A Chinese Postman Problem Based on DNA Computing

Zhixiang Yin,*,†,‡ Fengyue Zhang,† and Jin Xu†

Department of Control Science and Engineering, Hua Zhong University of Science and Technology, HuBei 430074, China and Department of Mathematics and Physics, Huainan Technology Institute, Annui 232001, China

Received May 14, 2001

DNA computing is a novel method for solving a class of intractable computational problems, in which the computing can grow exponentially with the problem size. Up to now, many accomplishments have been achieved to improve its performance and increase its reliability. A Chinese Postman Problem has been solved by means of molecular biology techniques in the paper. A small graph was encoded in molecules of DNA, and the "operations" of the computation were performed with standard protocols and enzymes. This work represents further evidence for the ability of DNA computing to solve NP-complete search problems.

1. INTRODUCTION

In 1995, Feynman gave a visionary talk describing the possibility of building computers that were submicroscopic.[1] Despite remarkable progress in computer miniaturization, this goal has yet to be achieved. Computer scientists rank computational problems in three classes: easy, hard, and uncomputable.[2] One of the major achievements of computer science in the last two decades is the understanding that many important computational search problems are NP-complete and thus are unlikely to have efficient algorithms that solve the problem exactly. Recently Adleman (1994) showed that DNA can be used to solve a computationally hard problem, the Directed Hamiltonian Path Problem, and demonstrated the potenial power of parallel, high-density computation by molecules in solution.[3] This parrallelism allows DNA computers to solve larger hard problems such as NP-complete problems in linearly increasing time, in constrast to the exponentially increasing time required by an electronic computer. After Adleman initiated the field of DNA computing in 1994, Lipton proposed DNA experiments to solve the satisfiability problem.[4] In 1997, Ouyan et al. presented a molecular biology based experimental solution to the "maximal clique" problem.[2] In 2000, Liu et al. designed a DNA model system, where a multibased encoding stategy is used in an approach to surface-based DNA computation.[5] In 2001, Wu analyzed and improved their surface-based method.[6] All their works use the tools of molecular biology, and all demonstrate the feasibility of carrying out computations at the molucular level. One of the formal frameworks for molecular computations is the Head's splicing system, which gives a theoretical foundation for computing based on DNA recombination.[7] In this paper, the Chinese Postman Problem is solved by DNA computing. It is well-known that the Chinese Postman Problem is an NP-complete problem.

* Corresponding author phone: 027-87543563, 13955406025; e-mail: zxyin66@263.net, zhixiangyin@sohu.com.
† Hua Zhong University of Science and Technology.
‡ Huainan Technology Institute.

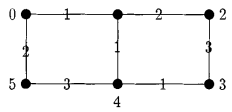

Figure 1.

2. THE CHINESE POSTMAN PROBLEM

All graphs considered in the paper are finite, undirected, and connected graphs. Let G be a connected graph and e_{ij} an edge of G. Let w_{ij} be a weight of e_{ij} (without loss of generality, we may assume w_{ij} to be a positive integer). We call a sequence constructed from a vertex with an alternating edge a closed walk, if its Start vertex is the same as the finished vertex; a call tour is a generalized Euler tour if it contains all of the edges of graph G at least once. Let G be a connected weighted graph. The Chinese Postman Problem is to find a generalized Euler tour that satisfies the sum of the weight of all edges and is minimum and the start vertex is fixed. We designed the following algorithm to solve it:

Step 1: Generate a random closed walk through the graph.

Step 2: Keep only those closed walks that begin with fixed vertices and end with fixed vertices (keep only all closed walks passing through a fixed vertex).

Step 3: Keep only those closed walks that enter all of the edges of the graph G at least once. (Namely, keep only those generalized with a Euler tour).

Step 4: Find the shortest closed walk; it will be our solution.

Step 5: Determine the postman path.

For terminologies and notations not defined in this paper, the readers are referred to ref 8.

3. ON THE CHINESE POSTMAN PROBLEM MODEL SYSTEM

The graph shown in Figure 1 with designated vertex "0" as the post office was solved with the algorithm above implemented at the molecular level. Note that the labeling of the vertices and the weight of the edges occurs in such a way that the closed walk or the generalized Euler tour passes through the edges only for convenience in this exposition

10.1021/ci010046r CCC: $22.00
Published on Web 02/02/2002

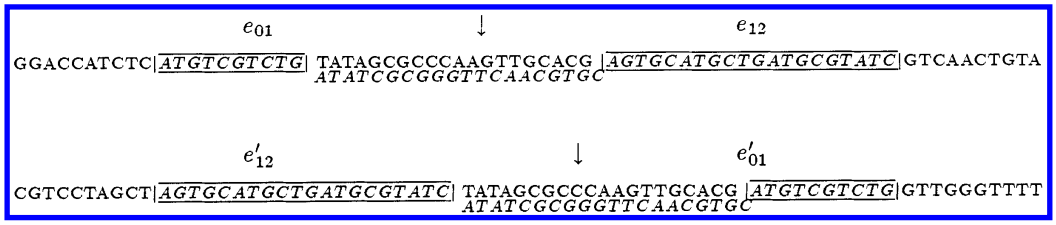

Figure 2.

and provides no advantage in the computation. The graph is small enough that the generalized Euler tour can be a visual inspection; however, it is large enough to demonstrate the feasibility of this approach. It seems clear that the methods described here could be scaled-up to accommodate much larger graphs. To implement Step 1 of the algorithm, each vertex *i* in the graph was associated with a random 20 bp (base pair) sequence of DNA denoted "i". For each edge e_{ij} in the graph, two oligoncleotides "e_{ij}" or "e_{ij}'" were created that contain three sections. For one oligoncleotide "e_{ij}", the first section is a complement strand of rear 10 bp of oligoncleotide "i"; the second section is an oligoncleotide strand, and its length is $10 \times w_{ij}$ bp; the third section is a complement strand of former 10 bp of oligoncleotide "j". With another oligonucleotide "e_{ij}'", the first section is a complement strand of former 10 bp of oligonucleotide "i"; the second section is an oligonucleotide strand, its length is $10 \times w_{ij}$ bp, and it is the same as the second section of "e_{ij}"; the third section is a complement strand of rear 10 bp of oligonucleotide "j". Notice that this construction preserves an unoriented edge. For example, oligonucleotides strand "e_{ij}" and oligonucleotides strand "e_{ij}'" preserve vertices *i* so it can arrive at *j* and *j* can also arrive at *i*.

0: CAACCCAAAACCTGGTAGAG
1: ATATCGCGGGTTCAACGTGC
2: CAGTTGACATGCAGGATCGA
3: AACCTGGTACCAAGCTTGAC
4: TGGTTTGGACTGGTCAAGTT
5: TATAGCGCATGCAGGATCGA

e_{01}: GGACCATCTC|ATGTCGTCTG|TATAGCGCCC
e'_{01}: AAGTTGCACG|ATGTCGTCTG|GTTGGGTTTT
e_{12}: AAGTTGCACG|AGTGCATGCTGATGCGTATC|GTCAACTGTA
e'_{12}: CGTCCTAGCT|AGTGCATGCTGATGCGTATC|TATAGCGCCC
e_{14}: AAGTTGCACG|CTGACGTGAC|ACCAAACCTG
e'_{14}: ACCAGTTCAA|CTGACGTGAC|TATAGCGCCC
e_{23}: CGTCCTAGCT|GTTACGTGAGTCTGACTGACGTGCATAGCT|TTGGACCATG
e'_{23}: GTTCGAACTG|GTTACGTGAGTCTGACTGACGTGCATAGCT|GTCAACTGTA
e_{34}: GTTCGAACTG|ACGATGCATG|ACCAAACCTG
e'_{34}:ACCAGTTCAA|ACGATGCATG|TTGGACCATG
e_{45}: ACCAGTTCAA|GTGCATGCGTATGAGCTGTCAGTCAGTCAC|ATATCGCGTA
e'_{45}: CGTCCAGCTT|GTGCATGCGTATGAGCTGTCAGTCAGTCAC|ACCAAACCTG
e_{50}: CGTCCTAGCT|ACTGCGTTAGCTGTAATGGT|GTTGGGTTTT
e'_{50}: GGACCATCTC|ACTGCGTTAGCTGTAATGGT|ATATCGCGTA

For each vertex *i* in the graph and for each edge e_{ij} in the graph, 100 pmol of oligonucleotides strand "i" and 100 pmol of oligonucleotides strand "e_{ij}", respectively, were mixed together in a single ligation reaction; the oligonucleotides "i" served as splints to bring oligonucleotides associated with compatible edges together for ligation (Figure 2).

Hence the ligation reaction resulted in the formation of DNA molecules encoding a random closed walk through the graph. The scale of this ligation reaction far exceeded what was necessary for the graph under consideration. For each edge in the graph, approximately 6×10^{13} copies of the associated oligonucleotide were added to the ligation reaction. Hence it is likely that many DNA molecules encoding the closed walk were created. In fact, the creation of a single such molecule would be sufficient. As a result, for this graph quantities of oligonucleotides less than an attomole would be sufficient. Alternatively, a much larger graph could have been processed with the picomole quantities used here.[3] To implement Step 2 of the algorithm, the product of Step 1 was amplified by PCR (Polymerase Chain Reaction) using rear 10 bp of oligonucleotide "0" and former 10 bp of oligonucleotide "0" as primer. Thus only those molecules encoding the generalized Euler tour that begins with vertex 0 and ends with vertex 0 were amplified. To implement Step 3 of the algorithm, the produce of Step 2 was affinity-purified with a biotin−avidin magnetic beads system. This was accomplished by first generating single-stranded DNA from the double-stranded DNA produce of Step 2 and then incubating the single-sranded DNA with the corresponding complement sequence of ATGTCGTCTG conjugated to magnetic beads. Only those single-standed DNA molecules that contained the sequence ATGTCGTCTG and hence encoded the closed walk through edge e_{01} or edge e_{01}' at least once denaturalize to the bound complement sequence of ATGTCGTCTG and were retained. This process was repeated successively with the corresponding complement sequence of AGTGCATGCTGATGCGTATC, CTGACGTGAC, GTTACGTGAGTCTGACTGACGTGCATAGCT, ACGATGCATG, GTGCATGCGTATGAGCTGTCAGTCAGTCAC, and ACTGCGTTAGCTGTAATGGT. Only those single-stranded DNA molecules contained the sequences ATGTCGTCTG, AGTGCATGCTGATGCGTATC, CTGACGTGAC, GTTACGTGAGTCTGACTGACGTGCATAGCT, ACGATGCATG, GTGCATGCGTATGAGCTGTCAGTCAGTCAC, and ACTGCGTTAGCTGTAATGGT. Hence the encoded closed walk goes through edge e_{01} or edge e_{01}', edge e_{12} or edge e_{12}', edge e_{14} or edge e_{14}', edge e_{23} or edge e_{23}', edge e_{34} or edge e_{34}', edge e_{45} or edge e_{45}', and edge e_{50} or edge e_{50}' at least once. Namely, we can find out all generalized Euler tours of *G*. To implement Step 4 of the algorithm, see the electrophoresis products of Step 3. It is for us to find the shortest DNA strand (for this graph Figure 1, the shortest DNA strand length must be 300 bp). Extract the DNA stand; afterward, this product was PCR-amplified and gel purified several times to enhance its purity. To implement Step 5 of the algorithm, we carry through sequencing to products of Step 4. Thereby, we can find out the solution of the Chinese Postman Problem. Similary with Adleman's experiment, we can accomplish the bioexperiment of Steps 1−4. Moreover, sequencing to implement Step 5, we must purify the products after each operation; it brings us convenience to sequencing by Sanger's method, and we can also use the method of sequencing by hybridization.[9] Consequently, we may read out postman paths of Chinese Postman Problem.

4. CONCLUSION

The potential of molecular computation is impressive. What is not clear is whether such massive numbers of inexpensive operation can be productively used to solve real computational problems, whereas (1) the weight of the edge may not be an integer, we only magnify w_{ij} so as to become an integer (notice each weight of the edge must be magnified the same number of times) or we directly round-off the number of these weights; (2) the weight of all edges may be too big to encode, and we only dwindle these weights into small numbers (each weight of an edge must be magnified the same number of times); Generally, we repeatedly apply to (1), (2) for conveniently encoding. Herein, the solution of the Chinese Postman Problem is not destroyed, but it is difficult for us to solve that one part of the weight of the edge which is very big and the other part of the weight of the edge which is very small. Nonetheless, for a certain intrinsically complex problem, such as the Chinese Postman Problem where existing electronic computers are very inefficient and where massively parallel searches can be organized to take advantage of the operations that molecular biology currently provides, it is conceivable that molecular computation might compete with electronic computation with our problem. We encode a graph with n vertices and only encode the n vertices, and oligoncleotides strands of all edges will be determined. Thus, the difficulty of encoding will be decreased. The approach in the paper is to solve the Chinese Postman Problem, which is very convenient. We can also solve the shortest path problem, Travelling Sale Problem (TSP Problem), with the method in this paper.

REFERENCES AND NOTES

(1) Feynmam, R. P. *In minaturization*; Gilbart, D. H., Ed.; Reinhold: New York, 1961; pp 282−296.
(2) Ouyang, Q.; Kaplan, P. D.; Liu, S.; Libchaber, A. DNA Solution of the Maximal Clique Problem. *Science* **1997**, *278*, 446−449.
(3) Adleman, L. M. Molecular Computation of Solutions to Combinatorial Problems. *Science* **1994**, *266*, 1021−1023.
(4) Lipton, R. J. DNA Solution of Hard Computational Problems. *Science* **1995**, *268*, 542−545.
(5) Liu, Q. et al. DNA Computing on Surfaces. *Nature* **2000**, *403*, 175−179.
(6) Wu, H. An improved surface-based method for DNA computation. *Biosystems* **2001**, *59*, 1−5.
(7) Head, T. Formal language theory and DNA: an analysis of the generative capacity of specific recombinant behaviors. *Bull. Math. Biol.* **1987**, *49*, 737−759.
(8) Bondy, J. A.; Murty, U. S. R. *Graph Theory with Applications*; The Macmillan Press Ltd.: 1976.
(9) Mirzabekov, A. D. DNA sequencing by hybridization a megasequencing method and adiagnostic tool. *TIBTECH* **1994**, *12*, 27−32.

CI010046R