

A DNA SEQUENCE APPROACH OF DIRECT-PROPORTIONAL LENGTH-BASED DNA COMPUTING BY DNA SEQUENCE GENERATOR



Kumar Parbhav Rajan

Research Scholar, Singhania University
Rajasthan, India

GNITED MINDS
Journals

Abstract

DNA computing has emerged as an interdisciplinary field that draws together molecular biology, chemistry, computer science, and mathematics. The fundamental of this unconventional computation has been proven to solve weighted graph problems, such as the shortest path problem and the travelling salesman problem. One of the fundamental improvements in DNA computing is direct-proportional length-based DNA computing for the shortest path problem. Generally, in DNA computing, the DNA sequences used for the computation should be critically approached in order to reduce error that could occur during computation. In this paper, a procedure to approach the DNA sequences for the direct-proportional length-based DNA computing is presented. The

procedure includes DNASequenceGenerator, a graph-based approach for approaching a set of good DNA sequences.

1. INTRODUCTION

A new computing paradigm based on DNA computing has appeared in 1994 when Leonard M. Adleman [Adleman, 1994] launched a novel *in vitro* approach to solve the so-called Hamiltonian path problem (HPP) with seven vertices by DNA molecules. While in conventional silicon-based computer, information is stored as binary numbers in silicon-based memory, he encoded the information of the vertices by generating randomized DNA sequences. The computation is performed by a series of primitive bio-molecular reactions involving hybridization, denaturation, ligation, magnetic bead separation, and polymerase chain reaction (PCR). The output of the computation, also in the form of DNA molecules can be read and “printed” by electrophoretical fluorescene method such as agarose gel electrophoresis or polyacrylamide gel electrophoresis (PAGE).

In the previous paper, an alternative molecular computing approach for weighted graph problem, which is called directproportional length-based DNA computing (DPLBDNAC) for the shortest path problem, has been proposed. Based on this approach, the cost of an edge is encoded as a directly proportional length oligonucleotides (oligos). After the initial pool generation and amplification, since numerous numbers of solution candidates are generated, by applying a series of bio-molecular operations, it is possible to extract the optimal combination which represents the solution to the shortest path problem. Also, the implementation of DPLB-DNAC is realized by laboratory experiments and several aspects of the experiments, such as the initial pool generation

method employed and the correctness of the proposed encoding rules are experimentally investigated.

Various kinds of methods and strategies for DNA sequence approach have been proposed to date. As reviewed by Shin *et al.* [Shin *et al*, 2005], those strategies are exhaustive search, random search, template-map strategy, graph method, stochastic methods, dynamic programming, biological-inspired methods and evolutionary algorithms.

DNA sequences used for the computation based on DPLB-DNAC should be carefully determined in order to reduce errors that could occur during the computation. In this paper, DNASequenceGenerator [Feldkamp *et al*, 2001], which is based on the graph method, is employed for approaching DNA sequences for DPLB-DNAC. DNASequenceGenerator used a directed graph to approach DNA sequences. The nodes in the graph represent base strands and a node has four strands that can appear as successors in a longer sequence as its child nodes. Then, by travelling the graph from root to leaf the DNA sequences can be approached. This approach also is able to find a set of orthogonal DNA sequences within a predefined error rate quickly as shown in Fig.1. The usefulness of the generated DNA sequences is demonstrated by experimentally implementation of the directproportional length-based DNA computing for the shortest path problem.

2. THE SHORTEST PATH PROBLEM

The input to the shortest path problem is a weighted directed graph $G = (V, E, \omega)$, a start node u , and an end node v . The output of the shortest path problem is a (u, v) path with the

smallest cost. In the case given in Fig.2, if u is V_1 and v is V_5 , the cost for the shortest path will be given as 100 and the optimal path is clearly shown as $V_1-V_3-V_4-V_5$. Even though the shortest path problem is belonging to the class P, it is worthy of being solved by DNA computing because numerical evaluation is involved during the computation.

3. PROCEDURE OF THE DNA SEQUENCE APPROACH

For the DNA sequence approach, several constraints adapted from [Innis and Gelfand, 1990] are considered as follows:

- Primers should be 17-28 bases in length
- Base composition should be 50-60% (G+C)
- Primers should end (3') in a G or C, or CG or GC to prevents “breathing” of ends and increases efficiency of priming
- Melting temperature T_m between 55-80°C are preferred
- Runs of three or more Cs or Gs at the 3'-ends of primers may promote mispriming at G or C-rich sequences (because of stability of annealing) and should be avoided.

These constraints should be critically considered during the DNA sequence approach. In this paper, the DNA sequences used for the computation are approached by using the DNASequenceGenerator, a program for constructions of DNA sequences, which can be freely downloaded at <http://ls11-www.cs.unidortmund.de/molcomp>. DNASequence-Generator uses a concept of uniqueness that, within a pool of sequences, allows any subsequence of a certain definable length to occur at most once in that pool.

The first step of the DNA sequence approach is to generate five unique single-stranded DNA sequences for each node in the graph which satisfy the previous constraints. During the generation of DNA sequences, the constraint for GC percentage (GC%) is chosen within the range of 50-55%. On the other hand, the melting temperature, T_m , which is calculated using the Sugimoto thermodynamic parameters [Sugimoto *et al*, 1996] is set around 60°C. Five DNA sequences and the complements for each node which satisfy the constraints are listed in Table 1.

In order to formulate the length constraint in the length-based DNA computing, let $|V|$ be the total number of nodes in the graph. V_i ($i = 1, 2, \dots, |V|$) and \bar{V}_i ($i = 1, 2, \dots, |V|$) be the 20-mer DNA sequences correspond to the i th node in the graph and its complement, respectively. Three rules to synthesize oligos for each edge in the graph are approached as follows:

- 1) For a connection between V_1 to V_j , synthesize the oligo for the edge as $V_1(20) + \bar{W}_{1j}(\omega - 30) + V_j(20)$
- 2) For a connection between V_i to V_j , where $i \neq 1, j \neq |V|$, synthesize the oligo for the edge as $V_i(20) + W_{ij}(\omega - 20) + V_j(20)$
- 3) For is a connection between V_i to $V_{|V|}$, synthesize the oligo for the edge as $V_i(20) + W_{in}(\omega - 30) + V_{|V|}(20)$

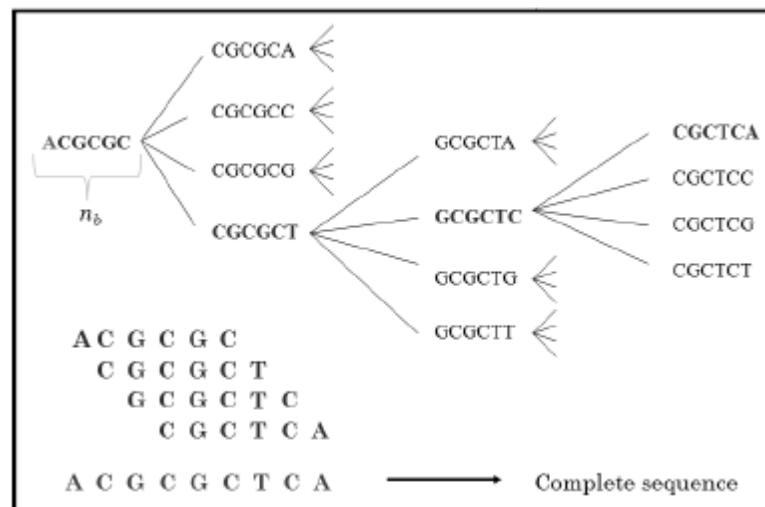
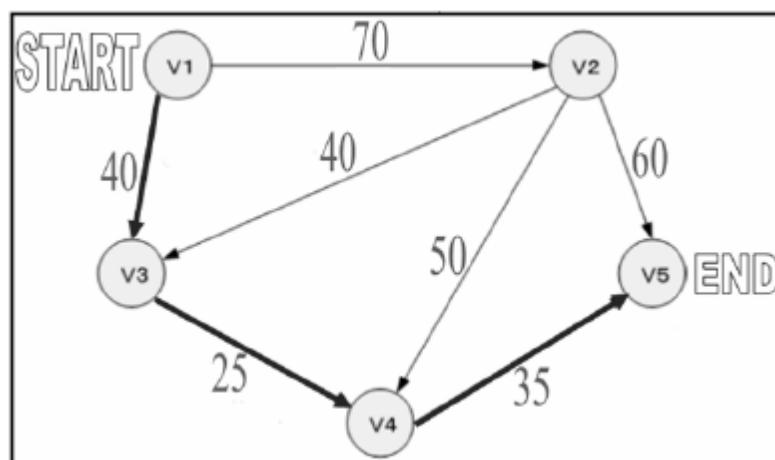


Fig 1. Graph method in DNASequenceGenerator

Fig.2. A weighted undirected graph $G = (V, E)$

Node	Sequences (5'-3')	Complements (5'-3')
V_1	AAAGCTCGTCGTTAGGAGC	GCTCCTAACGACGAGCTT
V_2	GCACTAGGGATTGGAGGTT	AACCTCAAATCCCTAGTGC
V_3	GCTATGCCGTAGTAGAGCGA	TCGCTCTACTACGGCATAGC
V_4	CGATACCGAACTGATAAGCG	CGCTTATCAGTCGGTATCG
V_5	CGTGGGTGGCTCTGTAATAG	CTATTACAGAGGCCACCCACG

Table 1. DNA sequences approached for nodes.

where V , W , and ‘+’ denote the DNA sequences for nodes, DNA sequences for weight, and ‘joint’ respectively. Furthermore, ‘ ω ’ denotes the weight value for corresponding DNA sequences for weight W_{ij} where W_{ij} denotes the DNA sequences representing a cost between node V_i and V_j . The value in parenthesis indicates the number of DNA bases or nucleotides for each segment. Since the DNA sequences for weight W_{ij} are not involved during hybridization of initial pool generation, the constraints for these sequences are relaxed for the generation of DNA sequences based on DNASequenceGenerator.

For easier understanding, Fig.3, Fig. 4, and Fig.5 visualize how each edge is encoded. The resultant DNA sequences for edges approached based on the rules are listed in Table 2. The synthesized oligos consist of three segments; two node segments and an edge segment.

4. EXPERIMENTAL IMPLEMENTATION

After the DNA sequences are approached, the oligos of the complement of the node sequences and edges sequences are synthesized. Then, all the synthesized oligos are poured into a test tube for initial pool generation via parallel overlap assembly (POA).

To describe graphically how the initial pool can be generated by POA, the double stranded DNA (dsDNA) in Fig. 6, represents the shortest path of the shortest path problem. Several single stranded DNAs (ssDNA), which are shown by thin lines, are required for the generation of that dsDNA. In this example, 3 cycles are required in order to generate the target dsDNA. The output of the first stage, second stage, and third stage of POA for generating the dsDNAs of the shortest path are shown in Fig.7, Fig.8, and Fig. 9, respectively.

The initial pool generation by POA was performed in a $100 \mu\text{l}$ solution containing $12 \mu\text{l}$ oligos (Proligo Primers & Probes, USA), $10 \mu\text{l}$ dNTP (TOYOBO, Japan), $10 \mu\text{l}$ $10\times$ KOD dash buffer (TOYOBO, Japan), $0.5 \mu\text{l}$ KOD dash (TOYOBO, Japan), and $67.5 \mu\text{l}$ ddH₂O (Maxim Biotech, Inc., Japan). The reaction consisted of 25 cycles and for each cycles, the appropriate temperature were as follows:

- 94°C for 30s
- 55°C for 30s
- 74°C for 10s

Next, the generated initial pool generation is subjected to amplification by PCR in order to amplify exponentially, DNA molecules that contain the start node V1 and end node V5. After the

PCR is accomplished, there should be a big numbers of DNA molecules representing the start node V_1 and end node V_5 travelling through a possible number of nodes. Four types of expected amplified dsDNAs after PCR are given in Fig.10.



Fig.3. DNA encoding for edges based on rule (i)

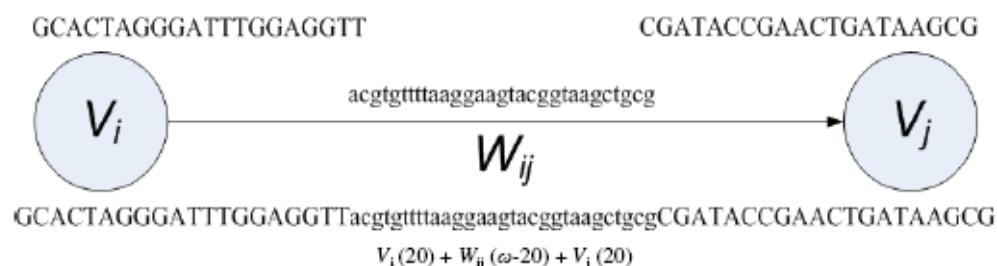


Fig.4. DNA encoding for edges based on rule (ii)

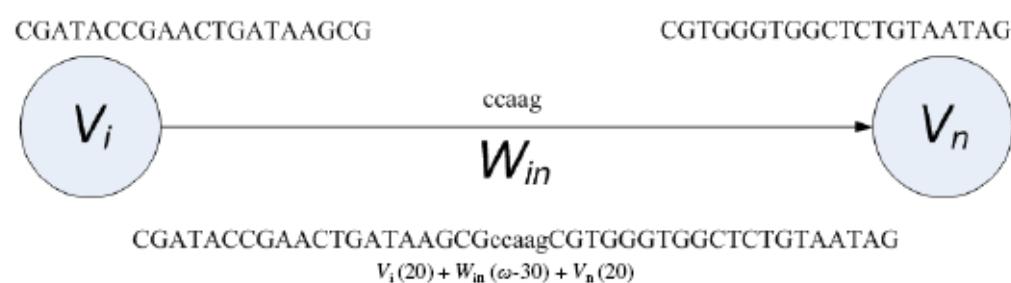


Fig.5. DNA encoding for edges based on rule (iii)

IDS
urnals

For amplification, PCR was performed in a $25 \mu\text{l}$ solution consists of $0.5 \mu\text{l}$ for each of forward and reverse primers, $1 \mu\text{l}$ template, $2.5 \mu\text{l}$ dNTP (TOYOBO, Japan), $2.5 \mu\text{l}$ 10x KOD dash buffer (TOYOBO, Japan), $0.125 \mu\text{l}$ KOD dash (TOYOBO, Japan), and $17.875 \mu\text{l}$ ddH₂O (Maxim Biotech Inc., Japan). The reaction consisted of 25 cycles and for each cycles, the appropriate temperature were as follows:

- 94°C for 30s
- 55°C for 30s
- 74°C for 10s

The sequences used as forward and reverse primers, as generated by DNASequenceGenerator, were CCTTAGTAGTCATCCAGACC (VI) and CCACTGGTTCTGCATGTAAC (5 V), respectively. Based on DPLB-DNAC, the output solution of PCR is brought for gel electrophoresis. During this reaction, the DNA molecules were separated in term of its length and hence, the shortest DNA molecules in terms of length in base-pairs (bp), which representing the shortest path could appear as the shortest band of the output of gel electrophoresis as shown in Fig.11.

Edges	DNA Sequences (5'-3')
$V_4-W_{45}-V_5$	CGATACCGAACTGATAAGCGccaagCGTGGGTGGCTCTGTAATAG
$V_3-W_{34}-V_4$	GCTATGCCGTAGTAGAGCGAccgtcCGATACCGAACTGATAAGCG
$V_1-W_{13}-V_3$	AAAGCTCGTCGTTAGGAGCacgtcggtcGCTATGCCGTAGAGCGA
$V_2-W_{23}-V_3$	GCACTAGGGATTGGAGGTTccgtctttaccaagtaatGCTATGCCGTAGAGCGA
$V_2-W_{24}-V_4$	GCACTAGGGATTGGAGGTTacgtgtttaaaggaaagtacggtaagctcgCGATACCGAACTGATAAGCG
$V_2-W_{25}-V_5$	GCACTAGGGATTGGAGGTTccgtcgcttaaggcagtacggactcgccCGTGGGTGGCTCTGTAATAG
$V_1-W_{12}-V_2$	AAAGCTCGTCGTTAGGAGCgggttaacgaagtctgtactatgggtattgcagGCACTAGGGATTGGAGGTT

Table 2. DNA sequences approached for edges.

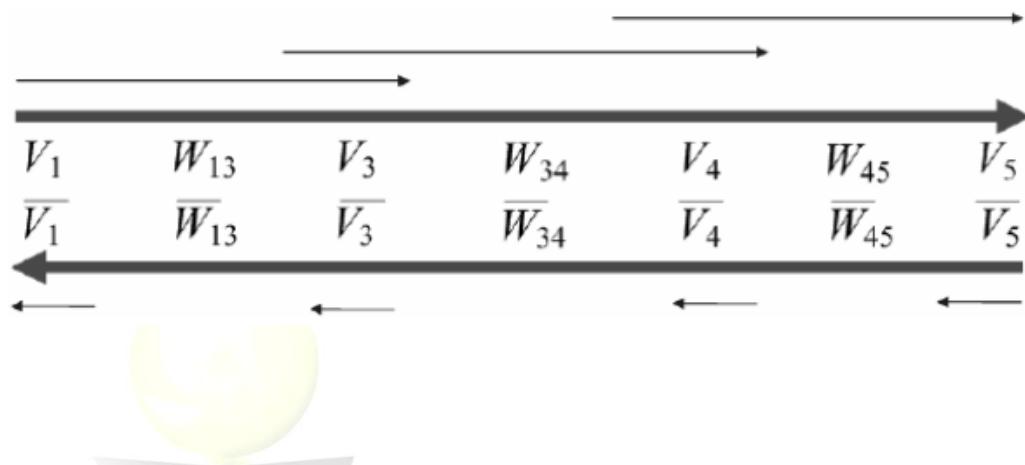


Fig.6. Example of a dsDNA representing the answer of the shortest path problem

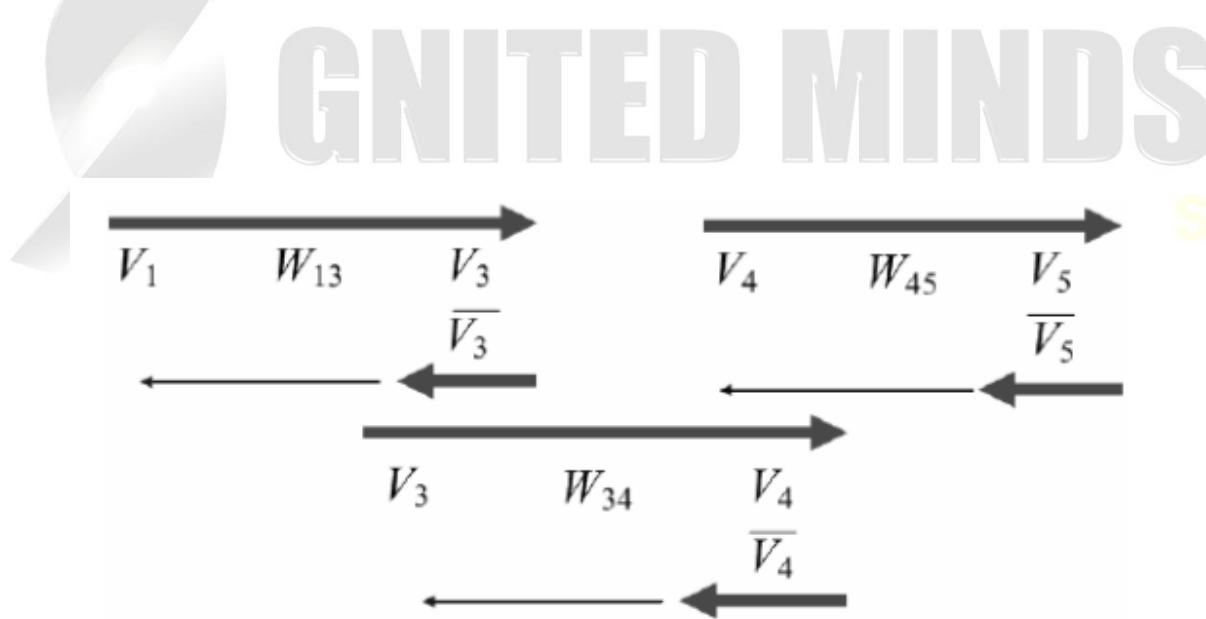


Fig.7. The first stage of POA

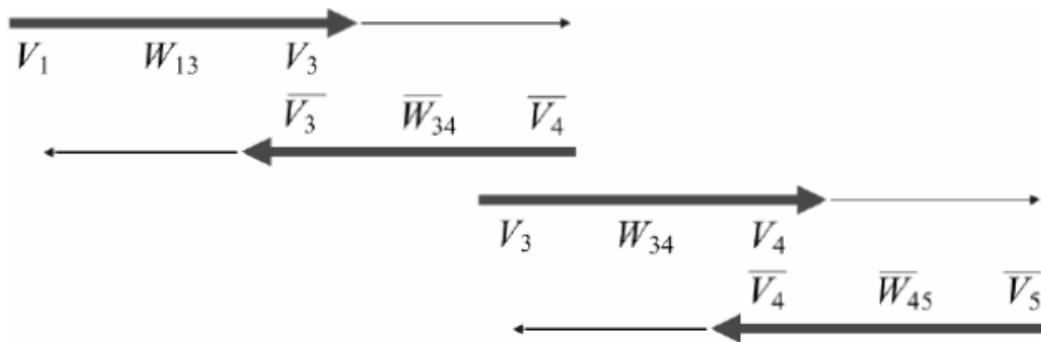


Fig.8. The second stage of POA

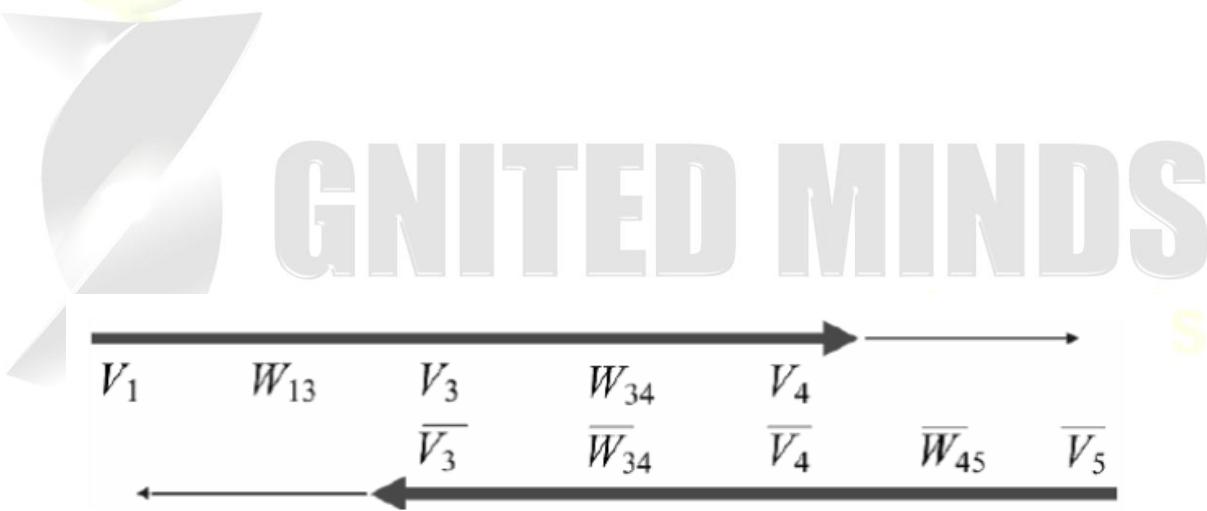


Fig.9. The third stage of POA

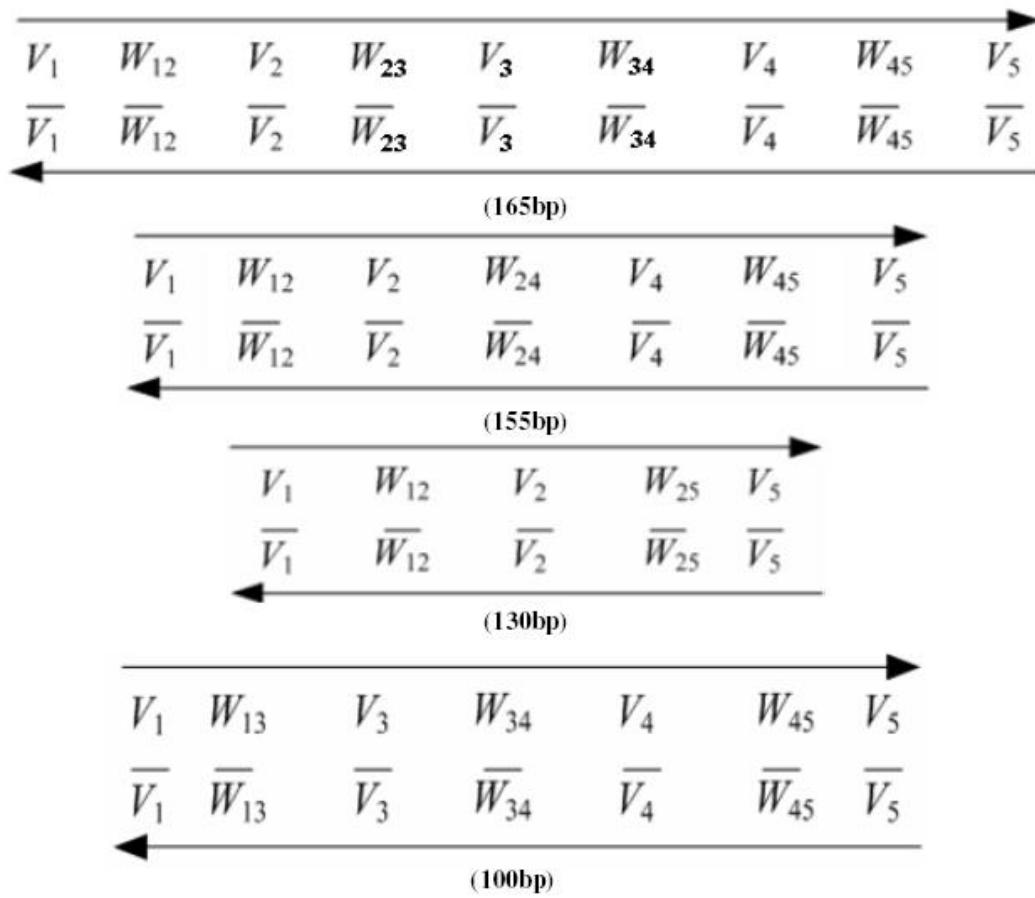


Fig.10. Examples of dsDNAs amplified by PCR. The length of dsDNAs in base base- airs (bp) is given in parenthesis and the arrowhead indicates the 3' end.

5. CONCLUSIONS

By using DNASequenceGenerator, a set of usable DNA sequences is generated for the experimental implementation of the direct-proportional length based DNA computing. The DNA sequences are belong to two groups: sequences for nodes and sequences for edges. The experimental results proved that the sequence approach strategy, assisted by

DNASequenceGenerator, can be employed for the implementation of direct-proportional length-based DNA computing.



Fig.11. Experimental results of gel electrophoresis on 10% polyacrylamide gel. Lane M denotes 20-bp ladder, lane 1 is the product of PCR, which is the output of the shortest path problem, and lane 2 is the product of parallel overlap assembly for initial pool generation.

REFERENCES

1. Adleman, L. "Molecular computation of solutions to combinatorial problems", *Science*, Vol. 266, 1994, pp. 1021-1024
2. Arita, M. Nishikawa, A. Hagiya, M. Komiya, K. Gouzu, H. and Sakamoto, K. "Improving sequence approach for DNA computing", *Proceedings of Genetic and Evolutionary Computation Conference*, 2000, pp. 875–882
3. Arita, M. and Kobayashi, S. "DNA sequence approach using templates", *New Generation Computing*, Vol. 20, 2002, pp. 263–277
4. Deaton, R. Garzon, M. Murphy, R.C. Rose, J.A. Franceschetti, D.R. and Stevens, Jr., S.E. "Reliability and efficiency of a DNA-based computation", *Physical Review Letters*, Vol. 80, No. 2, 1998, pp. 417-420
5. Deaton, R. Chen, J. Bi, H. Garzon, M. Rubin, H. and Wood, D.H. "A PCR based protocol for in vitro selection of noncrosshybridizing oligonucleotides", *Proceedings of the 8th International Workshop DNA Based Computers*, 2002, pp. 196–204
6. Feldkamp, U. Saghafi, S. Banzhaf, W. and Rauhe, H. "DNA sequence generator - A program for the construction of DNA sequences", *Proceedings of the 7th International Workshop DNA Based Computers*, 2001, pp. 179–188

7. Frutos, A.G. Thiel, A.J. Condon, A.E. Smith, L.M. and Corn, R.M. "DNA computing at surfaces: Four base mismatch word approach", *Proceedings of the 3rd DIMACS Workshop DNA Based Computers*, 1997, pp. 238

8. Hartemink, A.J. Gifford, D.K. and Khodor, J. "Automated constraint based nucleotide sequence selection for DNA computation", *Proceedings of the 4th DIMACS Workshop DNA Based Computers*, 1998, pp. 227-235 36.

9. Heitsch, C.E. Condon, A.E. and Hoos, H.H. "From RNA secondary structure to coding theory: A combinatorial approach", *Proceedings of the 8th International Workshop DNA Based Computers*, 2002, pp. 215-228

10. Ibrahim, Z. Tsuboi, Y. Ono, O and Khalid, M. "In Vitro Implementation of k-shortest Paths Computation with Graduated PCR", *International Journal of Intelligence Research*, Vol.1, No.2, 2005, pp.127-137

11. Ibrahim, Z. Tsuboi. Y and Ono, O. " Hybridizationligation versus Parallel Overlap Assembly: An Experimental Comparison of Initial Pool Generation for Direct-Proportional Length-Based DNA Computing " , *IEEE Transactions on NanoBioscience*, Vol. 5, No. 2, June 2006, pp. 103- 109

12. Innis, M.A. and Gelfand, D.H. "Optimization of PCRs, in PCR protocols", *Academic Press*, New York, pp. 3-12, 1990

13. Marathe, A. Condon, A.E. and Corn, R.M. "On combinatorial DNA word approach", *Proceedings of the 5th DIMACS Workshop DNA Based Computers*, 1999, pp. 75–89

14. Penchovsky, R. and Ackermann, J. "DNA library approach for molecular computation", *Journal of Computational Biology*, Vol. 10, No. 2, 2003, pp. 215–229

15. Ruben, A.J. Freeland, S.J. and Landweber, L. "PUNCH: An evolutionary algorithm for optimizing bit set selection", *Proceedings of the 7th International Workshop DNA Based Computers*, 2001, pp. 260–270

16. Shin, S.Y. Kim, D.M. Lee, I.H. and Zhang, B.T. "Evolutionary sequence generation for reliable DNA computing", *Proceedings of the IEEE Congress of Evolutionary Computation*, 2002, pp. 79–84

17. Shin, S.Y. Lee, I.H. Kim, D. Zhang, B.T. "Multiobjective evolutionary optimization of DNA sequences for reliable DNA computing", *IEEE Transaction on Evolutionary Computation*, Vol. 9, No. 2, 2005, pp. 143-158

18. Sugimoto, N. Nakano, S. Yoneyama, M. and Honda, K. "Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes", *Nucleic Acid Research*, Vol.24, 1996, pp.4501-4505

19. Tanaka, F. Nakatsugawa, M. Yamamoto, M. Shiba, T. and Ohuchi, A. "Developing support system for sequence approach in DNA computing", *Proceedings of the 7th International Workshop DNA Based Computers*, 2001, pp. 340–349

20. Zhang B.T. and Shin, S.Y. "Molecular algorithms for efficient and reliable DNA computing", *Proceedings of Genetic Programming*, 1998, pp. 735–742

