Available online at www.sciencedirect.com
ScienceDirect
Applied Mathematics and Computation 183 (2006) 79–84
APPLIED MATHEMATICS AND COMPUTATION
www.elsevier.com/locate/amc

A DNA procedure for solving the shortest path problem ☆

Zhaocai Wang [a,c], Dongmei Xiao [a,c], Wenxia Li [b,c,*,1], Lin He [c]

[a] Department of Mathematics, Shanghai Jiao Tong University, Shanghai 200030, PR China
[b] Department of Mathematics, East China Normal University, Shanghai 200062, PR China
[c] Bio-X DNA Computer Consortium, Shanghai Jiao Tong University, Shanghai 200030, PR China

Abstract

In this paper, we consider a procedure for solving the shortest path problem in the Adleman–Lipton model. The procedure works in O(n) steps for the shortest path problem of an edge-weighted graph with n vertices.

Keywords: The shortest path problem; NP-complete problem; Adleman–Lipton model; DNA computing

1. Introduction

In recent works for high performance computing, computation with DNA molecules, i.e., DNA computing, has considerable attention as one of non-silicon based computing. Watson-Crick complementarity and massive parallelism are two important features of DNA. Using the features, one can solve an NP-complete problem, which usually needs exponential time on a silicon based computer, in a polynomial number of steps with DNA molecules, e.g., Adleman [1] for Hamiltonian path problem – the first work for DNA computing, Lipton [11] for satisfiability (SAT) problem (the first NP-complete problem), Ouyang et al. [13] for the maximal clique problem, etc. Meanwhile, procedures for primitive operations, such as logic or arithmetic operations, have been also proposed so as to apply DNA computing on a wide range of problems [2–4,6–8,16,17]. However, most of the previous works in DNA computing do not require the consideration of the representation of numerical data in DNA strands. In fact, many practical applications in the real world involve edge-weighted graph problems such as shortest path problem, the travelling-salesman problem, etc. 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 have been some previous works to represent the numerical data with DNA. Narayanan et al. [12] presented a conceptual encoding method that represents costs with the lengths of DNA strands. Shin et al. [15] proposed a method for representing the real

☆ Supported by Bio-X DNA Computer Consortium No. 03DZ14025.
* Corresponding author. Address: Department of Mathematics, East China Normal University, Shanghai 200062, PR China.
E-mail address: wxli@math.ecnu.edu.cn (W. Li).
1 Supported by National Science Foundation of China #10371043 and Shanghai Priority Academic Discipline.

0096-3003/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.amc.2006.05.061

numbers in fixed-length DNA strands by varying the number of hydrogen bonds. Yamamura et al. [18] proposed a concentration control method which encoded the numerical data by means of the concentrations of DNA strands. Lee et al. [9] introduced a novel encoding method that utilizes a temperature gradient to design the sequences so that the DNA strands for higher-cost values have higher melting temperatures than those for lower-cost values.

In this paper, a DNA procedure is presented for figuring out solutions of the shortest path problem: for an edge-weighted graph $G = (V, E)$ find a path starting and ending at the specified vertices such that the total weights on the path is smallest. For instance, the edge-weighted graph G in Fig. 1 defines such a problem. We assume that the starting and ending vertices are 1 and 7, respectively. It is easy to see that the path $1 \to 2 \to 3 \to 4 \to 6 \to 7$ with total weights 8 is a solution to the shortest path problem for graph G in Fig. 1. We encode the numerical data by means of the lengths of DNA strands, the same way as that in [12]. A DNA procedure is formally presented by means of the DNA operations proposed by Adleman [1] and Lipton [11]. Since the shortest path discussed in the present paper is not required to go through each vertex, e.g., the solution for the graph G in Fig. 1 does not go through the vertex 5, we use the *append* operation in the design of data pool so that the shortest path can be found out by comparing the lengths of the DNA strands.

The rest of this paper is organized as follows. In Section 2, the Adleman–Lipton model is introduced in detail. Section 3 introduces a DNA algorithm for solving the shortest path problem and the complexity of the proposed algorithm is described. We give conclusions in Section 4.

2. The Adleman–Lipton model

Bio-molecular computers work at the molecular level. Because biological and mathematical operations have some similarities, DNA, the genetic material that encodes for living organisms, is stable and predictable in its reactions and can be used to encode information for mathematical systems.

A DNA (deoxyribonucleic acid) is a polymer, which is strung together from monomers called deoxyribonucleotides [14]. Distinct nucleotides are detected only with their bases. Those bases are, respectively, abbreviated as A (adenine), G (guanine), C (cytosine) and T (thymine). Two strands of DNA can form (under appropriate conditions) a double strand, if the respective bases are the Watson-Crick complements of each other – A matches T and C matches G; also $3'$ end matches $5'$ end, e.g., the singled strands $5'$ACCGGATGTCA$3'$ and $3'$TGGCCTACAGT$5'$ can form a double strand. We also call the strand $3'$TGGCCTACAGT$5'$ as the complementary strand of $5'$ACCGGATGTCA$3'$ and simply denote $3'$TGGCCTACAGT$5'$ by $\overline{\text{ACCGGATGTCA}}$. The length of a single stranded DNA is the number of nucleotides comprising the single strand. Thus, if a single stranded DNA includes 20 nucleotides, it is called a 20 mer. The length of a double stranded DNA (where each nucleotide is base paired) is counted in the number of base pairs. Thus, if we make a double stranded DNA from a single stranded 20 mer, then the length of the double stranded DNA is 20 base pairs, also written as 20 bp.

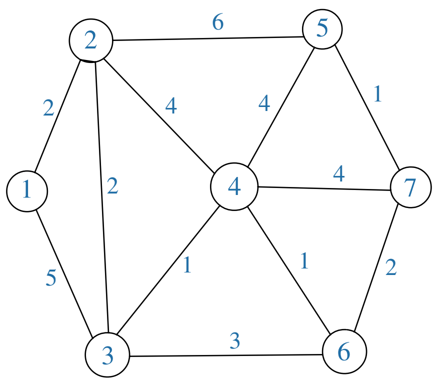

Fig. 1. An edge-weighted graph G with 7 vertices.