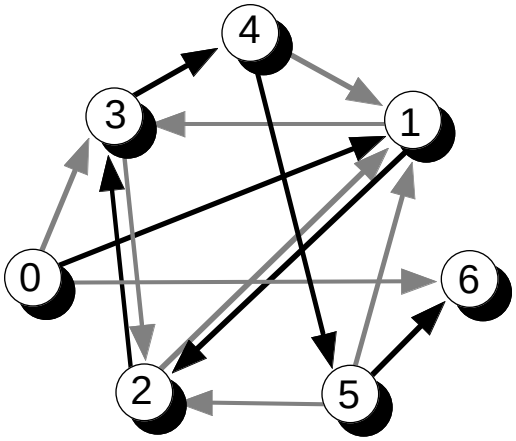

FIGURE 1. Adleman's graph for the HPP

a data or instruction stream, 6×10^{15} operations could be performed simultaneously. However, no DNA-based computations so far reported are actually fast. For example, Adleman spent a week finding a solution to a seven-city instance of the HPP. This is because the computation is performed by means of slow and laborious biological methods. Furthermore, the parallelism in a DNA solution is indeed huge, but it is not sufficient to represent all possible data streams for a large instance of the NP-complete problem. Therefore, to build up feasible DNA-based computers, both the speedup of each biological method and the great reduction in the number of data and instruction streams simultaneously represented with DNA molecules are needed

In the present study we describe a new experimental method for solving the HPP, which is much faster and requires much smaller number of DNA molecules than Adleman's method. Our method is based on a solid phase chemistry. DNA strands representing paths with no city visited twice are synthesized city-by-city from the start city on the surface of a solid support. An n-city instance of the HPP can be solved in $n-1$ extension cycles, which take about $45 \times (n-1)$ min. As well as the stepwise path synthesis, a pruning technique developed for the removal of looping paths helps the reduction of DNA molecules representing possible paths; thus definitely increasing the size of problems solvable on a DNA-based computer.

2. Materials and Methods

2.1. DNA oligomers. The same seven-city instance of the HPP as that given by Adleman (Fig. 1) was used in the present study. However, the set of base sequences that represent vertices (cities) and edges in the graph was different from that used by Adleman. The new set of base sequences that was suitable for the present method was designed and prepared.

Single-stranded DNA oligomer representing the start city was 40 bases long. It was composed of two parts: the 5′ end part of 25 bases long called a root sequence (RS) and the 3′ end part of 15 bases long specifying the start city, i.e. city 0 (v_0). Thus the start city oligomer is designated by RS-v_0. The 5′ end of RS-v_0 was biotinylated to be attached onto streptavidin-coated magnetic beads through the

biotin-streptavidin interaction. An RS part has the recognition site of a restriction enzyme *Eco*RI which allows RMVDC (remove duplicate cities) operation to get rid of looping path strands from the beads.

Single-stranded DNA oligomers representing other cities than the start city were also a 40-base oligomer composed of two parts but with a slightly different structure. The $5'$ end part was 15 bases long and used to specify city i (υ_i) in the graph. The $3'$ end part was a common sequence of 25 bases long called a tail sequence (TS), which was used to block coupling of more than one city to the end while executing EXT (extend) operation. These city oligomers are thus designated by υ_i-TS ($i = 1, 2, \cdots, 6$). The $5'$ end of υ_i-TS was phosphorylated for a ligation reaction in EXT operation. There is the recognition site of a restriction enzyme *Bam*HI in a TS part so that it can be removed by DEBLK (deblock) operation.

Base sequences specifying cities and those of RS and TS were designed to share no common subsequences of more than 5 bases long in order to prohibit inaccurate path extension and unfavorable association of single-stranded DNA oligomers in a reaction mixture. In addition, they were designed to have no stable secondary structure to ensure efficient annealing of complementary DNA oligomers. Their GC-content was not greatly biased, ranging from 0.33 to 0.67, so that all complementary DNA oligomers are subjected to annealing conditions with almost the same stringency.

The base sequences of the city oligomers used in the present study are as follows:

RS-υ_0: `5'-b-aaacgtctctccacgagcgcGAATT-Cggcctacttaagag-3'`
υ_1-TS: `5'-p-ttgttacgcagcccG-GATCCataggagactaagagagacg-3'`
υ_2-TS: `5'-p-attgccgaggtactG-GATCCataggagactaagagagacg-3'`
υ_3-TS: `5'-p-taccgaaagtatgaG-GATCCataggagactaagagagacg-3'`
υ_4-TS: `5'-p-tcgtgtgccgtgcaG-GATCCataggagactaagagagacg-3'`
υ_5-TS: `5'-p-tgtgaggttcgagaG-GATCCataggagactaagagagacg-3'`
υ_6-TS: `5'-p-ttgagcttgaaactG-GATCCataggagactaagagagacg-3'`,

where "b" and "p" designate a biotin and a phosphate group, respectively. The recognition sites of *Bam*HI and *Eco*RI are capitalized. The sites cleaved with these restriction enzymes are designated by "-".

A base sequence of single-stranded DNA oligomer representing an edge was the reverse complement of the two base sequences of connected cities lining up side by side. For example, the base sequence of edge oligomer connecting the start city (city 0) and city 1 is the reverse complement of the base sequence of υ_0-υ_1, and is designated by $rev(\upsilon_0\text{-}\upsilon_1)$. The base sequence of $rev(\upsilon_0\text{-}\upsilon_1)$ was thus `5'-cgggctgcgtaacaactcttaagtaggccg-3'`.

All DNA oligomers, synthesized on an automated DNA synthesizer and purified with an OPC cartridge, were obtained from Sawaday Technology, Inc. (Japan).

2.2. Experimental procedures for executing basic operations. The solid phase Hamiltonian path synthesis method was composed of seven basic operations: ASC, EXT, CAP, DEBLK, RMVDC, AMP and GPCR operations. ASC operation performs the attachment of the start city oligomer to a solid support. EXT operation extends path strands on a solid support by a city. CAP operation covers the $3'$ end of path strands on a solid support with an unreactive group to prohibit further extension of the paths. DEBLK operation removes the unreactive group to allow further extension of path strands. RMVDC operation removes path strands with a looping path, i.e. a path with cities visited twice, from a solid

support. AMP operation increases the number of path strands on a solid support. GPCR operation is a graduated PCR to display the order of cities on a path. All operations except GPCR were implemented by only simple and fast methods, which are amenable to full automation.

2.2.1. *ASC operation.* Single-stranded start city oligomer RS-v_0, which was biotinylated at the $5'$ end, was immobilized on streptavidin-coated magnetic beads. Twenty microliters (containing 20 μg) of Dynabeads M-280 Streptavidin (DYNAL A. S., Norway) were washed twice with 100 μl of binding and washing buffer (B&W buffer: TE buffer (10 mM Tris-HCl (pH 7.6), 0.1 mM EDTA) containing 1 M NaCl) using a Dynal MPC, and resuspended in 50 μl of B&W buffer. Ten pmol of biotinylated RS-v_0 in an equal volume of B&W buffer was added and incubated at 25℃ for 15 min while gently shaking the tube. The beads were then washed once with 100 μl of B&W buffer and twice with 100 μl of TE buffer at room temperature.

2.2.2. *EXT operation.* Beads were mixed with 1 pmol each of city oligomers v_i-TS ($i = 1, 2, \cdots, 6$) and edge oligomers *rev*(v_i-v_j) (i, j; all pairs of connected cities) in 50 μl of *Taq* ligation buffer (20 mM Tris-HCl (pH 7.6), 25 mM CH_3COOK, 10 mM $(CH_3COO)_2Mg$, 10 mM DTT, 1 mM NAD, 0.1% Triton X-100). To avoid a misligation reaction, the reaction mixture was heated up to 94℃ before an addition of 20 units of *Taq* DNA ligase (New England Biolabs, USA). The beads were then incubated at 65℃ for 10 min while gently shaking the tube. After the ligation reaction, the beads were washed twice with 100 μl of TE buffer at 94℃ to remove excess city and edge oligomers.

2.2.3. *CAP operation.* Beads were incubated at 37℃ for 10 min in 50 μl of cacodylate buffer (120 mM sodium cacodylate (pH 7.2), 1.2 mM $CoCl_2$, 0.6 mM 2-mercaptoethanol) containing 10 nmol of dideoxyribosyl thymidine triphosphate (ddTTP) and 20 units of terminal deoxynucleotidyl transferase (TdT) (TOYOBO, Japan). After the capping reaction, the beads were washed twice with 100 μl of TE buffer at room temperature.

2.2.4. *DEBLK operation.* Beads were mixed with 5 pmol each of *rev*(v_i-v_i-TS) oligomers ($i = 1, 2, \cdots, 5$) in 50 μl of H buffer (50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT). The reaction mixture was heated to 94℃ for 1 min and then cooled down to 37℃ in 5 min to facilitate annealing of the *rev*(v_i-v_i-TS) oligomers to path strands on the beads. After adding 200 units of *Bam*HI, the beads were incubated at 37℃ for 5 min. The beads were then washed twice with 100 μl of TE buffer at 94℃, making the path strands single-stranded.

2.2.5. *RMVDC operation.* Beads were suspended in 50 μl of Ex *Taq* buffer (TAKARA, Japan) containing 4 μl each of the 0.2 mM dNTPs and 5 pmol each of *rev*(v_i-v_i-TS) oligomers ($i = 1, 2, \cdots, 5$), which anneal to path strands on the beads. The reaction mixture was heated up to 94℃ and 1.25 units of Ex *Taq* DNA polymerase (TAKARA, Japan) was added to initiate the polymerization. The beads were then incubated at 80℃ for 1 min. After stopping the polymerization reaction with the addition of 2 μl of 500 mM EDTA, the beads were washed once with 100 μl of TE buffer at room temperature. The beads were then resuspended in 50 μl of H buffer and treated for 5 min at 37℃ with 200 units of *Eco*RI to remove looping path strands from the beads. After an *Eco*RI digestion, the beads were washed twice with 100 μl of TE buffer at 94℃, making the path strands single-stranded.

2.2.6. *AMP operation.* Beads were subjected to a PCR in 50 μl of Ex *Taq* buffer containing 4 μl each of the 0.2 mM dNTPs and 10 pmol each of RS and *rev*(TS) primers. PCR amplification was performed on UNOII Thermocycler (Biometra,

Germany) using the temperature profile of the initial denaturation at 94℃ for 1 min followed by 20 cycles of the denaturation at 94℃ for 20 sec, the annealing at 40℃ for 1 min, and the extension at 72℃ for 1 min. The hot start technique was used, so 1.25 units of Ex *Taq* DNA polymerase was added just after the initial denaturation step.

After the first PCR amplification, 1 μl of the supernatant was subjected to the second PCR in the same way as the first PCR but the RS primer biotinylated at the $5'$ end was used. After ethanol precipitation and removal of excess primers by centrifugal filtration, the PCR product was suspended in 50 μl of B&W buffer and then attached to 20 μg of streptavidin-coated magnetic beads through biotin at the $5'$ end of the RS-primer strand. The double- stranded path strands immobilized on the beads were denatured in 100 μl of TE buffer for 2 min at 94℃. Dissociated strands were washed away twice with 100 μl of TE buffer at 94℃.

2.2.7. *GPCR operation.* The length of a path and the order of cities on a path were analyzed using a PCR with appropriate primer pairs. To determine the length of a path, RS and *rev*(TS) primers were used. To determine the position of city i, RS (or RS-υ_0) and *rev*(υ_i) (or *rev*(υ_i-TS)) primers were used. For the analysis, the beads were suspended in 50 μl of Ex *Taq* buffer containing 4 μl each of the 0.2 mM dNTPs and 10 pmol each of the primers. PCR amplification was performed using the temperature profile of the initial denaturation at 94℃ for 1 min followed by 15-25 cycles of the denaturation at 94℃ for 20 sec, the annealing at 40℃ for 1 min, and the extension at 72℃ for 1 min. When RS-υ_0 and *rev*(υ_i-TS) primers were used, the annealing step was omitted. The hot start technique was used, so 1.25 units of Ex *Taq* DNA polymerase was added just after the initial denaturation step.

PCR products in the supernatant were resolved on a P/ACE5510 capillary electrophoresis system (Beckman, USA) using LIFluor dsDNA1000 Kit. The laser-induced fluorescence detection equipped with a P/ACE5510 system was used for the high sensitivity detection of PCR products. The intensity of fluorescence signal was calibrated using a double-stranded DNA solution, the concentration of which was determined by a UV absorbance at 260 nm. The peak area of 700 RFU·ETU (RFU: relative fluorescence unit; ETU: elution time unit) on an electropherogram corresponded to 20 fmol/ μl double-stranded DNA solution of 40 bp. The elution time of PCR products on an electropherogram was calibrated using the standard molecular weight marker supplied with LIFluor dsDNA1000 Kit.

3. Results

3.1. Solid phase Hamiltonian path synthesis. Sold phase DNA synthesis is an automated technique for the synthesis of polynucleotides, in which the chain grows while it is attached to a solid support. Our method for solving the HPP is very similar to the solid phase DNA synthesis (Fig. 2). Single-stranded DNA molecules representing paths without duplicate cities are synthesized city-by-city from the start city to the end city on the surface of a solid support. Excess DNA oligomers as well as enzymes and buffers are easily washed away without losses of path strands while the synthesis cycle is repeated. In the solid phase DNA synthesis, all nucleotides in a reaction mixture are coupled to the end of growing chains. In contrast, only those cities connected to the city at the end of growing paths are coupled in the solid phase path synthesis. This is a major difference

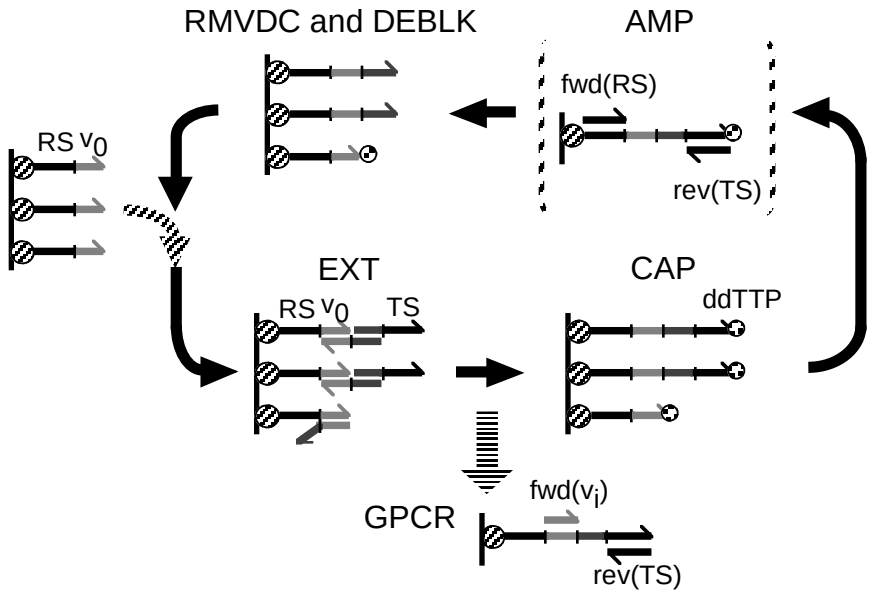

FIGURE 2. Solid phase Hamiltonian path sysnthesis

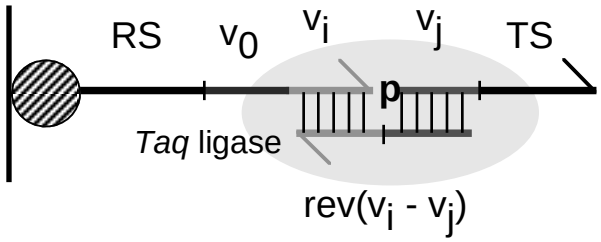

FIGURE 3. EXT operation

between the standard solid phase DNA synthesis and the solid phase Hamiltonian path synthesis.

As shown in Fig. 2, the process of the solid phase Hamiltonian path synthesis is initiated by attaching the $5'$ end of the start city oligomer RS-v_0 to the surface of a solid support. The start city oligomer is then coupled to connected city oligomers using EXT operation. *Taq* DNA ligase couples only connected cities to the $3'$ end of path strands because an edge oligomer $rev(v_i\text{-}v_j)$ stacks only a city oligomer v_j-TS at the $3'$ end of path strands terminating with v_i (Fig. 3). After the coupling reaction, the $3'$ end of path strands is capped with an unreactive group (ddTTP) using CAP operation in order to prohibit further path extension from the end.

When the length of paths increases to more than two cities, looping paths with duplicate cities begin to be synthesized. To remove the looping paths, RMVDC operation is performed after CAP operation. Single-stranded DNA oligomer $rev(v_i\text{-}v_i\text{-TS})$ is used to remove the duplication of city i. This oligomer has a tandem repeat of the complementary base sequence to city i. When a path has the duplication, the path strand and a tandem-city oligomer $rev(v_i\text{-}v_i\text{-TS})$ form a stable hybrid

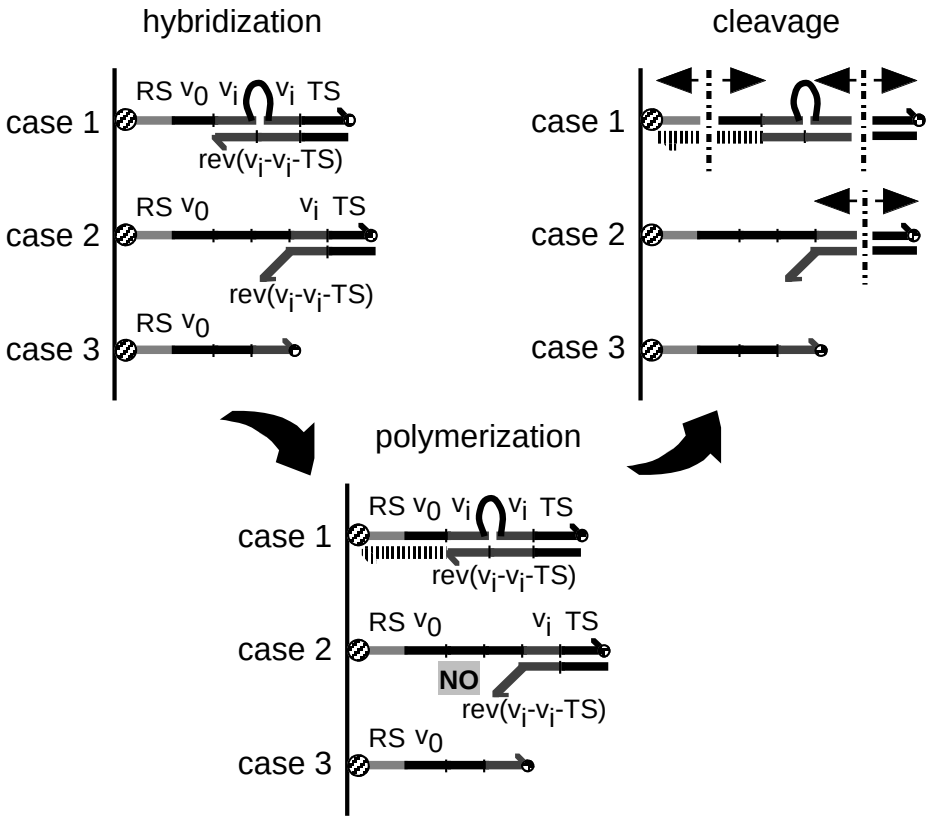

FIGURE 4. RMVDC and DEBLK operation

with a bulge loop (case 1 in Fig. 4). In the hybrid the 3′ end of $rev(\upsilon_i\text{-}\upsilon_i\text{-TS})$ is base-paired to the path strand. So extension of $rev(\upsilon_i\text{-}\upsilon_i\text{-TS})$ is initiated by *Taq* DNA polymerase, making an RS part double-stranded. The RS part at the root of a path strand has the recognition sequence of an *Eco*RI restriction enzyme. A path strand with the duplication is therefore removed from a solid support by *Eco*RI digestion. In contrast, when a path has no duplication, the path strand and $rev(\upsilon_i\text{-}\upsilon_i\text{-TS})$ oligomer form a stable hybrid without bulge loop (case 2 in Fig. 4). In the hybrid the 3′ end of $rev(\upsilon_i\text{-}\upsilon_i\text{-TS})$ is not base-paired to the path strand. So no extension of $rev(\upsilon_i\text{-}\upsilon_i\text{-TS})$ is initiated by *Taq* DNA polymerase, leaving the RS part single-stranded. A path without the duplication therefore remains on a solid support after executing RMVDC operation.

DEBLK operation follows RMVDC operation in order to advance to the next extension cycle. Annealing of $rev(\upsilon_i\text{-}\upsilon_i\text{-TS})$ oligomer to a path strand makes a TS part double-stranded. The TS part at the 3′ end of a path strand has the recognition sequence of a *Bam*HI restriction enzyme. The 3′ end of a path strand is thus activated to the next path extension by cutting the TS part off with *Bam*HI (case 2 in Fig. 4). In contrast, a path strand failed to be extended by EXT operation has no TS part, thus staying inactive to the next path extension (case 3 in Fig. 4). Restriction enzymes *Eco*RI and *Bam*HI used in RMVDC and DEBLK operations,

respectively, have the activity in the same buffer (H buffer). RMVDC and DEBLK operations can thus be executed at once as shown in Fig. 4. The simultaneous execution helps the reduction of the computation time.

The pruning of looping paths with RMVDC operation as well as the blocking of failed paths with CAP operation gradually reduce the number of path strands on a solid support. AMP operation should be carried out when the number of path strands remaining on a solid support is suspected to be too small to cover the combinatorial variety of intermediate paths. AMP operation increases the number of path strands on a solid support

EXT, CAP, (AMP), RMVDC and DEBLK operations are repeated until the length of paths extends to the number of cities (vertices) in the graph. Finally, the order of cities on the path is displayed by GPCR operation (graduated PCR) to get a solution to an instance of the HPP. Therefore, the HPP can be solved in the computation time proportional to the number of cities.

3.2. Accuracy and efficiency of EXT operation. EXT operation extends the length of paths by a city. Only connected cities are coupled to the 3′ end of path strands on a solid support because an edge oligomer $rev(v_i\text{-}v_j)$ stacks only a city oligomer v_j-TS at the 3′ end of path strands terminating with city i. The accuracy of this coupling reaction performed with a *Taq* DNA ligase determines the accuracy of EXT operation. For correctly solving the HPP using the present method, the high accuracy of EXT operation is essential. Thus we examined how accurate EXT operation was.

To examine the accuracy of EXT operation, we analyzed the order of cities on path strands using GPCR operation just after the first EXT operation. In Adleman's graph (Fig. 1) the start city (city 0) is connected to city 1, 3 and 6. So only such path strands as RS-v_0-v_1-TS, RS-v_0-v_3-TS and RS-v_0-v_6-TS are generated after the first EXT operation if a ligation reaction in EXT operation correctly couples the connected cities. As shown in Fig. 5, only those correct path strands were generated. GPCR operations using RS-v_0 and $rev(v_i$-TS) (i = 1, 3 or 6) primers gave the product of 80 base pairs eluted around an elution time of 14 min. In contrast, other GPCR operations using RS-v_0 and $rev(v_j$ -TS) (j = 2, 4 or 5) primers did not produce any significantly amplified DNA fragments. Therefore, EXT operation was shown to be accurate enough for DNA-based computation of the HPP using the present method when the ligation reaction was carried out at 65℃. At a lower temperature, the condition of the hybridization between edge and city oligomers is less stringent. So the accuracy of EXT operation might be reduced at a lower temperature. However, an EXT operation performed at 60℃ still showed the accurate path extension (data not shown).

In addition to the accuracy of EXT operation, its efficiency is important for the performance of DNA-based computation using the present method. The efficiency of EXT operation was measured by the peak area of a PCR product obtained by GPCR operation performed after the EXT operation. The peak area is likely to reflect the amount of a path strand generated by EXT operation because the number of PCR cycles used in the GPCR operation was as small as 15. Figure 5 shows that the peak area of a PCR product RS-v_0-v_1-TS was larger than that of the other two PCR products, RS-v_0-v_3-TS and RS-v_0-v_6-TS. Therefore, a city oligomer v_1-TS was more efficiently coupled to the 3′ end of RS-v_0 on beads than the other two city oligomers, v_3-TS and v_6-TS. The GC- content of the base sequence of city 1

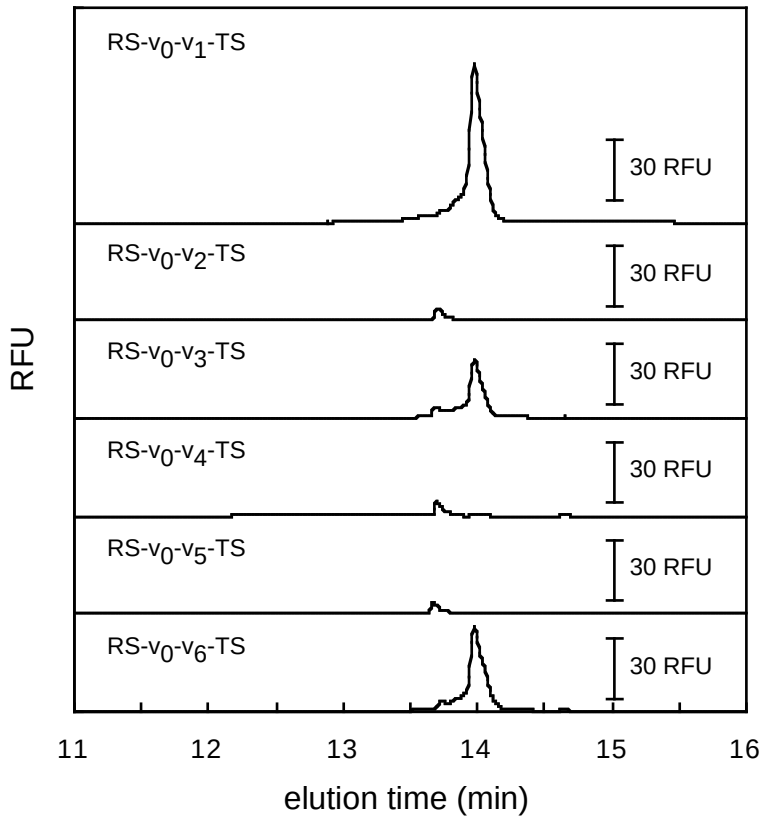

FIGURE 5. Electropherograms of graduated PCR products obtained by GPCR operation performed after the first EXT operation. The path sequence of a PCR product is designated at the upper left of each electropherogram.

is 0.6, which is higher than that of city 3 (0.4) and city 6 (0.4). An edge oligomer $rev(v_0\text{-}v_1)$ thus is able to form a more stable hybrid duplex with a city oligomer v_1-TS than with the other two city oligomers. This may cause the difference in the coupling efficiency.

The fact that the efficiency of EXT operation depends on the GC-content of a city oligomer means that the path search in a graph does not proceed uniformly; some paths are more preferentially generated than others. This biased search reduces the performance of DNA-based computation of the HPP using the present method. The design of city oligomers with more uniform GC-content may solve this problem.

3.3. Efficiency of CAP operation. For correctly solving the HPP using the present method, the efficiency of CAP operation is important in addition to the accuracy of EXT operation. CAP operation covers the $3'$ end of path strand on a solid support with an unreactive group to prohibit further extension of the path strand. The unreactive group is removed by DEBLK operation for a path strand with a TS part, while it is not removed for a path strand without a TS part. A

FIGURE 6. Electropherogram of graduated PCR products obtained by GPCR operation that follows EXT operation. Beads were subjected to CAP operation (broken line) or not (solid line) in advance of the EXT operation.

path strand that has failed to be extended by the previous EXT operation has no TS part. CAP operation thus prohibits the failed path strand from extending in the next EXT operation. The extension of the failed path strand causes errors in the solution of the HPP. Therefore, a high efficiency of CAP operation is essential for the solution of the HPP using the present method.

The efficiency of CAP operation was examined by treating beads with CAP operation in advance of EXT operation. The beads had the start city oligomer RS-v_0 on the surface. If the efficiency of CAP operation is high enough to cover completely the $3'$ end of a RS-v_0 strand with an unreactive group, no extension of the strand occurs in the following EXT operation. Figure 6 shows that the amount of an 80-bp PCR product was greatly reduced when the beads were treated with CAP operation in advance of EXT operation. The 80-bp PCR product corresponds to a path strand RS-v_0-v_1-TS, which is generated by extension of a path strand RS-v_0. The number of PCR cycles in the GPCR operation was as small as 15 to ensure the quantitative PCR. Therefore, a large reduction in the amount of the 80-bp PCR product means that the efficiency of CAP operation was sufficiently high to suppress the path strand extension although it was unlikely to be 100%.

3.4. Efficiency of DEBLK operation. An advance to the next extension cycle requires that DEBLK operation activates the $3'$ end of path strands by removing an unreactive group at the end. Without the activation, no EXT operation extends path strands. Generation of long path strands thus requires a high efficiency of DEBLK operation.

The efficiency of DEBLK operation was examined by repeating the solid phase path synthesis cycle. After the first EXT operation, one-fiftieth of beads were analyzed by GPCR operation to display the sequences of path strands on the beads. As shown in Fig. 7a, a GPCR operation using the primer pair of RS-v_0 and rev(v_1-TS) gave a product of 80 base pairs eluted at 14.0 min, which corresponds to a path

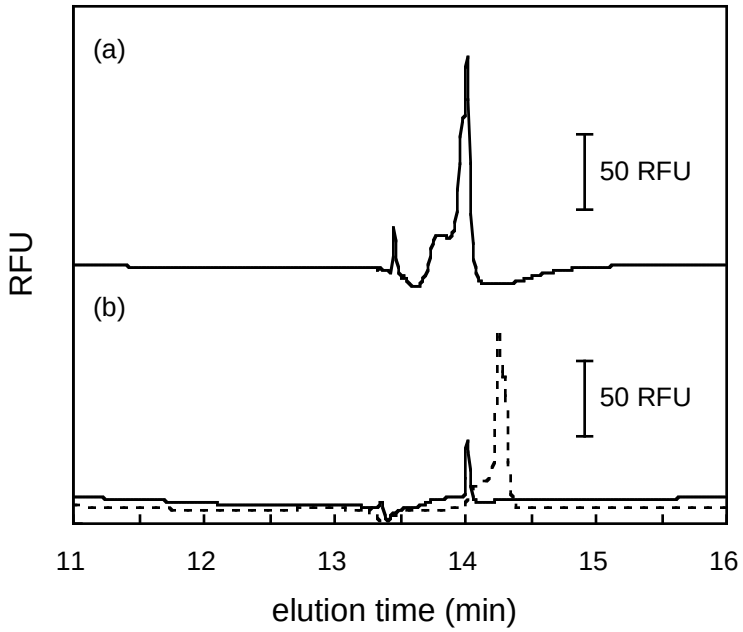

FIGURE 7. Electropherograms of graduated PCR products after the first cycle (a) and the second cycle (b) of the path synthesis. Graduated PCRs were performed using a primer pair of RS-v_0 and *rev*(v_1-TS) (solid line) or that of RS-v_0 and *rev*(v_2-TS) (broken line).

strand RS-v_0-v_1-TS. The rest of the beads were then subjected to CAP, DEBLK and EXT operations. After the second EXT operation, a GPCR operation using the primer pair of RS-v_0 and *rev*(v_2-TS) gave a product of 95 base pairs eluted at 14.3 min, which corresponds to a path strand RS-v_0-v_1-v_2-TS (Fig. 7b). The occurrence of the 95-bp strand proves that DEBLK operation activated the $3'$ end of path strands to some extent. On the other hand, a GPCR operation using RS-v_0 and *rev*(v_1-TS) primers gave a product of 85 base pairs, which corresponds to a path strand RS-v_0-v_1-TS (Fig. 7b). This fact means that some of path strands remained unactivated after DEBLK operation. The peak size of the 95-bp product is about four times as large as that of the 85-bp product. The number of PCR cycles used in the GPCR operation was as small as 15 to ensure the quantitative graduated PCR. Therefore, about 80% of the path strand on the beads were deblocked by DEBLK operation.

3.5. Free energy consideration on RMVDC operation. RMVDC operation utilizes the structure of a hybrid formed between a path strand and a tandem-city oligomer *rev*(v_i-v_i-TS). Elimination of path strands with duplicate cities by RMVDC operation requires the stable formation of the hybrid with a bulge loop (Fig. 4). Before starting extensive experiments on RMVDC operation, we theoretically examined the stability of a hybrid formed between a path strand and the tandem-city oligomer.

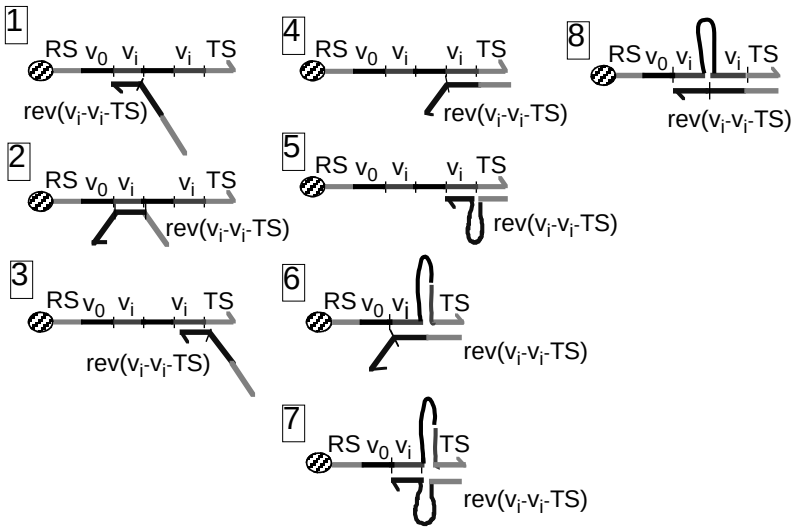

FIGURE 8. Possible hybrid structure formed between a path strand and a tandem-city oligomer $rev(v_i\text{-}v_i\text{-TS})$.

Figure 8 shows the structure of a possible hybrid formed between a path strand and a $rev(v_i\text{-}v_i\text{-TS})$ oligomer. In hybrids 1, 3, 5, 7 and 8, *Taq* DNA polymerase can initiate extension of the $rev(v_i\text{-}v_i\text{-TS})$ oligomer because the $3'$ end part of the oligomer is base-paired to the path strand. Among these hybrids, only hybrid 8 involves the duplicate parts on the path strand. Therefore, hybrid 8 must be the most stable among them to remove correctly a looping path using RMVDC operation in the way shown in Fig. 4.

The stability of a hybrid depends on the free energy of the hybrid. The lower the free energy is, the higher the stability is. The free energy of a hybrid is the sum of the free energy of each part composing the hybrid. In general, single-stranded regions are regarded as the standard state of the free energy calculation; thus a zero free energy is assigned to them. Double-helical regions have a negative free energy; these regions stabilize the hybrid. Loops such as bulge, internal and bifurcation loops have a positive free energy; thus the loops destabilize the hybrid. The free energy of a double-helical region is approximately proportional to the length of the region because the free energy is the sum of the free energy of each base-pair stack making the region up. The free energy of a loop increases monotonously with its size; thus a larger loop destabilizes the hybrid more extensively. From this rough consideration, the following order of the stability of the possible hybrids was derived. Hybrid 8 is most stable; hybrid 4 is less stable than it. Hybrids 5, 6 and 7 follow hybrid 4, leaving hybrids 1, 2 and 3 least stable.

To be more precise, hybrid 8 is most stable only when the length of v_i is large enough to compensate the free energy increase caused by the bulge loop formation. Otherwise, the most stable is not hybrid 8 but hybrid 4. Let the length of v_i be n bases. Then the free energy of the double-helical region of a v_i part is given by $-2 \times (n-1)$ kcal/mol, assuming that the base sequence of v_i has the uniform and average stability of base-stacking [**4**]. On the other hand, the free energy of a bulge loop of m bases long is given by $+1.5 \ln m$ for large m as shown in Fig. 9

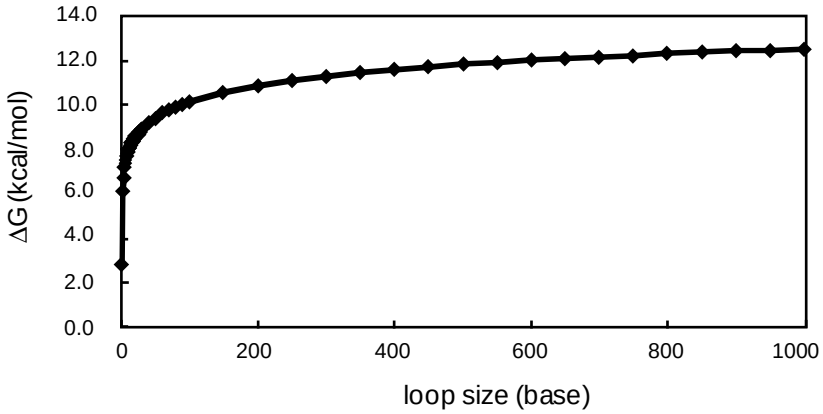

FIGURE 9. Free energy of a bulge loop formation at 25 ℃ in $[Na^+] = 1\,M$, pH 7.

[**5**]. Therefore, for a bulge loop of up to 1000 bases long, the v_i of 15 bases long is sufficient to make hybrid 8 most stable.

4. Discussion

The present method for solving the HPP is based on a solid phase chemistry and thus is similar to the surface-based approach to a memory for Boolean circuits [**6**]. The method has three advantages over Adleman's method [**1**], which has pioneered the field of DNA-based computation. First, the method is actually much faster than Adleman's method. Only simple and fast DNA operations were employed to solve the problem. Secondly, it can be performed on a fully automated machine. The full automation is essential not only for the speedup of computation but also for the error-free computation. Thirdly, the number of DNA molecules necessary for the computation is extensively reduced. Adleman's method generates all possible paths in a graph at once. The generated paths thus contain many straight and looping paths of any length, which do not always start with the start city nor end with the end city. In contrast, the present method stepwise generates only no-looping paths that start exactly with the start city. The extensive reduction allows the present method to be applied to a larger instance of the HPP.

The present method has been developed for solving the HPP. However, it could be applied to other computational problems. The method is composed of seven basic operations to manipulate DNA molecules on the surface of a solid support. These basic operations have some generality. Thus they can be used for the implementations of other DNA-based algorithms. For example, the satisfaction problem (SAT) could be solved efficiently by using some of these basic operations.

Experiments to examine the feasibility of the present method using Adleman's seven-city instance of the HPP is still going on. Especially, the full results of the experiments on RMVDC operation were not mentioned in the present study because many details on the experimental procedure of RMVDC operation still remain to be improved. These improvements are fixed soon and then we will be ready for starting the experiment to solve a larger instance of the HPP. The solid

phase method thus would go a step toward the realization of feasible DNA-based computers.

References

[1] L. M. Adleman, *Molecular Computation of Solutions to Combinatorial Problems*, Science **266** (1994), 1021-1024.
[2] R. J. Lipton, *DNA Solutions of Hard Computational Problems*, Science **268** (1995), 542-544.
[3] R. J. Lipton, E. B. Baum (eds.) *DNA based computers: Proc. of a DIMACS workshop, April 4, 1995*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science vol. 27, Amer. Math. Soc., Providence, RI, 1996.
[4] K. J. Breslauer, R. Frank, H. Blöcker, L. A. Marky, *Predicting DNA duplex stability from the base sequence*, Proc. Natl. Acad. Sci. USA **83** (1986), 3746-3750.
[5] C. DeLisi, D. M. Crothers, *Prediction of RNA Secondary Structure*, Proc. Natl. Acad. Sci. USA **68** (1971), 2682-2685.
[6] W. Cai, A. E. Condon, R. M. Corn, E. Glaser, Z. Fei, T. Frutos, Z. Guo, M. G. Lagally, Q. Liu, L. M. Smith, A. Thiel, *The Power of Surface-based DNA Computation*, Proc. RECOMB'97 (1997), 67-74.

Institute of Physics, Graduate School of Arts and Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan
E-mail address: morimo@genta.c.u-tokyo.ac.jp

Department of Information Science, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
E-mail address: arita@is.s.u-tokyo.ac.jp

Institute of Physics, Graduate School of Arts and Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan
E-mail address: suyama@dna.c.u-tokyo.ac.jp