Problem-Solving Method with Semantic Net Based on DNA Computing in Artificial Intelligence

Yusei Tsuboi, Zuwairie Ibrahim and **Osamu Ono**

Control System Laboratory
Graduate School of Science & Technology
Meiji University
1-1-1 Higashimita, Tama-ku, Kawasaki-shi, Kanagawa, 214-8671, JAPAN
Phone: +81-44-934-7289
Fax: +81-44-934-7909
e-mail: tsuboi@isc.meiji.ac.jp

Abstract

Semantic Net is among the problem solving systems in artificial intelligence fields. In this paper, we demonstrate how to design DNA-typed Semantic Net in order to apply DNA computing to artificial intelligence. Moreover, we propose a problem-solving method with DNA-typed Semantic Net. In this method, it is possible to reason out a reference object by using DNA computing algorithm. Proposed DNA-typed Semantic Net is used as a molecular knowledge based system. Vertexes and edges of the DNA-typed Semantic Net are encoded to four kinds of nucleotide. Single-stranded DNAs are hybridized and ligated to let them the double-stranded DNAs with the complementary sequences of input molecules and knowledge based ones. For the molecular knowledge based system, we estimate the computational complexity by using a simulation. Proposed problem-solving method is performed by DNA-based computer for a future generation of artificial intelligence.

1 Introduction

The development of *computers* has been in progress up to silicon-based computers. However, it is obvious that there is the limitation to make a micro-elemental device of which the computers are composed. In order to realize the further innovative development of the computers, there is an approach to molecular computing aiming at making an element device molecules. In particular, a technique by using DNA (Deoxyribo Nucleic Acid) is called DNA computing. DNA-based computers are a new computational paradigm with massive parallelism and complementality. In general, DNA molecules are composed of single or double-stranded DNA. A single-stranded primer has four based denoted by the symbols ‘A’, ‘T’, ‘C’, and ‘G’. A double-stranded oligo may be formed of two single primers due to hybridization reaction, because ‘A’ is based paired with ‘T’ and ‘C’ is based paired with ‘G’ complementalily [1]. The research of DNA computing started in 1994. It is widely known that Adleman [2] solved Hamilton Path with DNA molecules. The paper suggests that biomolecules solve the network problems by DNA self-assembly. In 2002, a finite automaton operates with DNA molecules in special cases [3]. The report describes applicability to replace modern computers with DNA computer as Turing machine [4]. In this field most of the papers, such as the two papers, have been proposed in order to calculate mathematic problems or to construct mainframe computers, and not to solve problems concerning artificial intelligence.

It is possible to reason out an object by Semantic Net [5] which is one of the problem solving systems in artificial intelligence. We propose a new method to apply a DNA computing strategy to artificial intelligence fields. Newly defined Semantic Net is used as a knowledge based system, and modeled into DNA molecules. Although a basic method for constructing a database with DNA molecules [6] has been already proposed, it does not fill a complicated network like Semantic Net. Artificial intelligence often needs a lot of computational efforts with knowledge bases. Similarly, the size of Semantic net increases exponentially with the amount of knowledge. However, by using Semantic Net based on DNA molecules, it is effective in reducing the computational complexity. In addition, a part of the parallel work of human’s brains is represented. For the molecular knowledge based system, we evaluate the proposed method with a simulation of computational complexity. These approaches indicate molecular biology should be applicable to engineering fields as a molecular application in the future.

2 A problem- solving method by using DNA computing algorithm

In this section, a new problem-solving method by using DNA computing is demonstrated in order to reason out an object. We describe the way to design DNA-typed Semantic Net, and outline chemical processes in DNA computing for the solution.

2.1 Newly defined Semantic Net

Semantic Net would mimic intellectual ability of people if they were based on associative memories. Its structure is a two-dimensional graph like a network. It is relatively easy for a man to deal with Semantic Net, because it represents an object (or concept) constructed from knowledge based on human's memories. The Semantic Net is made of three relations, Object: O, Attribute: A and Attribute value: V. In general, these list representations is denoted as follows.

$$\{<O, A_i, V_j> \mid i=1, 2,\ldots,m; j=1,2,\ldots,n\}$$

A basic Semantic Net as graph is designed with vertexes, edges and label representing their relations as shown in Figure 1. O is reasoned out by relation between A_i and V_j. Because the Semantic Net is simply standardized with vertexes and edge between them, it is suitable for a system to search for some objects in parallel and to be used as a knowledge based system. The size of Semantic Net increases exponentially with the attributes or attribute values.

Vertex ⇔ Object and attribute value
Edge ⇔ Attribute

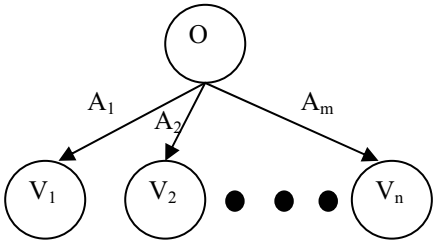
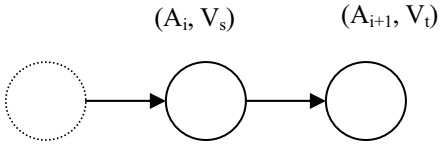


Figure 1: Semantic Net

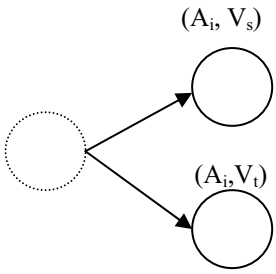
The other hand, if there is a complicated graph, it is imperative to transfer it into simple one. The graph of AND/OR type enables us to lessen the size of the graph and to understand it more easily. We demonstrate a new graph using the type of AND/OR and describe the way to design it. Thus, we do not use the normal existent Semantic Net like Figure 1, but the DNA-typed Semantic Net which is newly designed and defined in order to make the most of DNA computing as follows.

First, a tag as a name of a target object is set to an initial vertex in the graph. After we determine the number and the kinds of the attribute of the target object, both the attribute and attribute value are also set to another vertex following by the tag vertex. Second, the relation between vertexes and edges is represented using a new defined AND/OR graph. In Figure 2-a a directive edge in the terminal direction is connected between the vertexes in series except the following case. If there are two vertexes which have same attributes and different attribute values, each of directive edges is connected in parallel as shown in Figure 2-b.The edge denotes only connection between the vertexes in the directive graph. Finally, labels are attached to the vertexes, such as '(Tag)' and '(Attribute, Attribute value)'.

The vertexes denote either a name of the target or both the attribute and attribute value. In short, one path from an initial vertex to a terminal vertex means one object named on the tag. We define this graph as DNA-typed Semantic Net. The graph explains a target object is reasoned by the combinations between the vertexes. For example, we design the graph in the case of an apple named on the tag as shown in Figure 3. Now consider a method to reason out an object using DNA to present the vertexes, edges and paths in the DNA-typed Semantic Net above: (1) Design and modeling of DNA presenting vertexes and edges in the graph. (2) Reaction of DNA to form paths. (3) Analysis of the reaction to identify paths between initial and terminal vertexes by gel electrophoresis.



2-a: AND



2-b: OR

Figure 2: AND/OR design

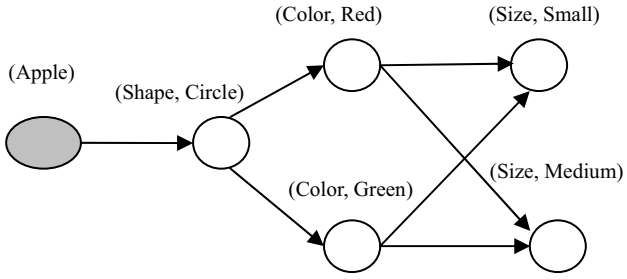


Figure 3: DNA-typed Semantic Net (Apple)

2.2 Design and modeling

In the graph of DNA-typed Semantic Net, every vertex and edge is modeled into a DNA strand as follows:
The vertexes, except the tag's one, are modeled into a short (0-12 nucleotide) piece of a single-stranded sequence by table 1. In table 1, a row shows attributes, a column shows attribute values and DNA sequence is designed by these relations so that it might not over lapping with the

other sequences at random. Every tag vertex of the objects has random sequences of unique number (200, 300, 400…) at the other end to distinguish the objects as shown in Figure 4. The edge $_{(Ai, Vs)\rightarrow(Ai+1, Vs)}$ from vertex$_{(Ai, Vs)}$ to vertex$_{(Ai+1, Vt)}$ are modeled into complementary sequences to the vertex sequences derived from the 3' 6- mer of vertex $_{(Ai, Vs)}$ and from the 5' 6-mer of vertex$_{(Ai+1, Vt)}$, as shown in Figure 5. It is a short (0-12 nucleotide) piece of single-stranded DNA except initial and terminal edges. These two DNA pieces are modeled respectively by the size which suits the end of the DNA pieces of the initial or the terminal exactly.

In this way, the DNA-typed Semantic Net is modeled into double-stranded DNA. Figure 6 shows one ((Apple)→(Shape, Circle)→(Color, Red)→(Size, Medium)) of the paths in the case of the apple is modeled into a double-stranded DNA.

Table 1: Design of attribute and attribute value

Shape		Color		·····		Size	
Circle	TCGATCTACTTA	White	TACGATTCGGAT	·····	·····	Large	TACTCGATACAT
Square	TTGCATCGTTAC	Red	ATCGTACCTGAT	·····	·····	Medium	CAGCTGAAATCA
				·····	·····		
Triangle	ATCCATGGATCG	Green	GTTATTCCCCAG	·····	·····	Small	AATTACGGGATA

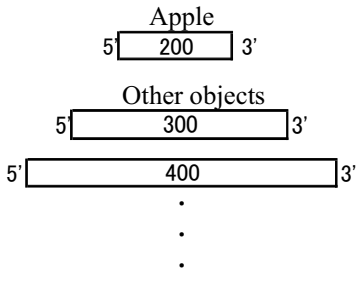


Figure 4: Tag design

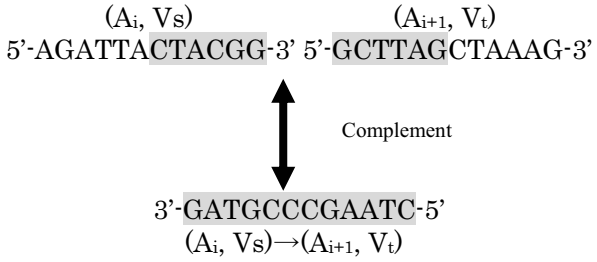


Figure 5: Edge design

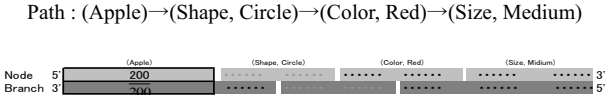


Figure 6: Double- stranded DNA (Apple)

2.3 Reaction

One to P reference objects are reasoned out by using DNA computing algorithm. We describe defined processes of the chemical reaction in DNA computing for the solution by a virtual operation. Figure 7 shows the reasoning system composed of the knowledge base and a premise for the solution. In the knowledge base, each of DNA pieces of edges and tags is synthesized as a knowledge based molecule. In the premise, the attribute values are extracted from 1-P reference objects separately under the previously determined attributes at 2.2. These labels are represented as '(Attribute, Attribute value)'. Using the (Attribute, Attribute value), a piece of a single-stranded DNA is synthesized as an input molecule by Table.1.

We have to amplify each of the knowledge-based molecules and the input molecules sufficiently with PCR (Polymerase Chain Reaction). PCR is a technique which is used to amplify the number of copies of a specific region of DNA, in order to produce enough DNA to be adequately tested. This technique can be used to identify with a very high-probability, disease-causing viruses and/or bacteria, a deceased person, or a criminal suspect.

Figure 8 shows a molecular knowledge based system composed of input molecules and knowledge based molecules in DNA computing algorithm. Now, we prepare virtual 1-P tubes to reason out 1-P reference objects. The knowledge based molecules divided into 1-P equal aliquots are put into each test tube. Each of 1-P input molecules are put into 1 to P test tubes respectively. Figure 9 shows the single-stranded DNAs will anneal to a complementary sequence under defined reaction conditions in each test tube. The DNA sequences representing input molecules and knowledge based molecules are mixed in the presence of a DNA ligase. This enzyme will form a covalent bond between two DNA molecules as long as they have complementary single-stranded overhangs. Thus, the sequences are ligated to form a duplex DNA which represents a path between an initial vertex and a terminal vertex. As a result, all possible double-stranded DNAs representing the paths are generated at random.

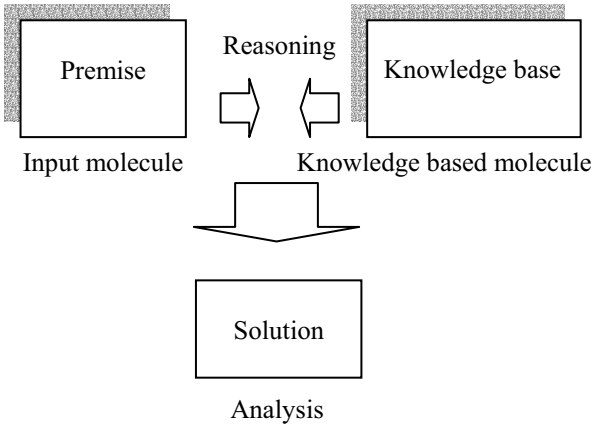


Figure 7: Reasoning system

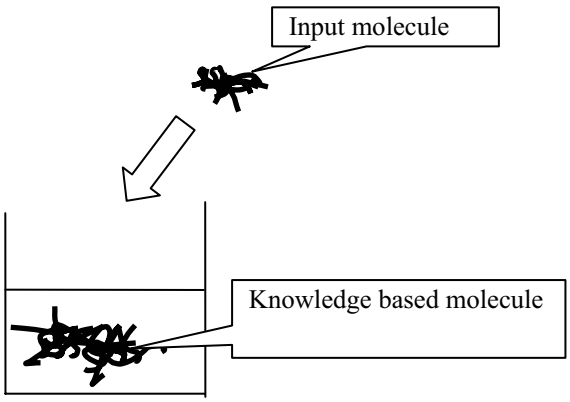


Figure 8: Molecular knowledge based system

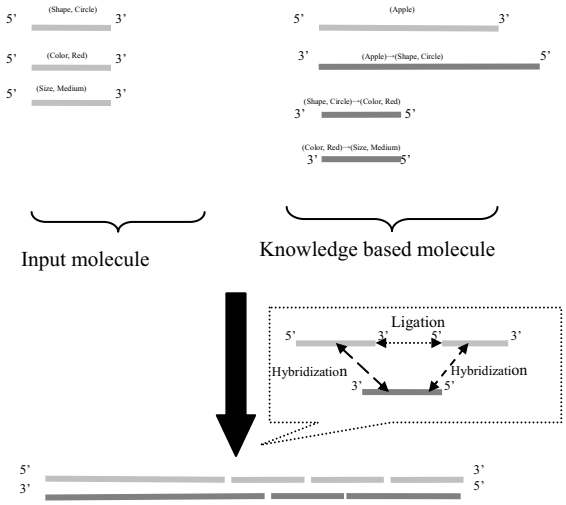
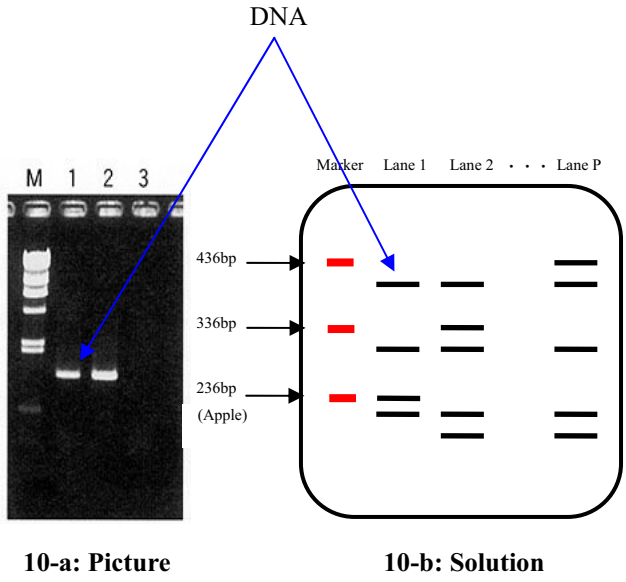


Figure 9: Annealing

10-b. It is possible to distinguish the objects by their size in respective lanes, because of unique number of the tag vertexes. The size of the necessary double strand is given by N_S. It is denoted as follows:

$$N\_S = S\_D \times N\_A + S\_T \quad (1)$$

where S_D is the size of a DNA piece, N_A ($1 \le N\_A \le 7$) is the number of attributes and S_T is the size of the tag. For instance, if a reference object is an apple such as Figure 3, N_D=12, N_A = 3 and S_T = 200, we find out double-stranded DNAs of 248 bp (base pair) exist in lane 1.



10-a: Picture **10-b: Solution**

Figure 10: Gel electrophoresis of reaction

2.4 Analysis

The generated DNAs are analyzed into necessary strands which is double-stranded DNA representing the target objects for the solution. If the necessary strands exist, it comes to that a reasoned reference object is one of the target objects. Generated DNAs are submitted to gel electrophoresis which separates the strands based on size (Figure 10-a). Gel electrophoresis refers to the technique in which molecules are forced across a span of gel, motivated by an electrical current. Activated electrodes at either end of the gel provide the driving force. A molecule's properties determine how rapidly an electric field can move the molecule through a gelatinous medium. It is possible to measure the size of double stranded DNAs by comparing with DNA markers. All the DNA markers of the same size as modeled double-stranded DNAs are prepared to distinguish generated DNA. Therefore, paths which have been are the same size as the marker will appear as bands on the gel.

In 1-P tubes, generated DNAs are put into 1-P lanes respectively. After gel electrophoresis is carried out, the result of the analysis is shown in lane 1 to P on Figure

3 Discussion

In proposed method, there are some chemical processes, PCR, annealing, hybridization, ligation and gel electrophoresis. We survey the several steps for the solution with virtual DNA molecules and we are sure that it is feasible algorithm in real chemical experiment based on the proposed method, because there are many remarkable papers [2][3][6]-[8] including this processes. They were primitively used and regarded as effective for the solution in DNA computing field. Thus, nowadays, we have focused on the proposing a new method for DNA computing application. In this section, we discuss on the effectiveness of proposed method, and view the capacity of DNA pieces for a large scale knowledge base, and view the computational complexity with a simulation result.

3.1 DNA molecules as a knowledge base

In this method, the knowledge (Attribute, Attribute Value) is made a DNA molecule as molecular knowledge, which is the effective in storing a lot of information in knowledge base. One piece, except tag piece, has

information of both the attribute and attribute value of the target object. we have to consider an effective way to select the symbol to avoid the error caused by mismatched hybridization which denotes that a sequence is hybridized with not its complementary sequence. Recently, the research concerning that has been in progress. Whereas some papers [9]–[12] report effective methods to resolve the error, in this research we indicated the effectiveness for the molecular knowledge based system through the produced equation below. We determine the size of a DNA piece representing a tag and/or the knowledge as the unique number and/or twelve nucleotides respectively. These size has the limitation of a capacity to store information of the knowledge. If the number of the target objects and determined attributes in the molecular knowledge based system increases to the limitation of the capacity, the knowledge is not fully able to be encoded into twelve nucleotides. Here, we discuss on the size of a piece of DNA molecule with the number of target objects and attributes to use Semantic Net as a molecular knowledge based system. In modern computing, it is widely known that the amount of information per unit value is defined as 'bit' (0/1). In DNA computing, we determine a unit as symbols, 'A', 'T', 'C' or 'G'. Each of the symbols is selected with the same probability. The amount of information of a DNA piece is given by

$$W(v_{IJ}) = -\log_4 \left(\Big/ \sum_{I=1}^{M} \sum_{J=1}^{N} v_{IJ} \right) \qquad (2)$$

where $W(v_{IJ})$ is a positive integer and v_{IJ} (I =1, 2,.., M ; J =1, 2,.., N) is the number of attributes A_J of target object O_I. $W(v_{IJ})$ equals to the size of one DNA piece. Although in general for organizations, DNA sequence is naturally designed to store their genetic information, for molecular knowledge based system DNA is regarded as a medium of knowledge information. It is possible to store knowledge information represented by combinations of the each symbol. Double-stranded DNA, one path with a meaning, is formed by bonding some knowledge based molecules with each other. A large scale knowledge base is searched for a specific object (or concept) by the DNA-typed Semantic Net in massive parallelism.

3.2 Evaluation

We outlined the way to encode newly defined DNA-typed Semantic Net into sequences and to reason out a reference object. We have to evaluate the advantage of the proposed method by using a DNA computer as compared with a silicon-based computer. It commonly says that it is difficult to evaluate a simulation of DNA computing algorithm on the silicon-based computer. DNA computer integrates software with hardware and calculates in parallel. If the algorithm is simulated on normal silicon-based computer, it will cause combinatorial explosions depending on the size of a problem. Thus, in order to demonstrate the advantage of the proposed method, we estimate computational complexity needed for the solution comparing the DNA computing with the modern computing. It is possible to reason out an object by the combinations between input molecules and knowledge based molecules. In short, the number of the combinations increases with the number of target objects and attributes. This figure 11 shows relations between the attributes and the combinations. X-axis is the number of attributes and y-axis (logarithmic scale) is the number of combinations. On a normal silicon based computer, the number of combinations is carried out by simulating the proposed method with silicon-based architecture and with DNA- based architecture separately when there are 3, 100, and 1000 objects in the molecular knowledge base,. With silicon-based architecture, blue, green and red lines are shown in the case of 3, 100 and 1000 objects respectively. Every three line increases exponentially with the number of attributes. The other hands, with DNA-based architecture, light blue line is shown in all the case of 3 100, and 1000 objects. This line increases logarithmically. The number of combinations does not depend on the number of the target objects. We are sure that it was because the proposed method did not require complicated mathematical calculation due to DNA self-assembly and DNA computers calculate in parallel. The simulation result suggested effective in reducing the computational time under ideal conditions, even if a large scale knowledge base is searched.

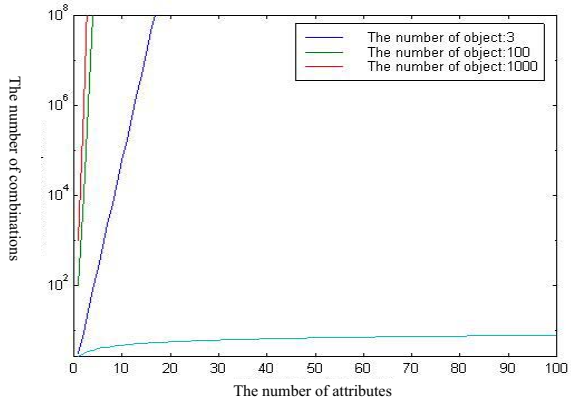


Figure.11 Simulation result

4 Conclusion

In this paper, we have presented a problem-solving method with DNA-typed Semantic Net and several procedures to reason out an object by using DNA computing algorithm. Furthermore, we have discussed on the effectiveness for the proposed method. It is demonstrated by following items.

(1) Newly defined Semantic Net is designed and modeled into duplex DNA to be used as a molecular knowledge based system.

(2) Reaction and analysis based on DNA computing algorithm was outlined with a virtual operation from theoretical point of view.

(3) We considered a capacity as a molecular knowledge base to store the knowledge information.

(4) The simulation result demonstrated the advantage of the proposed method by using a DNA-based computer.

It is important to propose effective problem-solving methods to be performed by a new type of computer like DNA-based computer for a future generation of engineering fields. Proposed problem-solving method suggests DNA computing be applicable to artificial intelligence fields. This approach will be utilized as a natural application system for solutions of pattern matching, deduction and retrieval problems in near future.

References

[1] J.D.Watoson and F.H.C.Crick, *Molecular Structure of Nucleic Acids*, Nature, Vol.171, pp.737--738, April.1953.

[2] L.M.Adleman, *Molecular computation of Solutions to Combinatorial Problems*, Science, Vol266, pp.1021-1024, 1994.

[3] Y. Benenson, T. Paz-Elizur, R.Adar, E.Keinan, Z.Livneh and E.Shapiro, *Programmable and Autonomous Computing Machine Made of Biomolecules*, Nature, Vol.414, pp.430--434, 2001.

[4] Turing.A.M, *On Computable Numbers, with An Application to the Entcheidungproblem*, Proc. Lond.Math. Soc. II Ser. 42, 230--265, 1936.

[5] M.R.Quillian, *The Teachable Language Comprehended, A Simulation Program and Theory of Language*, Comm.of ACM, vol.12, pp.459--476, 1969.

[6] M.Arita, M.Hagiya and A.Suyama, *Joining and Rotating Data with Molecules, IEEE International Conference on Evolutionary Computation*, pp.243--248, 1997.

[7] H.Lim, J.Yun, H.Jang, Y.Chai, S.Yoo, and B. Zhang, *Version Space Learning with DNA molecules*, 8th International Workshop on DNA-based Computer, pp.143--155, 2003.

[8] M.Yamamoto, N.Matsura, T.Shibata, Y.Kawazoe and A.Ohuchi, *Solution of Shortest Path Problems by Concentration Control*, 7th International Workshop on DNA-Based Computers, pp.143--155, 2003.

[9] Russel Deaton, Randy C.Murphy, Max Garzon, D.R. Franceschetti and S. E. Stevens, Jr.,*Good Encodings for DNA –based Solutions to Combinatorial Problems*, DNA Based Computers II DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol.44, pp.247--258, 1999.

[10] John A. Rose, Russel J, Deaton, Donald R.Franceschetti , Max Garzon, and S. Edward Stevens, Jr., *A Statistical Mechanical Treatment of Error in The Annealing Biostep of DNA Computation*, Proceeding of the Genetic and Evolutionary Computation Conference, pp.1829--1834, 1999.

[11] John SantaLucia., Hatim T. Allawi and P.Ananda Seneviratne, *Improved Nearest-Neighbor Parameters for Predicting DNA Duplex Stability*, Biochemistry, Vol.35, No.11, pp.355--356, 1996.

[12] John SantaLucia, Jr., *A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbor Thermodynamics*, Proc. of National Academy of Science U.S.A., vol.1460--1465, 1998.

[13] J.H.Holland, *Adaptation in Natural and Artificial systems*, A Bradford Book edition, The MIT Press, 1992, First edition in 1975.

[14] O.Ono, and K. Yamamoto et al, *Application and Evaluation of DNA Computing Simulation by List Based Processing*, 2002 IEEE CCA/CACSD Complex System Applications, pp825--829, 2002.