A Molecular Inference via Colorimetric Change

Jan J. Mulawka[1]

[1] Warsaw University of Technology, Institute of Electronic Systems, Warsaw, Poland,
email: JMulawka@elka.pw.edu.pl

Abstract. In this work the inference performed on DNA molecules is considered. The process known as self-assembling of DNA strands is fundamental to achieve longer chains of molecules encoding knowledge base. By proper organization of this process any logic operations can be accomplished. In particular, systems based on decision trees may be easy assembled. A severe problem of nano-systems is to provide communication between nano-world and real-world. To solve this problem it is proposed to make use of colorimetric change phenomenon. Here an inference procedure is proposed that allows to check the knowledge base in form of decision tree. Although DNA strands react in the nano-world, due to optical read-out the inference result is attained quickly and easy. Thus, complicated and expensive operations of genetic engineering laboratory are eliminated.

1 Introduction

Recent progress in nanotechnology has led to the development of nanoscale materials that have proven useful in the development of novel techniques for biology and medicine [1, 2, 3, 4]. For example molecular events can be sensed and detected in laboratory using three main formats: optical detection, electrical detection, and magnetic detection. Optical detection remains the most widely used mechanism for detecting biological binding events and for imaging in biological systems. In the future, the goal will surely be to enable single molecule detection in vivo, despite the large background present in a living system.

In nanotechnology one may distinguish an interesting process known as self-assembly of DNA molecules [10] in a water solution. A research in this area started in 1994 after pioneering work of Adleman [13] and is known DNA computing. Henceforth much efforts have been put to show different applications of self-assembling DNA strands. Also a number of studies have been conducted on molecular reasoning [5, 6, 7, 8, 11] where DNA approach is applied. A characteristic feature of these methods is that an inference path is created during hybridization of two complementary strands. The reasoning is based on classical logic where two states: true and false may be distinguished [9]. In some problems we are interested in proving just a hypothesis. In this situation by using given knowledge base we ask if the hypothesis can be proved based on known facts. It should be noted that in this case we are interested to get final result of the reasoning. Partial results are of less importance. Because the process is performed in nano-world (access to which is difficult due to very small dimensions) therefore of great importance are control and communication with the world of particles. To overcome this is very crucial

problem for success of nanotechnology. To this purpose different detection methods may be applied.

It has been observed in recent years that with respect to DNA detection, particles that are heavily functionalized with oligonucleotides exhibit unusual hybridization properties that have led to advantages with respect to assay sensitivity and selectivity. The binding affinity of deoxynucleosides (dA, dG, dC, dT) to gold nanoparticles was studied by many researchers. It has been demonstrated that when the particles are hybridized to complementary oligonucleotides, either in solution or immobilized on a flat surface, they exhibit extraordinarily sharp melting profile resulting in a change of solution color from red to blue which is known as colorimetric change. This phenomenon seems to be very promising to find an easy way to communicate between nano-world of particles and real-world. The primary objective of this contribution is to show how the colorimetric change may be used to develop a method of molecular inference.

2 Colorimetric responses used in the optical detection of DNA

In 1996 it was observed that oligocleotide-modified nanoparticles and sequence-specific particles could be used to generate materials with unusual optical and melting properties [1, 2, 3]. Studies indicated that 13-nm gold particles used in the assay changed the color of the solution from red to blue upon the analyte-directed aggregation of gold nanoparticles as depicted in Fig. 1. This simple phenomenon pointed toward the use of nanoparticles as DNA detection agents in a type of of "litmus test" for nucleic acid targets.

In Fig. 1 gold nanoparticles are represented by the circles. To the surface of the Au particles oligonucleotide strands are bound. In the solution there are also complementary target DNA strands. Let us denote these target strands by W and respective sequences by K, L as is shown in Fig. 2. In the presence of complementary target DNA, during hybridization gold nanoparticles will aggregate as depicted in Fig. 2.

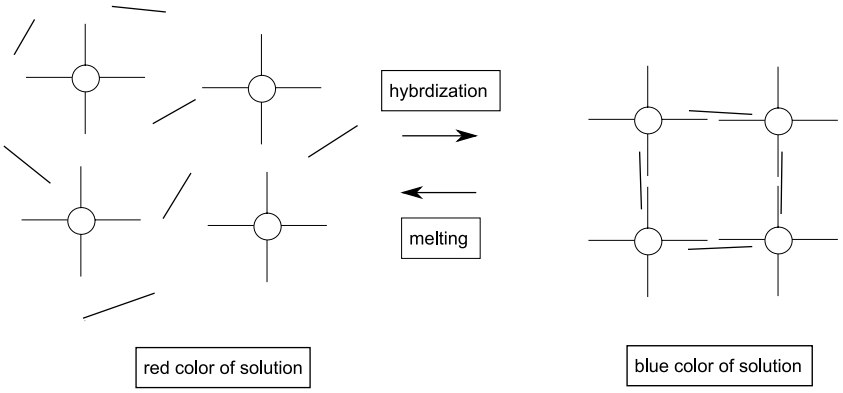


Figure 1. Change of solution color from red to blue after the hybridization of DNA strands

It was found that spotting the solution onto white support enhanced the colorimetric change and provided permanent record for each test. Also it should be noted that the melting profiles of the nanoparticle-labeled DNA aggregates were extraordinarily sharp, occurring over a temperature range much more narrow than the transition for unlabeled

or conventional fluorophore-labeled DNA. Thus, the colorimetric change may be utilized as a simple and inexpensive way of detecting nucleic acids. In further points we show how to use this phenomenon to develop an easy method of molecular reasoning.

3 Organization of the knowledge base

We assume that knowledge base is organized as wood of decision trees as is provided in Fig. 3. Each tree is composed of three kinds of nodes: root node, intermediate nodes and leaves nodes. A structure of particular decision tree consists of a number of node levels as is shown in Fig. 4. In this exemplary tree one may distinguish 7 nodes which are denoted as A, B, C, H.At the highest level a root (A) is located while at the lowest level leaves (H) are located. Between root and leaves there are intermediate nodes (B, C). Any number of leaves and intermediate nodes is possible to be in a tree. Also different connections between these nodes are possible.

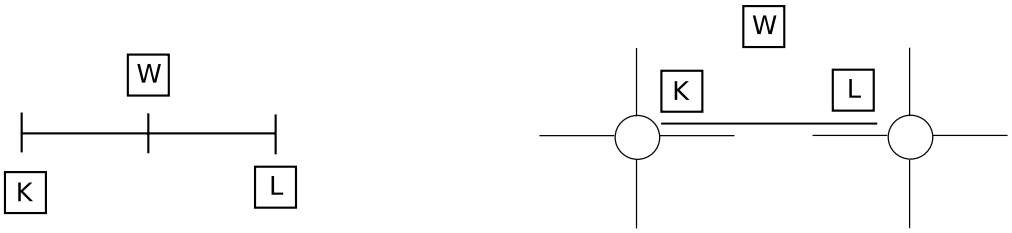


Figure 2. Structure of target DNA used to aggregate gold nanoparticles

Particular nodes in decision tree represent logic variables. The leaves represent conclusions or hypothesis. In our tree it is assumed that all conclusions are the same (H). Such restriction follows from the nature of DNA computing (in this case due the colorimetric change phenomenon which allows to distinguish a single logic state). Nodes in decision tree are connected by some branches. Nodes and branches create graph of the decision tree. The branches represent values of logic variables at particular nodes. In general graph may be complicated with a great number of nodes, branches and levels.

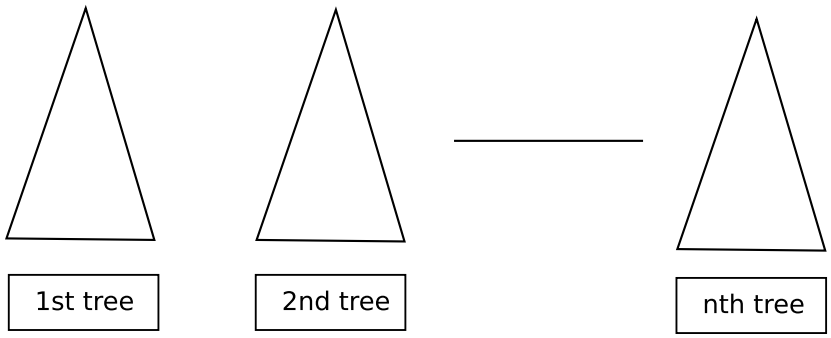


Figure 3. A knowledge base composed of n trees

A decision tree may be interpreted as a set of logic rules. A particular rule is created by an inference path from root node to a leaf passing through intermediate nodes. As an example consider the decision tree presented in Fig. 4. This tree consists of 7 nodes and 6 branches. There are two intermediate nodes and four leaves. The nodes are located in three levels. It can be easy shown that this tree represents a set of 4 rules which may be written in the form:

$$\begin{array}{llllll} \textbf{IF} & \textbf{A} = \textbf{a} & \textbf{AND} & \textbf{B} = \textbf{x} & \textbf{THEN} & \textbf{H} \\ \textbf{IF} & \textbf{A} = \textbf{a} & \textbf{AND} & \textbf{B} = \textbf{y} & \textbf{THEN} & \textbf{H} \\ \textbf{IF} & \textbf{A} = \textbf{b} & \textbf{AND} & \textbf{C} = \textbf{e} & \textbf{THEN} & \textbf{H} \\ \textbf{IF} & \textbf{A} = \textbf{b} & \textbf{AND} & \textbf{C} = \textbf{d} & \textbf{THEN} & \textbf{H} \end{array} \tag{1}$$

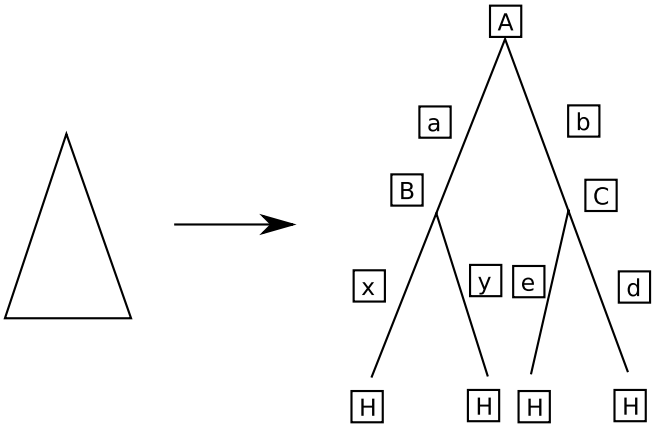


Figure 4. An example of simple decision tree structure

In general, since our knowledge base is composed of n trees, it may be written as a set of rules for particular tree similar to equations (1).

4 Inference process via the self-assembling of DNA molecules

To proceed any inference a knowledge base and data base are required. Our knowledge base is created by a set of above rules, however data base consists of a number of logic variables which create a set of initial facts. A process of inference may be considered as a production performed on set of initial facts. The primary objective of inference is to prove the conclusion (or hyphothesis). Since in our case all rules have the same conclusion H, we ask if based on this rules and some initial facts the conclusion H is true or false. In other words we are looking for an inference path from root to a leave in any tree. If such path there exists the conclusion H is true otherwise it is false.

In DNA computing data is represented by DNA strands. In previous contributions we described a number of molecular inference procedures. Here, we focus our attention on the method based on assembling decision tree [6, 8]. A severe problem of DNA computing is communication between nano-world and real-world. To overcome necessity to apply complicated and expensive equipment of genetic engineering laboratory we propose to take advantage of colorimetric change phenomenon.

In his approach we proceed as follows. During inference due to self-assembly of oligonucleotides longer chains are created. Thus, proving certain set of rules. Because all of rules in our knowledge base have the same conclusion H it is possible to treat this conclusion as a target DNA to aggregate gold nanoparticles. In order to extract the target W from the strand representing the conclusion H one may apply restriction enzyme as shown in Fig. 5. Restriction enzymes are endonucleases [10] produced by bacteria that typically recognize specific 4- to 8-bp sequences, called restriction sites, and then cleave both DNA strands at this side. Here the sequence R is marked out to recognize restriction site for a selected restriction enzyme.

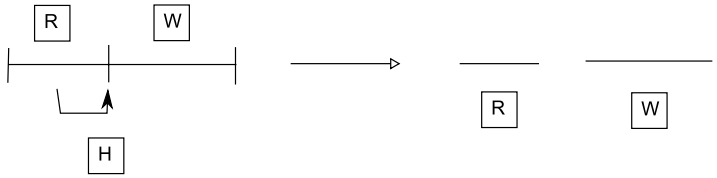


Figure 5. Extracting the target W from the conclusion H

Prticular steps of the inference process using colorimetric change phenomenon may be proceed as follows:

1. Prepare water solution of the knowledge base and the initial facts,
2. Mix the solutions in assay,
3. Perform hybridization and ligation process,
4. Wash the solution to remove unbound strands,
5. Cut leaves using restriction enzyme,
6. Aggregate gold nanoparticles,
7. Observe whether colorimetric change occurs (blue – true, red – false).

As is seen the above procedure is very simple and straightforward. A complicated operations of genetic engineering like Polimerase Chain Reaction (PCR), fluorophore labeling, electrophoresis, and others have been eliminated. Thus, expensive sophisticated equipment of genetic engineering laboratory is not required. Particular steps of this procedure may be performed in automatic way using standard operations of genetic engineering. Therefore the presented procedure is suitable for miniature implementation in form lab-on-a-chip.

5 Illustrative examples

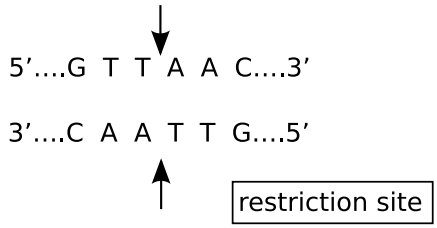
To explain discussed here molecular inference method we consider the following examples. First, let the knowledge base consists of a single decision tree as is depicted in Fig. 4. To encode this tree in form of DNA molecules we choose sequences of oligonucleotides as follows:

- Root – magnetic *A: 3' GCTTA 5'*
- Intermediate nodes *B: 3' ACATTGTACC 5', C: 3' ACTTCCTAGG 5'*
- Leaves *H: 3' GCTACTCCAATTG 5'*

- Initial facts *a: 3' CATGGCGAAT 5', b: 3' GATCCCGAAT 5', x, y: 3' GTTAACTGTAA 5', e, d: 3' GTTAACTGAAG 5'*

The target W is then composed of two oligonucleotide sequences:
K: 3' GCTAC 5'
L: 3' TCCAA 5'

The recognition sequence R is chosen to be:



As is required for restriction enzyme Hha I [10]. Note that after digestion we obtain molecule with blunt ends.

Let us assume now that we have two initial facts: a – true, y – true. Having encoded knowledge base as well as initial facts we ask if the hypothesis H can be proved based on facts a, y. Proceeding according to the above inference procedure we obtain the path starting from leaf (site 3') and terminating at the root as follows:
3' GCTACTCCAATTGACATTGTACCGCTTA 5'

From this strand we can easy retrieve the target W. Having digested this strand with the enzyme Hha I one can cleave the following oligonucleotide molecule:
3' GCTACTCCAA 5'
which sequence is the same as the strand W. Next by adding this oligonucleotide as target DNA to the assay with gold nanoparticles, a change of solution color from red to blue will be observed. Note that result of our test is optical, therefore straightforward and very fast. It is said to be a type of "litmus test" for nucleic acid targets.

In the second example suppose that initial facts are: a – true, d – true.

As follows from the graph of Fig. 4 in this case an inference path can not be created. And as a consequence the target strand W can not be produced. As a result gold nanoparticles are not aggregated and therefore color of the solution remains red which means false. So, in this case initial facts a, d are not sufficient to prove conclusion H.

6 Conclusions

Presented method of inference offers several advantages over other techniques described in [5, 8]. First of all it eliminates such operations like Polimerase Chain Reaction (PCR), fluorophore labeling, electrophoresis and others. As has been demonstrated the method is quick and easy. Its optical read-out does not require expensive and sophisticated instrumentation. The method is adequate for implementation in form lab-on-a-chip and may find a numerous applications in molecular biodiagnostics.

Bibliography

[1] R. Elghanian, J.J. Storhoff, R.C. Music, R. L. Letsinger, C.A. Mirkin, Selective colorimetric detection of polynucleotides based on the distance-dependent optical properties of gold nanoparticles, *Science*, 1997, 277, 1078-1081.

[2] J.J. Storhoff, R. Elghanian, R.C. Music, C.A. Mirkin, R.L. Letsinger, *J. Am. Chem. Soc.*,1998, 120, 1959-1964.

[3] N.L. Rosi, C.A. Mirkin, Nanostructures in biodiagnostics, *Chem. Rev.*, 2005, 105, 1547-1562.

[4] A.P. Alivisatos, Small May More, *Świat Nauki*, nr 11 (123), 2001, 57-63 (in Polish).

[5] P. Wąsiewicz, T. Janczak, J.J. Mulawka, A. Płucienniczak, The inference based on molecular computing, *Cybernetics and Systems Int. Journal*, vol. 31, no. 3, 2000, 283- 315.

[6] P. Wąsiewicz, T. Janczak, J.J. Mulawka , A. Płucienniczak, The inference via DNA computing, *Proc. Congress on Evolutionary Computation*, Washington, vol. 2, 1999, 988-993.

[7] J.J. Mulawka, P. Borsuk, P. Węgleński, Implementation of the inference engine based on molecular computing technique, *Proc. of the IEEE Int. Conf. on Evolutionary Computation*, Anchorage, 1998, 493 498

[8] J.J. Mulawka, T. Janczak, P. Borsuk, P. Węgleński, Reasoning via DNA based decision trees, *Proc. Int. Conf. on Rough Sets and Current Trends in Computing*, Warsaw 1998, 17-26.

[9] J.J. Mulawka, *Expert Systems*, WNT, Warszawa, 1996 (in Polish).

[10] P. Węgleński, *Molecular Genetics*, PWN, Warszawa, 1996 (in Polish).

[11] J.J. Mulawka, M.J. Oćwieja, Molecular inference via unidirectional chemica reactions, *Int. Conf. on Evolvable Systems: From Biology to Hardware*, Lausanne, Lecture Notes in Computer Science, 1478, Springer, 1998, 372-379.

[12] J.J. Mulawka, Molecular computing promise for new generation of computers, Proc. Polish-Czech-Hungarian Workshop on Circuit Theory, Signal Processing and Applications, Budapest, 1997, 94-99.

[13] L. Adleman, Molecular computation of solutions to combinatorial problems, *Science*, vol. 266, 1994, 1021-1024.

[14] R. J. Lipton, DNA solution of hard computational problems, *Science*, 268, April 28, 1995, 542-545.

[15] M. Ogihara, A. Ray, Simulating boolean circuits on a DNA computer, Technical Report TR 631, University of Rochester, Computer Science Department, August 1996.

[16] P. Wąsiewicz, A. Malinowski, R. Nowak, J.J. Mulawka, P. Borsuk, P. Weglenski, A. Płucienniczak, DNA computing: Implementation of data flow logical operations, *Future Generation Computer Systems*, EIsevier, 17, 2001, 361-378.

[17] J.J. Mulawka, P. Wąsiewicz, A. Plucienniczak, Logical operations with DNA strands, *Proc. Int. Conf. on Rough Sets and Current Trends in Computing*, Warsaw 1998, 27- 36.

[18] P. Wąsiewicz, P. Borsuk, J.J. Mulawka, P. Węgleński, Implementation of data flow logical operations via self-assembly of DNA, *Lect. Notes in Computer Science*, 1586, Springer, 1999, 174-182.

[19] E. B. Baum, Building an associative memory vastly larger than the brain, *Science*, 268, April 28, 1995, 583585.

[20] R. Nowak, P. Wąsiewicz, A. Malinowski, J.J. Mulawka, A. Płucienniczak, Processing DNA tokens in parallel computing, *Proc Int. Parallel and Distributed Processing Symposium*, San Francisco, 2001, 120-128.

[21] P. Wąsiewicz, R. Rudnicki, J.J. Mulawka, B. Lesyng, Adding numbers with DNA, *Proc. IEEE Int. Conf. on Systems, Man & Cybemetics*, Nashville, 2000, 265-270.

[22] P. Wąsiewicz, J.J. Mulawka, R. Rudnicki, B. Lesyng, Arithmetic operations on DNA - subtraction, *Seventh Int. Meeting on DNA based computers*, 2001

[23] J.J. Mulawka, P. Wąsiewicz, K. Piętak, Virus-enhanced genetic algorithms inspired by DNA computing, *Lect. Notes in Artificial Intelligence*, Subseries LNCS,1609, Springer, 1999, 529-537.

[24] P. Wąsiewicz, J.J. Mulawka, Molecular genetic programming, *Soft Computing*, Springer, vol. 5, no. 2, 2001, 106-113.

[25] J.J. Mulawka, T. Janczak, A. Malinowski, R. Nowak, DNA computing promise for information processing, *Universitatis Jagellonicae Acta Informaticae*, MMCCXLII, Z. 10, Kraków, 2000,113-130.