A novel DNA computing model based on RecA-mediated triple-stranded DNA structure

Fang Gang, Zhang Shemin, Dong Yafei, and Xu Jin

Abstract The field of DNA computing emerged in 1994 after Adleman's paper was published. Henceforth, a few scholars solved some noted NP-complete problems in this way. And all these methods of DNA computing are based on conventional Watson-Crick hydrogen bond of double-helical DNA molecule. In this paper, we show that the triple-stranded DNA structure mediated by RecA protein can be used for solving computational problems. Sequence specific recognition of double-stranded DNA by oligonucleotide-directed triple helix (triplex) formation is used to carry out the algorithm. We present procedure for the 3-vertex colorability problem. In our proposed procedure it is suggested that it is possible to solve more complicated problems with more variables by this model.

Keywords DNA computing, 3-vertex colorability problem.

It was well known that DNA computing was firstly presented in Adleman's paper[1], after which the field of DNA computing emerged. Henceforth, a few scholars solved some noted NP-complete problems in this way, such as Lipton[2]. In his paper, he showed how a large class of NP-complete problems could be solved by encoding the problem in DNA molecules. After this, Ouyang et al.[3] presented a molecular biology-based experimental solution to the "maximal clique" problem. Before long, the notion of computing by 3-Dimensional DNA structure was firstly proposed by Jonoska et al.[4]. Liu et al.[5] designed a DNA model system where multi-based encoding strategy was used in a one-word approach to surface-based DNA computation. Not long ago, Braich et al.[6] presented the solution of a 20-variable 3-SAT problem based on Sticker DNA computing model. And this computational problem may be the largest yet solved by nonelectronic means, though all these methods of DNA computing are based on conventional Watson-Crick hydrogen bond of double-helical DNA molecule. Naturally, sequence-specific recognition of double-stranded DNA by oligonucleotide-directed triple helix (triplex) formation always produces triple-stranded DNA. For homopurine/homopyrimidine triplex structures recognition of the third strand occurs through the major groove by formation of Hoogsteen-type or reversed Hoogsteen-type hydrogen bonds between the bases of the oligopyrimidine or oligopurine third strand and the purine-rich strand of the duplex target[7]. During homologous and general recombination in Escherichia coli, this unique formation is always mediated by RecA-protein. Since triple-stranded DNA can be formed readily at any location in double-stranded DNA with a homologous deoxyoligonucleotide in the presence of RecA-protein[8-10], we will explore the possibility and feasibility to compute by this unique structure.

In this paper we propose solutions to a famous NP-complete problem, the 3-vertex-colorability problem. Firstly, we will describe the formation of RecA-mediated triple-stranded DNA structure. Secondly, we describe our approach for solving the 3-vertex-colorability problem and present the detailed biological procedure. In the end we discuss the potential merits and drawbacks of computing by RecA-mediated triple-stranded DNA structure.
1 Formation of RecA-mediated triple-stranded DNA

Triple-stranded DNA structures have been proposed as intermediates in the strand exchange reaction during homologous recombination \(^{[9,11]}\). And the formation of this unique structure is always mediated by RecA-protein. It plays a central role in bacterial homologous recombination. Previous observations have shown that RecA-protein will form stable complexes with oligodeoxyribonucleotides (ODN) in the presence of ATP/γS, and that such a stable nucleoprotein filament is active in homologous pairing and promotes the formation of stable triple-stranded DNA structure \(^{[11-12]}\). In the first step of formation, RecA-protein binds ODN or ssDNA in the presence of ATP/γS. The resulting protein-DNA complex, called nucleoprotein filament, searches for homology on a dsDNA target. This search is fast and does not appear to involve base pair opening \(^{[13]}\). When homology is found, RecA aligns ssDNA with a duplex target and forms a synaptic complex, composed of RecA and the three DNA strands (Fig. 1). A molecular model of such a DNA structure has been calculated and denoted the parallel triplex or recombinant, R-form DNA \(^{[14]}\). It accommodates two identical DNA sequences in the third strand and the duplex target, designated by Zhurkin R (recombinant) and W (Watson) strands, respectively, in a parallel orientation. The complementary C (Crick) strand forms conventional Watson-Crick hydrogen bonds with the W strand. In the proposed model the R strand lies in the major groove of Watson-Crick duplex.

![fig1](https://example.com/fig1.png)

Fig. 1. Formation of RecA-mediated triple-stranded DNA.

The current interest in triple helical DNA arises from its potential in diagnostic and therapeutic clinical applications, such as sequence-specific drug targeting, DNA cleavage, mutagenesis and artificial gene regulation \(^{[12]}\). Recently, this unique DNA structure was used to isolate some DNA variants and detect SNPs \(^{[15,16]}\), and in this way the sequence-specificity of RecA-mediated triple-stranded DNA forming was demonstrated. Now, some properties of this unique structure inspire us to explore the possibility and feasibility of computing by it.

2 Solution of a 3-vertex-colorability problem

Let \( G \) be a graph with vertices \( V \) and edges \( E \). The graph \( G \) is 3-vertex-colorable if there is a surjective (onto) function \( f: V \rightarrow \{a, b, c\} \) such that if two vertices \( v, w \in V \) are adjacent (connected by an edge) then \( f(v) \neq f(w) \). The 3-colorability is defined similarly. This is an NP-complete problem which has been addressed before \(^{[17]}\). Here we present a solution to a simple example of 3-vertex-colorability problem. In Fig. 2, the graph is 3-vertex-colorable. For example, in Fig. 2, the 3 color group was \{1\}, \{2,4\}, \{3,5\}. If we presume that the vertex 1 is red, we will get two right solutions. To represent all possible color assignments to the vertices, the 3-color assignments to each of the vertices 1, 2, 3, 4, 5 are represented by three distinct 30-mer ODN sequences—one for a vertex colored Blue (\( B_i \)), \( B_i (i = 1, 2, 3, 4, 5) \), one for a vertex colored Yellow (\( Y_i \)), and one for a vertex colored Red (\( R_i \)). Each of the \( 3^5 \) possible solutions is represented by a 150 bp dsDNA library sequence whose template consists of the ordinal concatenation of one sequence representing one vertex-color (\( B_i \), \( Y_i \), or \( R_i \)). And these sequences (\( B_i \), \( Y_i \), or \( R_i \)) are also used as probes to separate the library sequences.

![fig2](https://example.com/fig2.png)

Fig. 2. A 3-vertex-colorable graph.

The corresponding sequences are generated to represent \( B_i \), \( Y_i \), and \( R_i (i = 1, 2, 3, 4, 5) \). Synthesizing the probe strands and full library strands on automated DNA synthesizer in the way that Braich et al. described \(^{[8]}\), then the procedure of the algorithm
starts:

Step 1: Mix and incubate the probe strands and RecA protein in a reaction mixture which contain ATPγS. Thus, the nucleoprotein filament will come into being.

Step 2: Mix the first couple of probes \(\{B_1, B_2\}\)—RecA protein complex, which represents that both of the two adjacent vertices—1 and 2 are colored Blues with the full library strands. Then some triple-stranded DNA structures will form. The probes were selected according to the adjacency matrix of the graph in Fig. 2, and they contained all the information of the graph.

Step 3: Separate the products from Step 2 by gel electrophoresis. Extract the strands that contain only one probe and do not contain triple-stranded DNA structure. Then deproteinizing these strands, the desired library strands are acquired.

Step 4: Repeat Step 2 and Step 3 with each of the remaining 20 couples of probes, which represent that two adjacent vertices are same colored. There are 7 edges in this graph, so there will be totally \(3 \times 7 = 21\) couples of probes. At the end of Step 4, all 21 couples of probes have been applied to separate the library strands. So the finally acquired library strands are the solution strands.

Step 5: Extract the final strands, PCR-amplify, and “read” the answer. The solution strands are

\[
B_1 Y_2 R_3 Y_4 R_5, \quad B_1 R_2 Y_3 R_4 Y_5, \quad Y_1 B_2 R_3 B_4 R_5, \quad Y_1 R_2 B_3 R_4 B_5, \quad R_1 B_2 Y_3 B_4 Y_5, \quad \text{and} \quad R_1 Y_2 B_3 Y_4 B_5.
\]

The algorithm is based on the fact that the library strands representing two adjacent vertices in the same colored include the strands representing all the unfeasible solutions. And the procedure of this algorithm is illustrated in Fig. 3.

## 3 Discussions

In the above proposition, owing to the stability of nucleoprotein filament and double-stranded DNA, secondary structure forming of probes and library strands will be prevented. Compared to other DNA computing methods, interprobe and interlibrary strand hybridization cannot happen. Thus the errors in computation will decrease. Because of this, the coding complexity will decrease too.

Although we believe that the biological procedure described is theoretically feasible, little specific experimentation is needed to confirm this proposition. In the procedure of solution to the problem, the techniques are available in common molecular biological laboratories. And we think it is considerably feasible. However, there are some problems, for example, in the biological procedure of solution to 3-vertex-colorability problem we do not clearly understand how the RecA protein and triple-stranded DNA complexes move in gel electrophoresis when the strands are separated. Theoretically, by using capillary electrophoresis technique desired sequences can be extracted. And specific experimentation should be carried out to clarify it. Although the coding complexity is expected to decreases the degree of sequence-specific recognition between nucleoprotein filament and doublestranded DNA is not very clear. If it is clarified some new coding strategies should be adopted. Furthermore, the efficiency of RecA-mediated triple-stranded DNA structure forming should be well characterized and understood.

Both the potential merits and drawbacks of computing by RecA-mediated triple-stranded DNA structure are obvious. And the potential merits of comput-
ing by triple-stranded DNA structure are attractive. Specific biological experimentation is needed to validate this new DNA computing model. Maybe unexpected questions will emerge when relevant experimentation is being carried out, and these questions will be answered with the development of laboratory techniques. In this way computing by triple-stranded DNA structure will be implemented and improved.

References

1 Adleman L. Molecular computation of solutions of combinatorial problems. Science 1994, 266; 1021—1024
2 Lipton RJ. DNA solution of hard computational problems. Science 1995, 268; 542—545
7 Frank Kamenetskii MD and Mirkin SD. Triplex DNA structures. Annu Rev Biochem 1995, 64; 65—95
8 Umlauf SW, Cox MM and Inman RB. Triple-helical DNA pairing intermediates formed by RecA protein. J Biol Chem 1990, 265; 16898—16912
10 Muller B, Bardett I and West SC. Unusual stability of recombination intermediates made by Escherichia coli RecA protein. EMBO J 1992, 7; 2685—2693
12 Shigemori Y and Oishi M. Specific cleavage of DNA molecules at RecA-mediated triple-stranded structure. Nucleic Acids Research 2004, 32(1); e4. doi: 10.1093/nar/gnh004
13 Yancey-Wrona JE and Camerini-Otero RD. The search for DNA homology does not limit stable homologous pairing promoted by RecA protein. Curr Biol 1995, 5; 1149—1158