Inter*net* Electronic Journal of
Molecular Design

February 2004, Volume 3, Number 2, Pages 102–109

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Nenad Trinajstić on the occasion of the 65th birthday
Part 8

Guest Editor: Douglas J. Klein

Chinese Postman Problem Using Molecular Programming

Zhi–xiang Yin[1,2] and Jin Xu[2]

[1] Department of Mathematics and Physics, Anhui University of Science and Technology, Anhui Huainan, 232001 China
[2] Department of Control Science and Engineering, Huazhong University of Science and Technology, Huibei Wuhan, 430074 China

Received: April 10, 2003; Revised: May 26, 2003; Accepted: June 5, 2003; Published: February 29, 2004

Citation of the article:

Z.–X. Yin and J. Xu, Chinese Postman Problem Using Molecular Programming, *Internet Electron. J. Mol. Des*. **2004**, *3*, 102–109, http://www.biochempress.com.

Inter*net* Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

Chinese Postman Problem Using Molecular Programming[#]

Zhi–xiang Yin[1,2,]* and Jin Xu[2]

[1] Department of Mathematics and Physics, Anhui University of Science and Technology, Anhui Huainan, 232001 China
[2] Department of Control Science and Engineering, Huazhong University of Science and Technology, Huibei Wuhan, 430074 China

Received: April 10, 2003; Revised: May 26, 2003; Accepted: June 5, 2003; Published: February 29, 2004

Internet Electron. J. Mol. Des. **2004**, *3* (2), 102–109

Abstract

Molecular programming (MP) has been proposed as an evolutionary computation algorithm at the molecular level. MP is different from other evolutionary algorithms in its representation of solutions using DNA molecular structures and its use of bio–lab techniques for recombination of partial solution. In this paper, the Chinese postman problem (CPP) has been solved by means of molecular programming. We make use of the encoding scheme of Shin for encoding real values. The new method is biologically plausible and has a fixed code length.

Keywords. Molecular programming; Chinese postman problem; weighted graph.

Abbreviations and notations

MP, molecular programming	SAT, the satisfiability problem
CCP, Chinese postman problem	PCR, polymerase chain reaction

1 INTRODUCTION

Adleman showed the potential of using biomolecules for solving computational problem and pioneered the field of DNA computing [1]. He solved the Hamiltonian path problem using DNA molecules, and Lipton came up with a method to solve the satisfiability (SAT) problem [2]. In 1997, Ouyan *et al*. presented a molecular biology based experimental solution to the maximal clique problem [3]. In 2000, Liu designed DNA model system; a multi–based encoding strategy is used in a one–word approach to surface–based DNA computation [4]. In 2001, Wu analyzed and improved their surface–based method [5]. In 2003, Yin *et al.* solved the 0–1 programming problem by surface–based DNA computing [6]. All their works use the tools of molecular biology and all demonstrate the feasibility of carrying out computations at the molecular level. One of the formal frameworks for molecular computations is Head's splicing system, which gives a theoretical

[#] Dedicated to Professor Nenad Trinajstić on the occasion of the 65th birthday.
* Correspondence author; phone: 86–27–87543563; fax: 86–27–87543130; E–mail: zxyin@mail.hust.edu.cn.

foundation for computing based on DNA recombination [7]. DNA computing represents the problem with DNA molecules and solves them with biological laboratory techniques. One advantage of DNA computing is massive parallelism. Another benefit is its enormous information storage capacity: 1 μ mol DNA contains approximately 10^{20} molecules, allowing a large problem space to be searched in almost constant time.

Despite significant progress, several problems remain and need to be resolved. Firstly, for a complex issue, a large amount of DNA is need in coding, which is hard to be achieved. Secondly, DNA computing is inaccurate, which can be caused by inaccurate hybridization, the effect of secondary structure of DNA molecule, the inaccuracy of experiment and large cost for biological lab experiments, all of these affecting the result of DNA computing. These problems can be alleviated by a software tool, which simulates DNA computing in advance to optimize DNA codes and experimental conditions [8]. Molecular programming (MP) has been proposed for this purpose. MP is a tool for programming DNA computers by means of artificial evolution [9]. On the other hand, MP can be regarded as a new evolutionary computation method that represents problem in DNA double strands and uses bio–lab techniques as search operators.

In 2002, Yin *et al*. solved the Chinese postman problem with DNA computing, but this method is based on the fact that all weights are integer [10]. In that paper, we used molecular programming to solve Chinese Postman Problem. Most of the problems solved by DNA computing are non–weighted graph. Such as Hamilton path problem [1] and maximal clique problem [2]. In contrast, the Chinese Postman Problem has weight (usually real valued) associated with edges. Shine proposed a method to represent weights in DNA codes to solve traveling salesman problem [8]. This method is based on the fact that hybridization between G/C pairs occurs more frequently those those between A/T pairs. It is because there are 3 hydrogen bonds between G and C, whereas 2 hydrogen bonds between A and T. Similar to the method, we can solve Chinese Postman Problem. For terminologies and notations not defied in this paper, the readers are referred to Ref. [11].

2 CHINESE POSTMAN PROBLEM

All graphs considered in this paper are finite, undirected and connected graphs. Let G be a connected graph and e_{ij} a edge of G. Let w_{ij} be a weight of e_{ij}. We call a sequence constructed from vertex and edge with alternating is closed walk, if its starting vertex is the same as finishing vertex. A 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 in which the sum of weight of all edges is minimum and the starting vertex is fixed. We designed the following algorithm to solve the Chinese Postman Problem:

Step 1. Generate random closed walks through the graph.

Step 2. Keep only those closed walks that begin from the fixed vertex and end to the fixed vertex (keep only all closed walks passing through the fixed vertex).

Step 3. Keep only those closed walks that enter all of the edge of the graph *G* at least once. (Namely, keep only those generalized Euler tours).

Step 4. Find the shortest closed walk and it will be the solution of the problem.

Step 5. Determine the postman path.

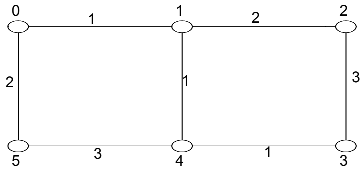

Figure 1. A graph of Chinese Postman Problem.

Figure 1 shows an instance of a graph that has 6 vertices. Let fixed vertex be 0, then the generalized Euler tour 0→1→2→3→4→1→4→5→0 and the sum of weight is 14.

3 CODING SCHEME ON CHINESE POSTMAN PROBLEM

Most existing DNA computing methods follow the Adleman's coding scheme to solve NP–complete problems. Here, the vertex codes are generated at random and then the edge codes, which link the vertices, are produced using the vertex codes. Ouyang *et al.* proposed a new coding scheme in which the edges have binary values coded in sequences. Recently, Shine proposed an improved DNA algorithm to solve traveling salesman problem [8], coding scheme of the method can overcome the difficult brought by real valued weights. According to the coding scheme, we represent edge sequences in two components: link sequences and weight sequences, represent the weight of edges by varying the amount of A/T pairs and G/C pairs in weight sequences.

Generally, DNA length and the G/C contents influence the ligation among DNA sequences. The longer the sequences the more often they get hybridized, thus leading to longer sequences of ligation. Similarly, the more G/C pairs the sequences have, the more probable they get hybridized. The reason is that the hybridizations between the G/C pairs are preferred to those between A/T pairs, since there are 2 hydrogen bonds formed between A and T and 3 hydrogen bonds between G and C. We encode the vertex with 4 components: 10–bp weight sequence ($\overline{W_{k2}}$), 10–bp position sequence (P_{i1}), 10–bp position sequence (P_{i2}), and 10–bp weight sequence ($\overline{W_{l1}}$), where the vertex position sequences P_{i1} and P_{i2} for all i are randomly generated. We also encode edge sequences with 4 components: 10–bp link sequence ($\overline{P_{i2}}$), 10–bp weight sequence (W_{l1}), 10–bp weight sequence (W_{l2}), and 10–bp link sequence ($\overline{P_{j1}}$) (see Table 1). The polarity of edge codes is the opposite (3'–5') to vertex codes. The position sequences represent a specific vertex; the weight sequence denotes a weight value in an edge. The weight sequences describe the proportion of edge weights, so the coding scheme can express the real value weights. The way describing the edge weights is explained in the fitness function [8].

Table 1. The coding scheme of fixed length

V_i : 5'–$GTAGCTATTA$	$AATAGCTACG$	$TTACATGTAA$	$TCTGCGATCC$ – 3'
$\overline{W}_{k2}$	P_{i1}	P_{i2}	$\overline{W}_{l1}$
V_j : 5'–$TTCTAGATCA$	$CAGCCTATGT$	$CCGAGTGTCA$	$AGTGGCCTTG$ – 3'
$\overline{W}_{l2}$	P_{j1}	P_{j2}	$\overline{W}_{m1}$
	Vertex sequences		
$AATGTACATT$	$AGACGCTAGG$	$AAGATCTAGT$	$GTCGGATACA$
$\overline{P_{i2}}$	W_{l1}	W_{l2}	$\overline{P_{j1}}$
	Edge sequences ($V_i \rightarrow V_j$)		

We use Figure 1 to illustrate the coding scheme. In step 1, the vertex position sequences P_{i1} and P_{i2} for all i is randomly generated. In steps 2 and 3, the whole edge sequences are designed. For the case of $V_i \rightarrow V_i$, designing link sequences $\overline{P_{i2}}$,$\overline{P_{j1}}$ and the weight sequence $W_l = (W_{l1}, W_{l2})$ and then they are combined. In Step 4, the vertex weight sequence is designed in a similar way; vertex weight sequences are generated by the edge weight sequences, and then are combined position sequences and weight sequences. In Step 5, the amount of G/C contents in edge sequences is optimized by a genetic algorithm [8]. This is done so that the edges with smaller weights have more G/C contents and thus have higher probability of being contained in the final solution. The coding scheme of Figure 1 is presented in Table 2.

The notation $0 \rightarrow 1 \rightarrow 2$ in Table 2(*a*) means that vertex 1 has in–edge from vertex 0 ($0 \rightarrow 1$) and out–edge to vertex 2 ($1 \rightarrow 2$). Note that the edge sequences with high weights have more G/C pairs than A/T pairs. Similarly, the edge sequences with low weights have more A/T pairs than G/C pairs. The link sequences contain the same number of A/T pairs and G/C. Thus, the weight sequences control the edge weight, and the link sequences have little effect on representing weight values.

4 MOLECULAR ALGORITHM ON THE CHINESE POSTMAN PROBLEM

To implement Step 1 of algorithm, for each vertex i in the graph and for each edge $i \rightarrow j$ in the graph, 50pmol of oligoncleotide strand "i" and 50 pmol of oligoncleotide strand "$i \rightarrow j$", respectively, were mixed together in a single ligation reaction, the oligoncleotides "$i \rightarrow j$" served as splints to bring oligoncleotides associated with compatible edges together for ligation (Table 3).

Table 2. Coding scheme of Figure 1

(a). Vertex sequence: weight and position sequence

1→0→1: 5'–CTAATGACTGCAACCCAAAACCTGGTAGAGATCTATGATA–3'
1→0→5: 5'–CTAATGACTGCAACCCAAAACCTGGTAGAGGCACGTGCAT–3'
5→0→1: 5'–CGACTACCGTCAACCCAAAACCTGGTAGAGATCTATGATA–3'
5→0→5: 5'–CGACTACCGTCAACCCAAAACCTGGTAGAGGCACGTGCAT–3'
0→1→2: 5'– CTAATGACTGATATCGCGGGTTCAACGTGCCACATGCGTC–3'
2→1→0: 5'–CGAGTACCGTATATCGCGGGTTCAACGTGCATCTATGATA–3'
0→1→0: 5'–CTAATGACTGATATCGCGGGTTCAACGTGCATCTATGATA–3'
2→1→2: 5'–CGAGTACCGTATATCGCGGGTTCAACGTGCCACATGCGTC–3'
0→1→4: 5'–CTAATGACTGATATCGCGGGTTCAACGTGCTATACATGAT–3'
4→1→4: 5'–CTAGAACTTGATATCGCGGGTTCAACGTGCTATACATGAT–3'
4→1→0: 5'–CTAGAACTTGATATCGCGGGTTCAACGTGCTCTATGATA–3'
4→1→2: 5'–CTAGAACTTGATATCGCGGGTTCAACGTGCCACATGCGTC–3'
2→1→4: 5'–CGAGTACCGATATCGCGGGTTCAACGTGCTATACATGAT–3'
1→2→3: 5'–CGAGTACCGTCAGTTGACATGCAGGATCGAGCCGGCCGCG–3'
1→2→1: 5'–CGAGTACCGTCAGTTGACATGCAGGATCGACACATGCGTC–3'
3→2→3: 5'–GGCCGCGCCGCAGTTGACATGCAGGATCGAGCCGGCCGCG–3'
3→2→1: 5'–GGCCGCGCCGCAGTTGACATGCAGGATCGACACATGCGTC–3'
2→3→2: 5'–GGCCGCGCCGAACCTGGTACCAAGCTTGACGCCGGCCGCG–3’
4→3→4: 5'–GTATACTACAAACCTGGTACCAAGCTTGACGAATTCATCT–3'
2→3→4: 5'–GGCCGCGCCGAACCTGGTACCAAGCTTGACGAATTCATCT–3'
4→3→2: 5'–GTATACTACAAACCTGGTACCAAGCTTGACGCCGGCCGCG–3'
3→4→5: 5'–GTATACTACATGGTTTGGACTGGTCAAGTTGGCCCGGCGC–3'
5→4→3: 5'–GGCCCGGCCGTGGTTTGGACTGGTCAAGTTGAATTCATCT–3'
3→4→3: 5'–GTATACTACATGGTTTGGACTGGTCAAGTTGAATTCATCT–3'
5→4→5: 5'–GGCCCGGCCGTGGTTTGGACTGGTCAAGTTGGCCCGGCGC–3'
1→4→1: 5'–CTAGAACTTGTGGTTTGGACTGGTCAAGTTTATACATGAT–3'
3→4→1: 5'–GTATACTACATGGTTTGGACTGGTCAAGTTTATACATGAT–3'
1→4→3: 5'–CTAGAACTTGTGGTTTGGACTGGTCAAGTTGAATTCATCT–3'
1→4→5: 5'–CTAGAACTTGTGGTTTGGACTGGTCAAGTTGGCCCGGCGC–3'
5→4→1: 5'–GGCCCGGCCGTGGTTTGGACTGGTCAAGTTTATACATGAT–3'
0→5→0: 5'–CGTAGCTCGATATAGCGCATGCAGGATCGAGCACGTGCAT–3'
4→5→4: 5'–GGCCCGGCCGTATAGCGCATGCAGGATCGAGGCCCGGCGC–3'
4→5→0: 5'–GGCCCGGCCGTATAGCGCATGCAGGATCGAGCACGTGCAT–3'
0→5→4: 5'–CGTAGCTCGATATAGCGCATGCAGGATCGAGGCCCGGCGC–3'

(b). Edge sequence: Link and weight sequence

0→1: 3'–GGACCATCTCTAGATACTATGATTACTGACTATAGCGCCC–5'
1→0: 3'–AAGTTGCACGTAGATACTATGATTACTGACGTTGGGTTTT–5'
1→2: 3'–AAGTTGCACCGTGTACGCAGGCTCATGGCAGTCAACTGTA–5'
2→1: 3'–CGTCCTAGCTGTGTACGCAGGCTCATGGCATATAGCGCCC–5'
2→3: 3'–CGTCCTAGCTCGGCCGGCGCCCGGCGCGGCTTGGACCATC–5'
3→2: 3'–GTTCGAACTGCGGCCGGCGCCCGGCGCGGCGTCAACTGTA–5'
3→4: 3'–GTTCGAACTGCTTAAGTAGACATATGATGTACCAAACCTG–5'
4→3: 3'–ACCAGTTCAACTTAAGTAGACATATGATGTTTGGACCATG–5'
4→5: 3'–ACCAGTTCAACCGGGCCGCGCCGGGCCGGCATATCGCGTA–5'
5→4: 3'–CGTCCTAGCTCCGGGCCGCGCCGGGCCGGCACCAAACCTG–5'
5→0: 3'–CGTCCTAGCTCGTGCACGTAGCATCGAGCTGTTGGGTTTT–5'
0→5: 3'–GGACCATCTCCGTGCACGTAGCATCGAGCTATATCGCGTA–5'
1→4: 3'–AAGTTGCACCATATGTACTAGATCTTGAACACCAAACCTG–5'
4→1: 3'–ACCAGTTCAAATATGTACTAGATCTTGAACTATAGCGCCC–5'

Hence the ligation reaction resulted in the formation of DNA molecules encoding random closed walks 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 attomle would be sufficient. Alternatively, a much larger graph could have been processed with the picomole quantities used here.

Table 3. Ligation

Vertex i ↓ Vertex j
GTAGCTATTAAATAGCTACGTTACATGTAATCTGCGATCCAATGTACATTAGACGCTAGGAAGATCTAGTGTCGGATACA
AATGTACATTAGACGCTAGGAAGATCTAGTGTCGGATACA
Edge $V_i \rightarrow V_j$

To implement the Step 2 of the algorithm, the product of Step 1 was amplified by polymerase chain reaction (PCR) using rear 10bp of position sequence oligonucleotide “0” and former 10bp of position sequence oligonucleotide “0” as primer. Thus only those molecules encoding generalized Euler tour that begin with vertex 0 and end 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.

First generating single–stranded DNA from the double–stranded DNA produce of Step 2 and then incubating the single–stranded DNA with corresponding complement sequence of TAGATACTATGATTACTGAC conjugated to magnetic beads accomplished this. Only those single–stranded DNA molecules that contained the sequence TAGATACTATGATTACTGAC and hence encoded closed walk that through edge 0→1 or edge 1→0 at least once, denaturalize to the bound complement sequence of ATGTCGTCTG and were retained. This process was repeated successively with corresponding complement sequence of GTGTACGCAGGCTCATGGCA, CGGCCGGCGCCCGGCGCGGC, CTTAAGTAGACATATGATGT, CCGGGCCGCGCCGGGC CGGC, CGTGCACGTAGCATCGAGCT and ATATGTACTAGATCTTGAAC. Only those single–stranded DNA molecules that contained the sequences GTGTACGCAGGCTCATGGCA, CGGCCGGCGCCCGGCGCGGC, CTTAAGTAGACATATGATGT, CCGGGCCGCGCCGGGC CGGC, CGTGCACGTAGCATCGAGCT and ATATGTACTAGATCTTGAAC.

Hence encoded closed walk that through edge $0 \rightarrow 1$ or edge$1 \rightarrow 0$, edge $1 \rightarrow 2$ or edge$2 \rightarrow 1$, edge $2 \rightarrow 3$ or edge$3 \rightarrow 2$, edge $3 \rightarrow 4$or edge$4 \rightarrow 3$, edge $4 \rightarrow 5$ or edge $5 \rightarrow 4$, edge$5 \rightarrow 0$ or edge$0 \rightarrow 5$, edge $1 \rightarrow 4$ or edge $4 \rightarrow 1$ at least once. Namely, we can find out all generalized Euler tour of Fig 2. To implement Step 4 of algorithm, we observed the number of hydrogen bonds of products of Step 3.

In this way we find the shortest DNA strand (for the graph Figure 2, the number of hydrogen bonds of the shortest DNA sequences must be Minimum). Extract the DNA stand, afterwards, this product was PCR–amplified and gel purified several times to enhance its purity. To implement Step 5 of algorithm, we carry through sequencing to products of Step 4. Thereby, we can find out the solution of Chinese Postman Problem. Similarly to Shine's experimentation, we can accomplish bio–experiment Step 1, Step 2, Step 3 and Step 4. Moreover, sequencing to implement Step 5, we purified the products after each operation, which brings us convenience in sequencing by Sanger's method. For this task we can also use the method of SBH (sequencing by hybridization) [12]. Consequently, we may read out postman paths of the Chinese Postman Problem.

5 CONCLUSIONS

The potential of molecular computation is impressive, but what is not clear is whether such massive numbers of inexpensive operation can be productively used to solve real computational problems. We presented a weight encoding scheme for molecular programming and demonstrated its performance on the Chinese postman problem. In this encoding, the relative values of G/C contents against A/T contents are taken into account to represent real–valued weight of the edge in the graph. Since G/C pairs have 3 hydrogen bonds and A/T pairs 2, we can represent real–valued X according to the G/C contents.

We have shown that the method is effective for reliable DNA computing applied to Chinese postman problem. The method can easily be modified to be used in other graph problem in which edges are associated with real–valued costs. Nonetheless, for certain intrinsically complex problem, such as 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 to our problem.

For a graph with *N* vertices, in the course of encoding, oligonucleotide sequences of all vertices and edge will be determined after encoding position sequences of the *N* vertices and weight sequences of all edges. The approach in the paper may be used to solve the Chinese Postman problem. Using the method proposed in the paper we can also solve the shortest path problem, TSP.

Acknowledgment

This project is supported by the National Natural Science Foundation of China (Grant No. 60274026) and Anhui province educatuional committee Foundation (Grant No. 2003Kj098).

5 REFERENCES

[1] L. M Adleman, Molecular Computation of Solutions to Combinatorial Problems, *Science* **1994**, *266*, 1021–1023.
[2] R. J Lipton,.DNA Solution of Hard Computational Problems, *Science* **1995**, *268*, 542–545.
[3] Q. Ouyang, P. D. Kaplan, S. M. Liu, DNA Solution of the Maximal Clique Problem, *Science* **1997**, *278*, 446–449.
[4] Q. H. Liu, DNA Computing on Surfaces, *Nature* **2000**, *403*, 175–179.
[5] H. Y. Wu, An improved surface–based method for DNA computation, *Biosystems* **2001**, *59*, 1–5.
[6] Z. X. Yin, F. Y. Zhang, and J. Xu, A general 0–1 programming problem Based on DNA Computing, *Biosystems* **2003**, *70*, 73–79.
[7] T. Head, Formal language theory and DNA: an analysis of the generative capacity of specific recombinant behaviors, *Bull. Math. Biol* **1987**, *49*, 737–759.
[8] S. Y. Shine, Solving Traveling Problrms using Molecular Programming, *IEEE* **1999**, *35*, 994–1000.
[9] B. T. Zhang, and S. Y. Shin, Code Optimization for DNA Computing of Maximal Cliques, Advances in Soft Computing–Engineering Design and Manufacturing, Springer–Verlag, 1998.
[10] Z. X. Yin, F. Y. Zhang, and J. Xu, A Chinese Postman Problem Based on DNA Computing, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 222–224.
[11] J. A. Bondy, and U. S. R. Murty, Graph Theory with Applications, The Macmillan Press LTD 1976
[12] A. D. Mirzabekov, DNA sequencing by hybridization a megasequencing method and adiagnostic tool, *TIBTECH* **1994**, *12*, 27–32.

Biographies

Zhi–xiang Yin is professor of mathematics at the Anhui University of Science and Technology. After obtaining a Ph.D. degree in mathematical modeling of DNA computing from the Huazhong University of Science and Technology, Dr. Zhi–xiang Yin undertook DNA computing research at the Anhui University of Science and Technology.