

DNA COMPUTING AND ITS APPLICATIONS: SURVEY

ROHANI BINTI ABU BAKAR AND JUNZO WATADA

Graduate School of Information, Production and System
Waseda University
2-7 Hibikino, Wakamatsu, Kitakyushu 808-0135, Japan
rohani@ump.edu.my; junzow@osb.att.ne.jp

Received December 2007; accepted February 2008

ABSTRACT. *The aim of the paper is to make a review of DNA computing achievements on current stage, especially on new approaches or methods contributing to solve either theoretical or application problems. Starting with the NP-problem that Adleman solved by means of wet DNA experiment in 1994, DNA becomes one of the most appropriate alternatives to overcome the silicon computer limitations. Today, many researchers concentrate on this research subject either to improve available methods used in DNA computing itself or to suggest a new way to solve engineering or application problems with a DNA computing approach. This paper gives an overview of research achievements in DNA computing and touches on the achievements of improving methods employed in DNA computing as well as in solving application problems. At the end of discussion we address several challenges that DNA computing are facing in the society.*

Keywords: DNA computing, Molecular computing, DNA computing application, Biological inspired computing

1. Introduction. DNA computing is one interdisciplinary research area that is growing fast since DNA molecules are implemented in a computational process. One of the main objectives of this research area is to produce, in near future, a biologically inspired computer based on DNA molecules to replace or at least beneficially complement with a silicon based computer. Since R. Feynman has suggested to construct a computer from molecules in 1964 [1]. It took 30 years till Adleman in 1994 making proof of the principle study that DNA molecules can solve an NP problem of Hamiltonian Path Problem (HPP) through bio-chemical procedure [2].

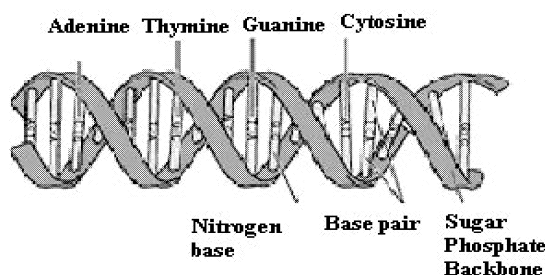
DNA is a basic storage medium for all living cells. The main function of DNA is to absorb and transmit the data of life for billions years. Roughly, it is around 10 trillions of DNA molecules could fit into a space, the size of a marbles. Since all these molecules can process data simultaneously, theoretically, we can calculate 10 trillions times simultaneously in a small space at one time. DNA computing is more generally known as molecular computing. It is interdisciplinary field where it is a combination of biology, chemistry, mathematics and computer science. Computing with DNA offers a completely new paradigm for computation. The main idea of computing with DNA is to encode data in a DNA strand form, and laboratory techniques of molecule biology, called as bio-operations, will be involved to manipulate DNA strands in a test tube in order to simulate arithmetical and logical operations. It is estimated that a mix of 1018 DNA strands could operate 10^4 times faster than the speed of a today's advanced supercomputer [3].

Since then, DNA computing is the area of exciting multidisciplinary research. Rozenberg et al. in 1999 distinguished two major lines of research in DNA computing as (i) the theoretical line concerned with models, algorithms and paradigms for DNA computing and (ii) the experimental line concerned with the design of laboratory experiment to test the biochemical feasibility [4]. Even though there is still a long way to implement DNA algorithm in real life, but researchers are interested in modeling and testing the solution in

a case study in order to challenge the limitation of DNA itself. Today, many researchers in this field develop models and do the laboratory experiment especially in challenges of biochemical feasibility. In the mean time, some people concern with developing a real DNA computer and building DNA algorithms to solve engineering or application problems.

The paper is organized as follows. Section 1 is devoted to the introduction of this research topic. A definition of DNA computing is briefly provided in this section. In order to have a better understanding of a DNA computing approach, Section 2 will discuss the basic structure and techniques of DNA computation which are available in molecular biology for DNA processing. These techniques can be considered as a basic toolbox for the experiment of DNA computing. From there, achievements in improving experimental and theoretical methods will be discussed in Section 3. However, we will put emphasis on the achievements in improving models, algorithms and paradigms in solving engineering and application problems. In Section 4, we are going to address some challenges and promise in this area. We complete our discussion with concluding remarks in Section 5.

2. Molecular Biology Components. A deoxyribonucleic acid or DNA molecule is found in every cellular organism as a storage medium for genetic information. It is a polymer constructed from monomers called nucleotides distinguished by their chemical group of bases that attached to them. Consisting of three main components; sugar, phosphate and base. There are four different bases; adenine, guanine, cytosine and thymine, abbreviated A; G; C and T, respectively. Since nucleotides differ only in terms of their bases, abbreviations of bases are used to identify them. In mathematical modeling it can be illustrated as $X = \{A, C, G, T\}$. Figure 1 illustrates a basic DNA structure.



It consists of four distinct bases; Adenine (A), Thymine (T), Guanine (G) and Cytosine (C). A will always bind with T and G always with C that represent Watson-Crick complementary law.

FIGURE 1. Basic structure of DNA molecules

Every single nucleotide is linked together end-to-end to form DNA strands. A short single-stranded polynucleotide chain, usually less than 30 nucleotides long, is called an *oligonucleotide* (or, shortly, oligo). The DNA sequence has a polarity: a sequence of DNA is distinct from its reverse. The two distinct ends of a DNA sequence are known under the name of the 5 - end and the 3 - end, respectively. The most important feature of DNA is the ability to bind together between two single strands called as Watson-Crick complementarity.

Bonding between single strands occurs when two different bases attract with each other. In this complementarity, A will bond only with T and G only with C. Pairs of (A,T) and (G,C) are, therefore, known as a complementary base pair. The two pairs of bases form hydrogen bonds between each other with two bonds between A and T and three between G and C. A different number of hydrogen bonds between these two pairs give

a different strength for them. In other words, a pair of G and C is stronger than a pair of A and T. It is one of the characteristics that are always considered in designing DNA strands for DNA computing. The classical double helix of DNA is formed as shown in Figure 1 when two separate strands bond. Two requirements must be fulfilled for this to occur; firstly, the strands must be complementary, and secondly, they must have opposite polarities. In DNA computing, we adopt several bio-chemistry operations to replace current computational tools. In next subsection, we will briefly discuss basic operations that adopted from bio-chemistry operations as a computing tool in this field.

2.1. Operations.

Synthesis. Synthesis is a process of designing and restructuring information in DNA sequence form. In DNA computing, designing and synthesizing information in the DNA sequence form is an important process where wrong design might leads to wrong result.

Ligation and hybridization. DNA ligation is a process of joining two single linear DNA fragments together. More specifically, DNA ligation involves creating a phosphodiester bond between 3 -hydroxyl of one nucleotide and the 5 -phosphate of another. Meanwhile hybridization is a process of combining complementary single-stranded nucleic acids into a single molecule. Nucleotides will bind to their complement under normal conditions, so two perfectly complementary strands will bind to each other at the end of the process.

Polymerase chain reaction (PCR). PCR is a process that quickly amplifies the amount of specific molecules of DNA in a given solution using primer extension by a polymerase. DNA polymerases execute several functions including the repair and duplication of DNA. Each cycle of the reaction doubles the quantity of the molecules, giving an exponential growth in the number of operations.

Gel electrophoresis. Gel electrophoresis is a technique to sort DNA strands based on their length or weight through a gel such as agarose gel, in electrical field based on the fact that DNA is negative charge. Larger or longer strands will travel slowly into positive charge and after some time, the strands spread to different bands according their size of length.

3. Achievements in Current Research. In this twenty-four year, after DNA computing was invented by Adleman [2], a lot of achievements have been reported by researchers of DNA computing either in theoretical or practical parts. Today, researchers in this field more concentrate on developing methods for testing biochemical feasibility with wet experiment, there are some groups who concentrate on developing a DNA computer itself and developing algorithms to solve engineering or application problems. Even though developing a real DNA computer is still a long way in front of us, but developing and building algorithms in solving today's application problems are important too in order to test and simulate the stability and reliability of DNA computing algorithm.

In order to solve today's application problems, researchers faced some limitation of manipulation in this field especially in some routine steps in DNA computing techniques. In this section, several significant achievements in this field are discussed. These achievements are explained in two categories: achievements in biochemical feasibility and achievements in solving engineering or application problems.

3.1. Achievements in testing biochemical feasibility of biochemical experiments. In this section, because a lot of researchers deal with a variable of numerical values in solving application and engineering problems, we focus our discussion on achievements of DNA computing methods in presenting numerical values.

In particular, there are four ways in presenting numerical value in DNA computing. Presenting numerical value is an important part to solve weighted graph problem in this research area. Incorrect way to present a weighted graph value can be driven a wrong

result at the end. The four ways of presenting numerical value in DNA computing are widely employed as (i) constant length based [2][5], (ii) direct proportional length based [6],(iii) concentration control [7] and (iv) gradient temperature method [8].

In Adleman's experiment in 1994, he employed the constant length based method to present distances between two cities [2]. However, in this experiment, Adleman did not put the labels on the arcs to represent the distances between cities. As well as Lipton in 1995 [9] proposed an enhancement of Adleman's model but still did not deal with any information yet regarding distances between cities. The first models that deal with information or labels on arcs were done by Narayanan and Zorbalas in 1998 [5] in solving a weighted graph problem.

Narayanan and Zorbalas have proposed to employ constant based length to represent an information of arcs, in their case, distances between cities. In this algorithm, for example, distance 1 will be presented by 3-mer of DNA meanwhile distance 2 will be presented by 6-mer of DNA and so on. As a result, the longer distance will be presented by the longer DNA strand and shorter distance by shorter strand. At the end of algorithm, the shortest strand will correspond to the optimal result for the problem.

However, in this technique, the number of distances is limited to employ, because if we employ too many distances, the size of strands will become longer and longer. It is not advisable to employ too long DNA strands in DNA computing because some error comes out during other processes such as mutation or error on reading. In Narayanan and Zorbalas' study, there are not implemented in any laboratory experiment. So that, Ibrahim et al. took an initiative to make an experiment for solving weighted graph problem by implementing direct-proportional length based in 2004 [6] by solving shortest path problem with considering 5 cities and 7 edges with costs.

Ibrahim et al. have proposed a new alternative approach to overcome constant-proportional length based disadvantage, named as direct-proportional length based technique. In this technique, the cost of an edge is encoded as a direct-proportional length oligos. After an initial pool generation and amplification, since numerous numbers of feasible candidates are generated, by using the standard bio-molecular laboratory operations, it is possible to extract the optimal combination which represents a solution to the shortest path problem.

On the other hand, Yamamoto et al. proposed to employ concentration control in solving weighted graph problem in 2000 [7]. In this technique, since chemical reactions are controlled using a DNA concentration, the concentrations of DNA are used as input and output data. Yamamoto et al. believed this technique can reduce the experiment operation costs in detecting process of DNA computing, because this technique has to extract and analyze only relatively intensive bands. In this technique, the concentrations of complementary oligonucleotide encoding vertices are set to the same values, and the relative concentration D_{ij} of each oligonucleotide encoding edge $i \rightarrow j$ with cost C_{ij} is calculated as explained in [7].

In 2004, Lee et al. proposed a novel encoding method to solve weighted graph problem [8]. This method was utilized a temperature gradient to overcome a drawback of previous methods. Melting temperature method uses fixed-length DNA strands and represent costs by melting temperatures of given DNA strands. In this technique, cost or weight for each arc is designed with various melting temperatures according to the values. A smaller value or cost is represented by a DNA sequence with a lower melting temperature, and therefore a more economical path has a lower melting temperature. On the other hand, each city sequence is designed with the same melting temperature because city sequences should contribute equally to the thermal stability of paths. Finally, road sequences that connect two cities are generated using the sequences of departure cities, arrival cities and costs.

However, until now, the researchers still continuing to find out most optimal solution in presenting numerical values. Solution for this problem will be open a new horizon in this field to solve engineering and application problems.

3.2. Achievements in improved or proposed models, algorithms and paradigms in solving engineering or application problems. Although a lot of researchers are still focusing on solving shortest part problems, there are some other researcher groups that take an initiative to solve other applications cryptography [11][12][22], scheduling [16][10][19][20][26][27], clustering [28][29], encryption [18][21][13], forecasting [23] and even tried to employ it in signal and image processing application [15][17]. In this section, we will discuss briefly recent application problems that have been solved using DNA computing method.

Zhixing et al. proposed an DNA computing based algorithm to solve job scheduling problem in 2006 [20]. The authors illustrated the working operation problem in order to explain their proposed model where six tasks are considered. With the attention of solving this problem, Zhixing et al. have mimiced the method used in HPP. It is not the first time DNA is employed to solve scheduling problems. In 2005, Watada et al. [16] proposed a DNA algorithm to schedule elevator systems and this work was polished by Jeng et al. [24] and Muhammad et al. [10] in 2006. In 2007, Bakar et al. [26][27] proposed another DNA computing model to solve re-arrangement of flexible manufacturing systems (FMS) in production line. However, because of limitation to represent numerical values in DNA, all the researchers, so far only consider a small or medium size of scheduling problems to illustrate their proposed solution.

On the other hand, some other researchers in this field are working on proposing DNA algorithm employed in information security technology [14]. For example, Boneh et al. [21] and Adleman et al. [18] have proposed a model to break a Data Encryption Standard (DES) as a alternative way for encryption data technology. DNA cryptography has been proposed by Gehani et al. [11], Kartalopoulos [12] and Tanaka et al. [13] as a new born cryptography field. Beside DNA cryptography and DES, there are some development in DNA steganography and DNA certification. Recently, DNA is employed as a intrusion detection model for computer and telecommunication systems by Boukerche et al. [30]. Among all DNA computing models proposed in this research area DNA certification is most matured and the application is most widely studied [14].

In optimization fields, several methods or models have been proposed to solve application problems. Most of these problems require huge processing time and considering a number of feasible combinatorials in order to find out optimum solution. For example, Bakar et al. have proposed a model to solve clustering problem in mutual distance [29] and proximity approaches [28]. On the other hands, Jeng et al. have introduced a merging technique between DNA computing and fuzzy set to forecast a money exchange rate [24]. Meanwhile, Kim et al. have solved optimum re-arrangement of clique density in a company using DNA computing [25].

4. Challenges and Promises. The main issue in implementing DNA computing technologies to solve real application is how we can present numerical values especially when a number of numerical values are related in DNA strands form. Recently various researches have done and still investigating in order to solve this problem. Researchers have proposed several techniques as discussed before to solve this problem. However, at the time, all proposed solutions are only suitable for the limited number of numerical values and not tested for a number of numerical values. This problem still is open to solve. Solving problems in presenting numerical values in DNA strands form will enable DNA computing more practically to solve a lot of engineering and real application problems.

Developing robust method in wet lab experiment to solve engineering problems are critically essential in DNA computing. One of important process in wet lab experiment is to reading an end result during the experiment. Recently, gel electrophoresis, where the strands will be sorted by their bands, is the most popular technique for this step. However, gel electrophoresis has their own limitation where this limitation should be disadvantage for DNA technique. One of the drawback of gel electrophoresis technique is coming from the fact we cannot analyze the gel images in one time for all bands when we are dealing with a number of base pairs. It is that because some bands especially the earlier one might not exist in the buffer reader yet. So that, we are only able to read a certain part of bands in one time. It will be difficult to made analysis process pursued properly.

Another important technique in wet technologies of DNA computing is PCR. PCR 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 diseased person, or a criminal suspect. However, traditional PCR itself has several limitations that may affect results in DNA computing. Thus, several researchers focus on enhancing this technique to overcome this limitations. As a result, real time PCR is employed in several wet experiment in order to enhance the readability of end results from the experiment real time PCR.

Even though current difficulties found in translating theoretical DNA computing models into real life are not sufficiently overcome, there is still potential for other areas of development. DNA computing offers a new approach to solve combinatorial problems such as NP-hard problems in parallel. This advantage offers a potential to solve problems that faced by a traditional machine in processing a number of tasks. Thus, considering this benefit, researchers are able to solve a problem, especially one dealing with a number of calculations such as optimization of clustering , scheduling problem and so on.

On the other hand, DNA is capable to store a lot of information in small space compared to digital way of storing information. Back to today's situation we are dealing with huge size of information. Today, our information not only in word or document yet, but also in images, video format and so on in these formats require a huge size of storage. So, DNA seems to offer a right choice to solve today storage problem. As a started research dealing with huge size of storage, Tsiftaris et al. [15] have proposed a solution to employ DNA in signal processing field and Tsuboi et al. in image processing [17].

Adams stated in his study [31] three main reasons why DNA computation is practical, firstly, there are a specific computer will be easier to design and implement, with less need for functional complexity and flexibility; secondly, DNA computing may prove entirely inefficient for a wide range of problems, and directing efforts on universal models may be diverting energy away from its true calling; thirdly, the types of hard computational problems that DNA based computers may be able to effectively solve are of sufficient economic importance that a dedicated processor would be financially reasonable.

With so many possible advantages over conventional techniques, DNA computing has bright development potential for practical use. Future work in this field should begin to incorporate cost-benefit analysis so that comparisons can be more appropriate with existing techniques.

5. Concluding Remarks. In this paper we reviewed current technologies that are available in DNA computing research field. We also reviewed some current methods and algorithms that have been developed in DNA computing recently. In recent years, researchers in DNA computing have started to employ these approaches in solving industrial and engineering problems. However, other researchers are concentrated in developing new techniques in wet technologies. As sequences, DNA technologies have improvement significantly from last decade in theoretical aspects as well as applications.

As a conclusion, DNA computing is one of the newest exciting areas to be explored by researchers. There are a lot of opportunities in expanding and manipulating DNA characteristics and operations to solve real application especially industrial engineering and management engineering problems. If we considered all the advantages offered by DNA computing, it should become an alternative way to solve difficulty faced by current silicon computer. However, until now, there are still some obstacles in employing this method into engineering problems as we discussed in Section 4.

Acknowledgment. The first author would like to thank to Universiti Malaysia Pahang (UMP) and Kementerian Pengajian Tinggi Malaysia for a study leave.

REFERENCES

- [1] R. P. Feynman, *Miniaturization*, New York, Reinhold, pp.282-296, 1961
- [2] L. M. Adleman, Molecular computation of solutions to combinatorial problems, *Sciences*, vol.266, no.5187, pp.1021-1024, 1994.
- [3] L. Kari, From micro-soft to bio-soft: Computing with DNA, *Biocomputing and Emergent Computation: Proc. of the BCEC97*, World Scientific, Skovde, Sweden, pp.146-164, 1997.
- [4] G. Rozenberg and A. Salomaa, DNA computing: New ideas and paradigms, *Lecture Notes in Computer Science*, Springer-Verlag, vol.7, pp.188-200, 2006.
- [5] A.Narayanan and S.Zorbalas, DNA algorithm for computing shortest paths, *Proc. of the Genetic Programming*, Morgan Kaufman, pp.718-723, 1998.
- [6] Z. Ibrahim, Y. Tsuboi, O. Ono and M. Khalid, Direct-proportional length-based DNA computing for shortest path problem, *International Journal of Computer Sciences & Applications*, vol.1, no.1, pp.46-60, 2004.
- [7] M. Yamamoto, N. Matsuura, T. Shiba, Y. Kawazoe and A. Ohuchi, DNA solution of the shortest path problem by concentration control, *Lecture Notes in Computer Science*, vol.2340, pp.203-212, 2004.
- [8] J. Y. Lee, S. Y. Shin, T. H. Park and B.-T. Zhang, Solving traveling salesman problems with DNA molecules encoding numerical values, *Biosystems*, vol.78, no.1-2, pp.39-47, 2004.
- [9] R. J. Lipton, DNA solution of hard computational problems, *Sciences*, vol.268, no.5210, pp.542-545, 1995.
- [10] M. S. Muhammad, Z. Ibrahim, O. Ono and M. Khalid, Application of length-based DNA computing for complex scheduling problem, *International Journal of Information Technology*, vol.12, no.3, pp.100-110, 2006.
- [11] A. Gehani, T. LaBean and J. Reif, DNA-based cryptography, *Lecture Notes in Computer Science*, vol.2950, pp.167-188, 2004.
- [12] S. V. Kartalopoulos, DNA-inspired cryptographic method in optical communications, authentication and data mimicking, *Proc. of the IEEE on Military Communications Conference*, vol.2, pp.774-779, 2005.
- [13] K. Tanaka, A. Okamoto and I. Saito, Public-key system using DNA as a one-way function for key distribution, *Biosystems*, vol.81, no.1, pp.25-29, 2005.
- [14] G. Cui, L. Qin, Y. Wang and X. Zhang, Information security technology based on DNA computing, *Proc. of the 2007 IEEE International Workshop on Anti-counterfeiting, Security, Identification*, Xiamen, China, pp.288-291, 2007.
- [15] S. A. Tsiftaris, A. K. Katsaggelos, T. N. Pappas and E. T. Papoutsakis, How can DNA computing be applied to digital signal processing?, *Signal Processing Magazine, IEEE*, vol.21, no.6, pp.57-61, 2004.
- [16] J. Watada, S. Kojima, S. Ueda and O. Ono, DNA computing approach to optimal decision problems, *Proc. of the 2004 IEEE International Conference on Fuzzy Systems*, vol.3, pp.1579-1584, 2004.
- [17] Y. Tsuboi and O. Ono, Applicability of DNA computing algorithm solving image recognition in intelligent visual mechanics, *Proc. of the 29th IEEE Annual Conference, Industrial Electronics Society*, vol.3, pp.3800, 2003.
- [18] L. M Adleman, P. W. K. Rothmund, S. T. Roweis and E. Winfree, On applying molecular computation to the data encryption standard, *Journal of Computational Biology*, vol.6, no.1, pp.53-64, 1999.
- [19] F. Zhang, B. Liu, W. Liu and J. Xu, A DNA computing model based on acryditeTM gel technology to solve the timetable problem, *IEEE/ICME International Conference on Complex Medical Engineering*, pp.1632-1635, 2007.

- [20] Y. Zhixiang, J. Cui, Y. Yang and Y. Ma, Job shop scheduling problem based on DNA computing, *Journal of Systems Engineering and Electronic*, vol.17, no.3 pp.654-659, 2006.
- [21] D. Boneh, C. Dunworth and R. Lipton, Breaking DES using a molecular computer, *Technical Report, CS-TR-489-95*, Princeton University, 1995.
- [22] G. Xiao, M. Lu, L. Qin and X. Lai, New field of cryptography: DNA cryptography, *Chinese Science Bulletin*, vol.51, no.12, pp.1413-1420, 2006.
- [23] D. J.-F. Jeng, J. Watada, B. We and J.-Y. Wu, Fuzzy forecasting with DNA computing, *Lecture Notes in Computer Science*, vol.4287, pp.324-336, 2006.
- [24] D. J.-F. Jeng, I. Kim and J. Watada, Bio-soft computing with fixed-length DNA to a group control optimization problem, *Soft-Computing, Springer*, vol.12, no.13, pp.223-228, 2007.
- [25] I. Kim, D. J.-F. Jeng and J. Watada, Redesigning subgroups in a personnel network based on DNA computing, *International Journal Innovative, Computing, Information and Control*, vol.2, no.4, pp.885-896, 2006.
- [26] R. B. A. Bakar and J. Watada, A bio-soft computing approach to re-arrange a flexible manufacturing robot, *Lecture Notes in Computer Science*, vol.4694, pp.308-315, 2007.
- [27] R. B. A. Bakar and J. Watada, Biological computation of optimal task arrangement to multiple robot flexible machining cell, *Proc. of the International Conference on Soft-Computing and Human Sciences*, Kitakyushu, Japan, pp.61-66, 2007.
- [28] R. A. B. Bakar, J. Watada and W. Pedrycz, A proximity approach to DNA based clustering analysis, *International Journal Innovative, Computing, Information and Control*, vol.4, no.5, 2008.
- [29] R. A. B. Bakar, J. Watada and W. Pedrycz, DNA approach to solve clustering problem based on a mutual order, *Biosystems*, vol.91, no.1, pp.1-12, 2008.
- [30] A. Boukerche, K. R. L. Juca, J. B. Sobral and M. S. M. A. Notare, An artificial immune based intrusion detection model for computer and telecommunication systems, *Parallel Computing*, vol.30, no.5-6, pp.629-646, 2004.
- [31] J. C. Adams, On the application of DNA based computing, *available online at: [http : //publish.uwo.ca/ jadams/dnaapps1.htm](http://publish.uwo.ca/~jadams/dnaapps1.htm)*, Last reviewed 14th January, 2008.