DNA Strand Displacement Computing Model for the SAT Problem

Jing Yang, Xue Pang, Zhen Tang, Xinmu Yang and Congcong Liu

School of Mathematics and Big Data, Anhui University of Science and Technology, Huainan, Anhui, China
Email: jyangh82@163.com

Abstract. DNA origami is widely used in DNA computing by its programmability and nano-addressability. This study presents the solution for the satisfiability (SAT) problem utilizing DNA strand displacement and origami DNA: firstly, map all the solutions of the SAT problem to the origami base; secondly, add the initiator chain to make it fully react; finally, determine the feasible solution by whether there is fluorescence on the DNA origami base, so as to search for the final solution. And through a specific case, we verify the feasibility of the model.

Keywords. DNA origami; satisfiability problem; DNA strand displacement; DNA computing.

1. Introduction
In 1994, Professor Adleman used DNA molecule to solve the NPC problem --the Hamilton path problem [1]. Since then, computing model using DNA molecule as a computational medium has been developed. In 2006, Rothemund [2] proposed for the first time a new DNA self-assembly method -- DNA origami. By base-pairing more than 200 specially designed staple strands (each about 32bp) with DNA long strands (M13MP18), he successfully folded rectangles, triangles, five-pointed stars and smiling faces in a cover paper published in the journal of Nature. As the continuous development of molecular biology technology, DNA origami has changed from one-dimensional structure to two-dimensional structure and then to three-dimensional structure [3-6], playing an increasingly important role in NP problems. DNA Origami is a kind of assemble stable and highly complex nanostructures than traditional DNA self-assembly. Therefore, it has the advantages of simple design, high assembly efficiency and nanometer addressable. In 2011, Han D et al. designed and constructed self-assembled nanostructures, using DNA Origami to assemble a series of DNA nanostructures with high curvature [7]. In 2015, Zhang et al. used DNA Origami to construct a finite size wire-frame DNA nanostructure with programmability and high complexity [8]. In 2016, Chao encoded the vertices of undirected graphs with DNA nano-origami structures and self-assembled the sticky ends between nanostructures to solve the graph coloring problem [9]. In 2017, Tikhomirov published an article about micron-scale DNA origami arrays in Nature [10]. They successfully used DNA origami to replicate images of the Mona Lisa and a rooster. And it was found that self-assembly into an array was unaffected by changes in tile surface patterns. In 2018, Chao built a DNA Navigator system. The system can perform single-molecule parallel depth-first searches in the root tree of ten vertices defined on the DNA Origami basis. They used it to solve the maze problem [11]. In 2019, Tang made DNA origami to establish a dynamic NAND gate calculation model [12]. In 2020, He used DNA polymerase to design DNA single-stranded scaffold into geometric structure [13]. It provides cost-effective method for small-scale DNA origami prototype production. In the same year, Khosravi designed DNA origami nanopore [14]. This model has good stability.
DNA strand displacement (DSD) is a process of hybridization between single strand of DNA and a partially complementary double strand of DNA, in which the single strand of the original structure is replaced and released, and a new double strand structure is finally produced, the longer strand replaces the shorter strand, which is used as the output strand. The basic process of DSD is shown in Figure 1. DSD is widely used in science and technology, nanotechnology, cryptography, biomedicine and other fields. The DNA strand replacement design has a simple structure, easy reaction operation, high efficiency, and high yield. In 2015, Li designed and synthesized a new type of photoelectrochemically active material for DNA detection. He created a photoelectrochemical sensor for DNA detection based on hybrid chain reaction signal amplification for the first time, which is expected to further improve the material for bioassay and clinical diagnosis [15]. In 2017, George used DNA strand replacement to design a logic inverter gate [16]. The logic inverter gate has modular characteristics and broad application prospects. It uses visual DSD software to simulate DNA circuits. In 2020, Li proposed a new idea of odd judgment logic circuit based on DNA strand replacement reaction technology. It is a method for solving practical mathematical problems and is widely used in a variety of logic circuits and computing systems. Therefore, it plays an important role in biological computer [17].

2. The Satisfiability Problem

2.1. The Satisfiability Problem

SAT problem is the first known NPC problem, which is a fundamental problem of computer science and artificial intelligence. It has a wide range of applications in computer vision hardware testing cryptography and artificial intelligence.

Definition: Given \( n \) boolean variables \( x_1, x_2, \ldots, x_n \) and \( m \) clause \( C_1, C_2, \ldots, C_m \). A clause is defined as follows:

\[
C_j = x_{i1} \lor x_{i2} \lor \cdots \lor x_{ij}, \quad j = 1, 2, \ldots, m
\]  

(1)

where ‘\( \lor \)’ is the logical ‘or’, each \( x_{ij} \) is either a variable or the negation of a variable in \( C_j \).

A formula \( \alpha \) in conjunctive normal form is defined as follows:

\[
\alpha = C_1 \land C_2 \land \cdots \land C_m
\]  

(2)

where ‘\( \land \)’ is the logical ‘and’, and each \( C_i \) is a clause.

An assignment is defined as a mapping from Boolean variables set to \{False, True\}. For a given conjunctive normal form with \( n \) variables, if there are one or more sets of variables that make the conjunctive normal form true, this problem is called satisfiability problem (abbreviated as SAT problem). Since each variable has 2 values, there are \( 2^n \) possible solutions to an SAT problem with \( n \) variables. Many scholars are devoted to the study of satisfying algorithms. Since DNA computation can be used in combinatorial optimization, many scholars try to use DNA molecules to solve the satisfiability problem. In 1995, Lipton mapped the solution space of the satisfiability problem into a contact network graph, and used a series of operations on the encoded oligonucleotide fragments to eliminate the solutions that did not satisfy the STA problem, then to solve the satisfiability of the SAT problem [18]. In 2000, Sakamoto encoded the logical operational constraints of DNA molecules by using the hairpin structure of the single-stranded molecule. Then the 3-SAT problem was solved by DNA self-assembly [19]. In 2002, Braich used the sticker model to design a semi-automated device to solve the DNA
solution of the satisfiability problem of 20 variables [20]. In 2008, Wang proposed a biological algorithm based on DNA LCR (Ligase Chain Reaction). And this algorithm is used to solve the SAT problem [21]. This method has strong search ability, high efficiency, and low error rate. In 2015, Yang Jing et al. constructed the molecular beacon tile by using molecular beacon, and solved the 3SAT problem by using DNA self-assembly model [22]. In 2016, Qian proposed a DNA computing model. And they used this DNA chip to solve the SAT problem [23]. In 2020, Cui presented the SAT problem model based on probe machine [24].

2.2. Biological Steps
Step 1: There are \(2^n\) possible solutions to the SAT problem with \(n\) variable. Add different origami bases and the same auxiliary chain in \(2^n\) test tubes;

Step 2: Construct all possible solutions of the first clause, add the initiation strands of the search solution to \(2^n\) test tubes, after a period of DNA strand displacement reaction, observation whether there is fluorescence on the origami substrate in the test tube. If there is fluorescence, it is the solution of this clause, otherwise it is not the solution of the clause;

Step 3: Repeat step 2 to find the solutions of all clauses. The common solution of each clause is the final solution of SAT.

3. Case Analysis

3.1. Model Composition
Without loss of generality, take the 3-SAT problem as an example, \(F = (x_1 \lor \bar{x}_2 \lor x_3) \land (x_1 \lor x_2) \land (\bar{x}_1 \lor x_3)\). The model consist of DNA origami bases, hairpin structures, hairpin structures labeled with quenching group and fluorescent group, initiation chain, and auxiliary chains. The initiation chain needs to be artificially added to the test tube, and the auxiliary chain is stable in a large amount in the test tube. For variable \(x_i\) to be 0, the hairpin structure is not labeled with quenching groups and fluorescent groups, while for variable \(x_i\) to be 1, the hairpin structure is labeled with quenching groups and fluorescent groups (figure 2).

![Model composition structure](image)

**Figure 2.** Model composition structure.

3.2. Model Solving
For the first clause of the satisfiability problem, there are 3 variables \(x_1, x_2,\) and \(x_3\), and there are \(2^3\) possible solutions, each of which is \(1(0, 0, 0), 2(0, 0, 1), 3(0, 1, 0), 4(0, 1, 1), 5(1, 0, 0), 6(1, 0, 1), 7(1, 1, 0), 8(1, 1, 1)\), recorded as the No1 ~ 8 solution, so 8 kinds of origami bases are required, as shown in figure 3. In all test tubes, you only need to add an initiation chain, the initiation chain will react with the stem of the hairpin structure, and the hairpin structure will be opened. After the hairpin structure is opened, it will react with the auxiliary chain in the solution, and the auxiliary chain will displace the initiation chain. At this time, the distance between the quenching group and the fluorescent group is extended. The quenching group cannot absorb the fluorescence emitted by the fluorophore, the fluorescent signal can be detected, as shown in figure 4. The displaced initiation chain will react with the rest of the variables, and finally will react completely.
The first clause $x_1 \lor \bar{x}_2 \lor x_3$, $\bar{x}_4$ means when $x_1=0$, $\bar{x}_2=1$; when $x_1=1$, $\bar{x}_2=0$. Take solution No. 3 (0, 1, 0) and solution No. 5 (1, 0, 0) as examples. Solution No. 3 means $x_1 = 0$, $x_2 = 1$, $x_3 = 0$. Adding the initiation chain to the test tube. The initiation chain and auxiliary chain will open all the hairpin structures. When $x_2=1$, $\bar{x}_2=0$, $x_1$ and $x_3$ are all 0, so no fluorescence is detected, that is, solution No. 3 is not the solution of the SAT problem. Solution No. 5 means $x_1 = 1$, $x_2 = 0$, $x_3 = 0$. Adding the initiation chain to the test tube. The initiation chain and auxiliary chain will open all the hairpin structures. When $x_2=0$, $\bar{x}_2=1$, $x_1 = 1$ and $x_3 = 0$, so the fluorescence can be detected, that is, solution No. 5 is the solution of the first clause of the SAT problem. The specific response is shown in figure 5. Similar No. 1, No. 2, No. 4, No. 6, No. 7, and No. 8 are all solutions to the first clause of the SAT problem.

![Figure 3. Eight kinds of origami bases for the first clause.](image)

For the second clause $x_1 \lor x_2$, of the satisfiability problem, there are only two variables $x_1$ and $x_2$, and there are $2^2$ possible solutions, each of which is $1(0,0), 2(0,1), 3(1,0), 4(1,1)$, which is recorded as the $1 \sim 4$ solution, so 4 kinds of origami bases are required respectively, as shown in figure 6. Take solution No. 1 (0, 0) and solution No. 3 (1, 0) as examples (figure 7). Solution No. 1 means $x_1 = 0$, $x_2 = 0$. Adding the initiation chain to the test tube, the initiation chain and auxiliary chain will open all the hairpin structures and no fluorescence can be detected, that is, solution No. 1 is not the solution of the SAT problem. Solution No. 3 means $x_1 = 1$, $x_2 = 0$. Adding the initiation chain to the test tube, the initiation chain and auxiliary chain will open all the hairpin structures, and fluorescence can be detected in the test tube, that is, solution No. 3 is the solution of the second clause of the SAT problem. The specific response is shown in figure 8. Similar No. 2 and No. 4 are all solutions to the second clause of the SAT problem.

![Figure 4. Reaction path diagram.](image)

![Figure 5. The reaction diagram of solution No. 3 and No. 5.](image)
For the third clause $\bar{x}_1 \lor x_3$ of the satisfiability problem, there are only 2 variables $x_1$ and $x_3$, and there are $2^2$ possible solutions, each of which is 1(0, 0), 2(0, 1), 3(1, 0), 4(1, 1), which is recorded as the No. 1~ 4 solution, so 4 kinds of origami bases are required, as shown in figure 8. Take solution No. 2 (0, 1) and solution No. 3 (1, 0) as examples. Solution No. 2 means $x_1 = 0, x_3 = 1$. Add the initiation chain to the test tube. The initiation chain and auxiliary chain will open all the hairpin structures. When $x_1 = 0, \bar{x}_1 = 1$, and $x_3 = 1$, so the fluorescence can be detected, that is, solution No. 2 is the solution of the third clause of the SAT problem. Solution No. 3 means $x_1 = 1, x_3 = 0$. Add the initiation chain to the test tube. The initiation chain and auxiliary chain will open all the hairpin structures. When $x_1 = 1, \bar{x}_1 = 0$, and $x_3 = 0$, so no fluorescence is detected, that is, solution No. 3 is not the solution of the SAT problem. The specific response is shown in figure 9. Similar No. 1 and No. 4 are all solutions to the third clause of the SAT problem.

Combining the three clauses of the satisfiability problem $F = (x_1 \lor \bar{x}_2 \lor x_3) \land (x_1 \lor x_2) \land (\bar{x}_1 \lor x_3)$, we can find (0, 1, 1), (1, 0, 1), (1, 1, 1) are the final solutions of the SAT problem. The model uses the complexity of DNA computation to transform the time complexity of the problem into space complexity. The solution space of the problem depends on the number of variables.
4. Conclusion
This paper uses DNA origami and DNA strand displacement to construct a model to solve the satisfiability problem. Firstly, map all the solution spaces of each clause to the origami base. Secondly, add the initiation chain and let it fully react in the test tube. Finally, observe whether there is fluorescence on the DNA origami substrate. If there is fluorescence, it is the solution of a certain clause of the SAT problem, otherwise it is not the solution of the SAT problem. The model constructed in this paper is limited by the DNA origami base, and the number of variables for the SAT problem will also be limited. However, the model designed in this paper has a simple reaction self-assembly structure, and the reaction is easy to operate and realize. Without the addition of initiation chain, the hairpin structure on the original origami substrate in the test tube and the auxiliary chains in the solution remain stable, which greatly reduces the generation of false solutions and has good stability. The reaction can be completed by adding only one initiation chain in the test tube. We need to judge whether there is fluorescence on the DNA origami substrate instead of counting the number of fluorescence signals, which reflects that the model is easy to find solutions.

Acknowledgments
This project is supported by National Natural Science Foundation of China (No.61702008, No.62072296), Anhui Postdoctoral Fund (No.2019B331), Anhui Natural Science Foundation (No.1808085MF193) and 2019 Outstanding Young Talents in Colleges and Universities Overseas Visiting Research Project (No. gxgwfz2019015).

References
[1] Adleman L 1994 Molecular computation of solutions to combinatorial problems Science 266 (5187) 1021-1024.
[2] Rothemund P W K 2006 Folding DNA to create nanoscale shapes and patterns Nature 440 (7082) 297.
[3] Winfree E 1998 Algorithmic Self-Assembly of DNA (California: California Institute of Technology) PHD Thesis.
[4] Yin P, Hariadi R F, Sahu S, et al. 2008 Programming DNA tube circumferences Science 321 (5890) 824-826.
[5] Yin P, Choi H M, Calvert C R, et al. 2008 Programming biomolecular self-assembly pathways Nature 451 318-322.
[6] LaBean T H 2009 Nanotechnology: Another dimension for DNA art Nature 459 (7245) 331-332.
[7] Han D, Pal S, Nangreave J, et al. 2011 DNA origami with complex curvatures in three-dimensional space Science 332 (6027) 342-346.
[8] Zhang F, Jiang S X, Wu S Y, et al. 2015 Complex wireframe DNA origami nanostructures with multi-arm junction vertices Nature Nanotechnology 10 (9) 779-84.
[9] Yu Y, Su S and Chao J 2016 A solution to the coloring problem of design drawings based on “DNA origami” Journal of Nanjing University: Natural Science Edition 52 (4) 656-661.
[10] Tikhomirov G, Petersen P and Qian L L 2017 Fractal assembly of micrometre-scale DNA origami arrays with arbitrary patterns Nature 552 (7683) 67-71.
[11] Chao J, Wang J B, Wang F, et al. 2019 Solving mazes with single-molecule DNA navigators Nature Materials 18 (3) 1-10.
[12] Tang Z, Yin Z X, Sun X, et al. 2019 Dynamically NAND gate system on DNA origami template Computers in Biology and Medicine 109 112-120.
[13] He K, Li Z, Liu L, et al. 2020 Assembly of a DNA origami Chinese Knot by only 15% of the staple strands ChemBioChem 21 (15) 2132-2136.
[14] Khosravi R, Ghasemi R H and Soheilifar R 2020 Design and simulation of a DNA origami nanopore for large cargoes Molecular Biotechnology 62 423-432.
[15] Li C X, Wang H Y, Shen J, et al. 2015 Cyclometalated iridium complex-based label-free photoelectrochemical biosensor for DNA detection by hybridization chain reaction amplification *Analytical Chemistry* 87 (8) 4283-4291.

[16] George A K and Singh H 2017 DNA strand displacement based logic inverter gate design *IET Micro & Nano Letters* 12 (9) 611-614.

[17] Li J X, Wang Y F and Sun J W 2020 Odd judgment circuit of four inputs based on DNA strand displacement *Journal of Nanoelectronics and Optoelectronics* 15 (3) 415-424.

[18] Lipton R J 1995 DNA solution of hard computational problems *Science* 268 (5210) 542-545.

[19] Sakamoto K 2000 Molecular computation by DNA hairpin formation *Science* 288 (5469) 1223-1226.

[20] Braich R S, Chelyapov N, et al. 2002 Solution of a 20-variable 3-SAT problem on a DNA computer *Science* 296 (5567) 499-502.

[21] Wang X, Bao Z, Hu J, et al. 2008 Solving the SAT problem using a DNA computing algorithm based on ligase chain reaction *Bio System* 91 (1) 117-125.

[22] Yang J, Yin Z X and Cui J Z 2015 Satisfiability problem based on self-assembly model of molecular beacons *Journal of Bionanoscience* 9 (3) 197-202.

[23] Nie Q and Yin Z X 2016 A DNA computing method on DNA chip for SAT problem *Journal of Fuyang Teachers College* 33 (003) 53-55.

[24] Cui J Z, Yin Z X, Yang J and Geng X Y 2020 Probe machine based computing model for solving satisfiability problem *Bio-inspired Computing: Theories and Applications* 117.