

# An Improved Binary Particle Swarm Optimization Algorithm for DNA Encoding Enhancement

Mohd Saufee Muhammad, Krishna Veni Selvan and  
Sharifah Masniah Wan Masra

Department of Electronics Engineering  
Faculty of Engineering, Universiti Malaysia Sarawak  
94300 Kota Samarahan, Sarawak, Malaysia  
msaufee@feng.unimas.my, kveni52@yahoo.com,  
wmmasnia@feng.unimas.my

Zuwairie Ibrahim and Amar Faiz Zainal Abidin  
Center for Artificial Intelligence and Robotics (CAIRO)  
Department of Mechatronics and Robotics  
Faculty of Electrical Engineering  
Universiti Teknologi Malaysia  
81310 Skudai, Johor, Malaysia  
zuwairie@fke.utm.my, amarfaiz@fke.utm.my

**Abstract**—The accuracy of DNA computing highly depends on the DNA strands used in solving complex computations. As such, many approaches are proposed to design DNA oligonucleotides that are stable and unique. In this paper, an improved binary particle swarm optimization (IBPSO) algorithm is proposed and implemented. Four objective functions which are H-measure, similarity, hairpin and continuity are employed to define the uniqueness of designed sequences. The DNA words are constrained within a predefined range of GC-content and melting temperature. The performances and the ability of the algorithm to enhance the characteristics of generated DNA code words are analyzed. The results obtained show that this algorithm executes better sequences and did perform better compared to other optimization techniques. Moreover, it converges faster than the previously suggested binary particle swarm optimization algorithm.

**Keywords**—DNA oligonucleotides; DNA computing; improved binary particle swarm optimization (IBPSO)

## I. INTRODUCTION

Deoxyribonucleic acid or DNA is the biological macromolecule which carries genetic information in all living species. The molecular compositions of DNA are made up of subunits called nucleotides. Each nucleotide consists of a phosphate group, a sugar (deoxyribose) and one of four possible nitrogen-containing bases, adenine (*A*), thymine (*T*), guanine (*G*) or cytosine (*C*). In 1953, Watson and Crick [1] proposed the idea that the DNA is a double stranded molecule held together in a double helical structure. Their second idea was that the two strands of DNA are complementary to each other. Adenine forms two hydrogen bonds and pairs only with thymine while guanine forms three hydrogen bonds and pairs only with cytosine. These base pairing conditions are well known as Watson-Crick complementary base pairing. Thus, the two DNA strands in the double helix form are in opposite directions, one in the 5'-3' direction and the other in the 3'-5' direction [2]. DNA strings are so extensive that can have an enormous information storage capability. This characteristic of DNA guided to a novel idea of DNA-based computations, as a solution for complex problems.

## II. DNA COMPUTING

DNA computing is a new highly potential method for various complex computational problems. It is a simulation of DNA structures and computing using some molecular biological technology [3]. DNA computing is also generally known as molecular computing. Computing with DNA leads to an overall new standard for computation as it comprises a combinations of biology, chemistry, mathematics and computer science. In DNA computation, the information stored in DNA strand is encoded and several laboratory molecular biology operations are used to manipulate the DNA strands in a test tube. The arithmetical and logical operations will then be carried out to solve the problem [4]. An initial research on DNA computing was pioneered by Adleman in 1994 [5]. He solved a directed Hamiltonian path problem (HPP) through DNA computing techniques [6]. Since then, DNA computing field has been well known and many researchers had carried out further researches to diverse its applications and methods [7] - [11].

## III. DESIGN CRITERIA

Perfect hybridization between a DNA sequence and its base-pairing complement is required for a successful DNA computing, in order for the information stored in DNA molecules to be retrieved easily. Therefore, a good DNA library is crucial and it can be achieved by considering some design criteria into the design process. In this paper, four objective functions, which are H-measure, similarity, hairpin, and continuity and two constraints, GC-content and melting temperature are employed as the design criteria. The selections and the calculations of all the design criteria applied in this research are based on Shin *et al.* [12]. The objective functions and constraints are described as follows:

### A. H-measure

H-measure calculates the number of nucleotides that are complementary among two sequences so as to prevent cross-hybridization [7]. Two sequences are arranged in opposite directions and their positions are shifted during this calculations.

---

This research is supported by the Ministry of Higher Education (MOHE), Malaysia under Fundamental Research Grant Scheme (FRGS) funding.