DNA Computing Approach to Evolutional Reasoning Algorithm by Using Restriction Enzyme

Yusei Tsuboi, Zuwairie Ibrahim, and Osamu Ono

Institute of Applied DNA Computing
Graduate School of Science & Technology
Meiji University
1-1-1 Higashimita, Tama-ku, Kawasaki-shi, Kanagawa, 214-8671 JAPAN
e-mail: tsuboi@isc.meiji.ac.jp

Abstract

In this paper, we describe how to encode SMC and propose a new reasoning algorithm as experimental protocol of DNA computing techniques, which are enhanced by a concept derived from ADCA. When the SMC is encoded into DNA sequences, the enzyme recognition site involves in each encoded ssDNA representing nodes and edges. The reasoning algorithm shows an object is reasoned out. The ssDNAs are mixed and anneal to complementary sequences under defined conditions in a test tube. In order to retrieve a correct dsDNA which means an object reasoned, a following cycle is repeated. First, the restriction enzyme such as EcoR I cuts a specific part of dsDNAs, reading enzyme recognition sites of the sequence. Second, the cut products are annealed. Finally, the generated products are analyzed by gel electrophoresis. Even if once annealing process runs, the cutting process enables to evolutionarily reuse the same DNA molecules for computation. This repetition stops, when correct strands remain at the analysis process. If the strands do not remain in spite of several repetitions of the cycle, we can say "No". As an output, DNA chips will display the name of the object reasoned.

Keywords: DNA computing, Semantic Network, Algorithm, AI application

1 Introduction

In 1994, L. Adleman's [1] ground-braking work demonstrated the way to use DNA molecules for computational purposes. This experience also contributed into a better understanding where to go with DNA machines, namely, to try to develop memory machines that are machines with very large memory that implements rather simple search operations.

In 2004, a semantic model was proposed and described theoretically for DNA-based memories by Tsuboi, *et al* [2]. This model, referred to as 'semantic model based on molecular computing' (SMC) has the structure of a graph formed by the set of all attribute-attribute-value pairs contained in the set of represented objects, plus a tag node for each object. The objects representing double-stranded DNAs (dsDNAs) will be formed via parallel self-assembly, from encoded single-stranded DNAs (ssDNAs) representing the attribute attribute-value pairs (nodes), as directed by splinting ssDNAs representing relations (edges) in the network. The computational complexity of the implementation is estimated via simple simulation, which indicates the advantage of the approach over a simple sequential model. According to this report, if such reasonable computation is realized in *vitro*, a huge number of DNA molecules will be needed in advance as the size of the graph increases. Thus, we have to generates massive initial pools in the first implementation step and then filter the candidate solutions which satisfy the given conditions. To successfully decrease the initial pools, there are a few different initial pool generation methods, parallel overlap assembly (POA) introduced by Stemmer [3], the mix and sprit method introduced by Faulhammer *et al*. [4], with their own advantages and disadvantages. In addition to these methods, the adaptive DNA computing algorithm (ADCA) with a feedback structure was introduced by Yamamoto*, et al*. [5] in 2004. We strongly support the ADCA because this algorithm requires only simple and reliable operations: annealing, cutting by a restriction enzyme, polymerase chain reaction (PCR) and gel electrophoresis. The ADCA is applied to a shortest path problem for a mobile robot. The simulation results indicated to extremely reduce the number of DNA molecules needed as compared with a simple Adleman's model.

In this paper, we propose an evolutional reasoning algorithm which uses the SMC and the ADCA. We will review the ADCA in section 2 and the SMC in section3. Section 4 explains the evolutional reasoning algorithm by using a restriction enzyme, as experimental protocols. Section 5 presents the discussion and conclusion.

2 Encoding Scheme

Many works on DNA computing have employed the encoding scheme presented by Adleman. In Adelman's experiment, each of edges and nodes within the small

Hamiltonian path graph presented by him is represented by ssDNAs of 20 nucleotides respectively. These codes were constructed at random; the length 20 is enough in order to ensure that the codes are "sufficiently different". The edge $u \rightarrow v$ from node u to node v is described to be Watson-Crick complementary to the node sequences derived from the 3' 10- mer of the node u and from the 5' 10-mer of the node v. For instance, the codes of the node 1, 2 and the edge 1→2 specified below

node1: 5'TATCGGATCGGTATATCCGA3'
node2: 5'GCTATTCGAGCTTAAAGCTA3'
edge1→2: 3'CATATAGGCTCGATAAGCTC5'

As for an encoding scheme of ADAC, let us suppose that we have code α, β and complementary code $\overline{\alpha}$, $\overline{\beta}$ which are parts of recognition site of a restriction enzyme EcoR I. α and β are respectively assigned 'AATTC' and 'G'. That is to say, $\overline{\alpha}$ and $\overline{\beta}$ are respectively assigned 'TTAAG' and 'C'. A dsDNA involving codes α, β, $\overline{\alpha}$, $\overline{\beta}$ is cut at the set part by EcoR I as shown in Figure 1.

α and β code are embedded between the sequences derived from the 3' 10- mer of the node u and from the 5' 10-mer of the node v. $\overline{\alpha}$ and $\overline{\beta}$ are attached to the 5'end of the edge $u \rightarrow v$. In this way, the codes of the node 1, 2 and the edge 1→2 are modified below,

node1: 5'TATCGGATCG | α β | GTATATCCGA3'
node2: 5'GCTATTCGAG | α β | CTTAAAGCTA3'
edge1→2: 3'CATATAGGCTCGATAAGCTC| $\overline{\alpha}$ $\overline{\beta}$ |5'

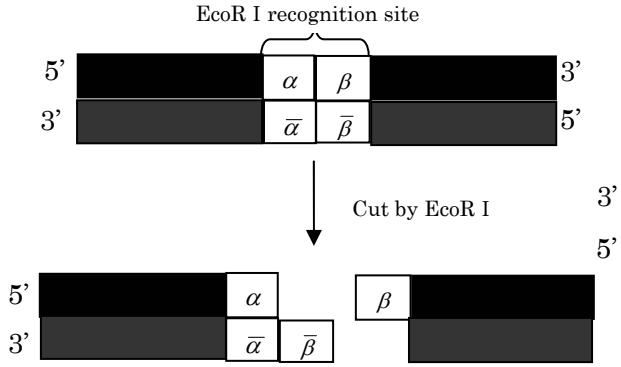

Figure 1 A dsDNA is cut by enzyme EcoR I

3 Semantic Model Based on Molecular Computing (SMC)

Figure 2 describes an SMC formed by the union of a set of some objects. It is made of three relations: object, O; attribute, A; and attribute-value, V. This list representation is denoted as follows:

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

A tag as a name of an object is set to an initial node in the graph. Both the attribute and attribute-value are also set to another node following by the tag node. The nodes denote either a name of the object or both the attribute and the attribute-values. In short, one path from an initial node to a terminal node means one object named on the tag. The model represents an object, as reasoned out by the combinations between the nodes connected by the edges. An SMC contains all attributes common to every object as well as each attribute-value. Attribute layers consist of attribute-values, lined up. If an object has no value of a certain attribute, the attribute value is assigned '*no value*'.

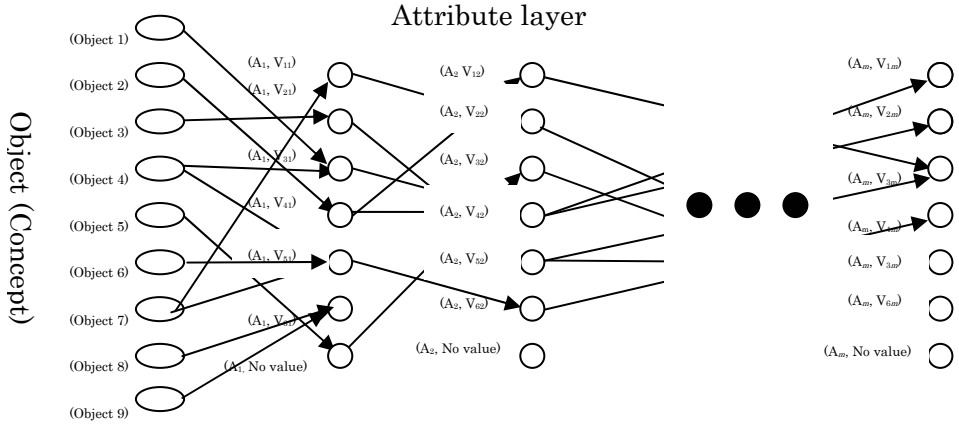

Figure 2 A semantic model based on molecular computing (SMC), which collectively models a set of objects, given a total number of attribute layers, m.

4 Methodology

In this section, we propose a reasoning algorithm by using a restriction enzyme, which are enhanced by a concept derived from the ADCA.

4.1 DNA Representation of SMC

Each of the nodes and edges within an SMC may be represented by a DNA library. In the DNA library, a row shows attributes and a column shows attribute-values. DNA sequence is designed by these relations so that it might not be overlapping with the other sequences at random. Firstly, with Adleman's encoding scheme, except for initial and terminal edge, each nodes and edges is assigned ssDNA oligonucleotide of length 20. As for tag nodes, the sequences are also assigned by random 20 bases. The initial and terminal edges are respectively represented by the size which suits the end of the DNA pieces exactly. Here, an important thing is that every sequence is designed according to these relations to prevent mishybiridization via other unmatching sequences. Next, we innovate the concept of ADAC encoding scheme in the Adleman's encoding scheme. The code α, β $\overline{\alpha}$ and $\overline{\beta}$ of the enzyme recognition site of EcoR I involves in each encoded ssDNA representing nodes and edges, except for the initial node, the terminal node and the terminal edge. Figure 3 shows DNA representation of one of the object within the SMC.

4.2 Algorithm

The reasoning algorithm shows an object is reasoned out by DNA computing techniques. Semantic information of some reference objects is stored in a semantic memory as knowledge bases. The algorithm reveals that which reference object several input objects are classified into with DNA molecules all at once. Figure 4 explains overall procedures of DNA operations required for solutions.

A set of DNA representing reference objects and an input object fragments, formed by the combinations of oligonucleotides, are synthesized as follows:

- Reference object

The ssDNA of each edge and tag node in the network is synthesized as a set of *knowledge based molecules.*

- Input object

Attribute-values are extracted from an input object according to determined attribute A_i. Using the attributes and the attribute-values, an ssDNA is synthesized as a set of *input molecules* with the sequence defined by the DNA library.

The ssDNAs are mixed and anneal to complementary sequences under defined conditions in a test tube. In order to retrieve a correct dsDNA which means an object reasoned, the generated products are analyzed from DNA length by gel electrophoresis. If they remain, say "Yes": the object is reasoned out, otherwise, we should consider about two cases. The one is "No": the object is not reasoned out. The other is that the number of DNA molecules prepared in advance is too few to form the correct dsDNAs. To distinguish the two cases, a following cycle is repeated. First, the EcoR I cuts the all the analyzed products, reading the enzyme recognition site of the sequences. Second, the cut products are annealed again. Figure 5 illustrates ligation & hybridization process after the cutting. Finally, the generated products are analyzed as well. Even if once annealing process runs, the cutting process enables to evolutionarily reuse the same DNA molecules for computation. This repetition stops, when correct strands remain at the analysis process. If the strands do not remain in spite of several repetitions of the cycle, we can say "No".

4.3 Output

DNA Chips, output in readable format is accomplished by attaching the cloned, coding sequences to an array. Thus, each spot would represent an object. Readout occurs directly form sensing fluorescent tags attached to a tag sequence of the correct strands, as probes.

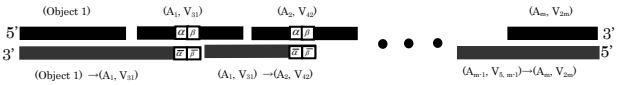

Figure 3 DNA representation of one of the objects

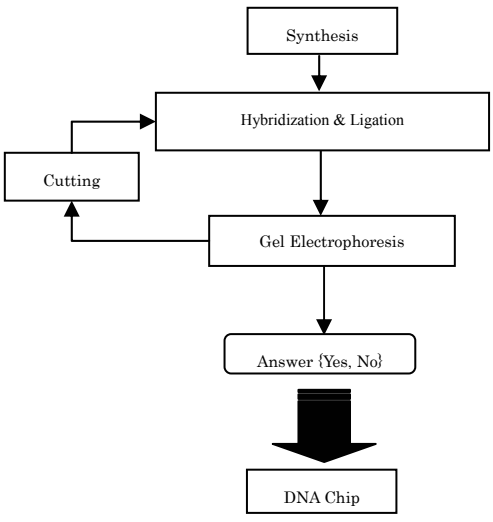

Figure 4 Overall procedures of DNA Operations

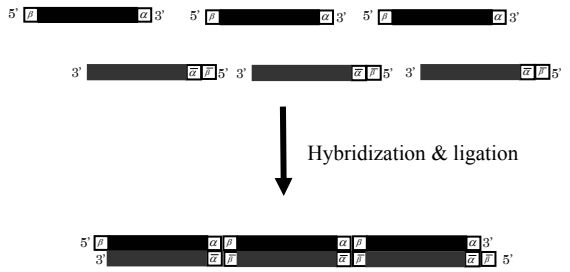

Figure 5 Hybridization & ligation process after cutting

5 Discussion & Conclusion

We have been discussed an evolutional reasoning algorithm by using DNA computing techniques. This paper provides a necessary DNA computing-chemical process including experimental operations. An adopted protocol is very simple. Standard genetic engineering techniques such as annealing, ligation, cutting and gel electrophoresis are required. Several experiments have been conducted to assess the performance of DNA-based databases realized by biochemical reactions. Within the SMC, attributes and attribute-values are represented by random (0-20) oligonucleotides {A, T, C, G}. There are two issues to consider length and sequences design of DNA. The one is that if a set of knowledge increases, oligonucleotides, length 20 are not fully assigned to each of nodes and edges in the network. Such length limits to represent a lot of objects. This issue would be resolved simply by assigning longer oligonucleotides length. The other is that in practical sequences design, we have to consider an effective way to select proper sequence to avoid mismatched, error hybridization will have to be devised. Word design strategies for DNA-based computation have been investigated so far. Substantial progress has been reported on this issue [6]-[10]. We expect that this issue will be satisfactorily resolved in the near future. Thus, these issues are very crucial to obtain a correct answer. In the light of these issues, we will have to design best sequences with adequate length, when a laboratory experiment is done.

The SMC is one of the models for applying DNA computing techniques to the research field of artificial intelligence. Its application has many incredible advantages, whereas the reasonable performance demands a huge number of molecules. The proposed algorithm will enable to minimize useless molecules synthesized. In addition, it repeats cutting, annealing and analysis processes to reach the solution only, which interests us in terms of an intelligent mechanism based on DNA computing. It is expected that the proposed algorithm would extend many AI applications, knowledge bases, pattern matching, etc. As a future work, to achieve reliable performance, some parameters of reactions-temperatures, concentrations of oligonucletides, times of reactions, etc. will be experimentally tested.

Reference

[1] L. Adleman, “Molecular computation of solutions to combinatorial problems”, *Science* 266, pp.583-585, 1994.

[2] Y. Tsuboi, Z. Ibrahim, and O. Ono, “Semantic model for artificial Intelligence based on molecular computing”, Proc. of *The 8th International Conference on Knowledge-Based Intelligent Information and Engineering Systems*, Lecture Notes in Artificial Intelligence (LNAI), Vol. 3213, pp. 32-39, 2004.

[3] W. Stemmer, “DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution, Proc. of Natl. Acad. Sci., pp.10747-10751, 1994.

[4] D. Faulhammer, A. R. Cukras, R. J. Lipton, and Laura F. Landwever, “Molecular computation: RNA solutions to chess problems, Proc. of *Natl. Acad. Sci.*, pp1385-1389, 2000.

[5] O. Ono, Y. Tsuboi, T. Yamamoto, Z. Ibrahim, and S. Watanabe, “Computation algorithm based on adaptive DNA molecules by using polymerase chain reaction”, Pcoc. of *20th Fuzzy System Symposium* pp. 83-86, 2004 (in Japanese).

[6] R. Daton, M. Garzon, R. C. Murphy, J. A. Rose, D. R. Franceschetti, and S. E. Stevens Jr, “Reliability and efficiency of a DNA-based computation”, *Phys. Review Letters* 80, pp.417-420, 1998.

[7] R. Deaton, C. R. Murphy, M. Garzon, D. R. Franceschetti, and S. E. Stevens Jr.,”Good encodings for DNA-based solutions to combinatorial problems”, Proc. of *DNA Based Computers DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, Vol.44, pp.247-258, 1999.

[8] J. A. Rose, R. Deaton, D. Franceschetti, M. Garzon, and S. E. Stevens Jr., “A statistical mechanical treatment of error in the annealing biostep of DNA computation”, Proc. of *Genetic and Evolutionary Computation Conference*, pp.1829-1834, 1999.

[9] J. SantaLucia, H. Allawi, and P. Seneviratne, “Improved nearest-neighbor parameters for predicting DNA duplex stability”, *Biochemistry,* Vol.35, No.11, pp.355-356, 1996.

[10] R.Penchovsky, and J. Ackermann, “DNA library design for molecular computation”, *Journal of Computational Biology*, Vol.10, No.2, pp.215-229, 2003.