

Ensembles of DNA Letters for the Design of Unique DNA Library Using a Modified Version of Multi-criteria VEDEPSO Optimizer

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

Department of Electronics Engineering

Faculty of Engineering

Universiti Malaysia Sarawak

94300 Kota Samarahan, Kuching, Sarawak, Malaysia

kveni52@yahoo.com, msaufee@feng.unimas.my, wmmasnia@feng.unimas.my

Abstract—The DNA code words designing is a multi-criteria combinatorial optimization task. The designed words should be as unique as possible, thermodynamically stable, non-self hybridized, non-cross hybridized with others and have good chemical properties. In this paper, the DNA words designing approach implied concurrent minimizations of four objective functions, H-measure, similarity, hairpin and continuity. The designations is subjected to a predefine range of melting temperature and GC-content. A novel multi-population optimizer, M-VEDEPSO, is employed to design sets of DNA strands. The algorithm runs for 10 times and as a result, each population has lower average fitness values compared to the fitness values obtained using the conventional VEDEPSO algorithm. The results obtained from the algorithm are indicated by 12 randomly selected non-dominated particles/individuals. These solutions are obtained via Pareto dominance concepts.

Keywords—fitness; Pareto; population; thermodynamic

I. RESEARCH BACKGROUND

DNA code words designing is a process of arranging the four DNA alphabets, letters or codes known as adenine (A), thymine (T), guanine (G) and cytosine (C) within a predefined length. These arrangements are then evaluated using some combinatorial constraints. DNA words designing are well known as a complex multi-criterion constrained optimization task. Thus, many research works had been proposed to solve this complex optimization problem. For instance, exhaustive search [1], random search [2], graph method [3], dynamic programming [4], template-mapping [5, 6], theoretical approach [7–10], *in vitro* methods [11–13], evolutionary algorithm [14, 15], genetic algorithm [16–22], simulated annealing [23, 24], taboo search [25], bee swarm genetic algorithm [26], artificial fish swarm algorithm [27], ant colony optimization [28], particle swarm optimization [29–32] and hybrid algorithms [33–37]. In this paper, a hybrid methodology of discrete particle swarm optimizer (PSO) and discrete differential evolution (DE) is practiced for solving the combinatorial optimization.

II. IMPORTANCES OF DNA LIBRARY DESIGN

Ensembles of DNA alphabets create a unique DNA library which is mainly used for molecular computing, DNA nanostructure design, DNA tagging in chemical libraries and DNA microarray design [38].

DNA strands are massively parallelism and have an enormous information storage capacity. Unlike the conventional silicon based computations, molecular computing is a highly reliable DNA based computers for solving difficult computations or NP-hard problems. It involves some *in vitro* laboratory experiments to extract the solutions. Unique DNA dictionaries are used to represent each solution in molecular computing so as to ease the computation procedures. Perfect hybridizations of each DNA strand are required to avoid unnecessary circumstances during *in vitro* process which could lead to wrong computation results. Hence, it is vital to employ a highly unique DNA library in DNA computing [18].

Generally in DNA nanostructure design, the DNA sequences created using DNA bases are assembled into huge structures with well-defined criteria such as 2d-crystals, cubes, cages and some common nanomechanical devices. These structures are built in the form of multi-branched complex building blocks, which are designed from partial hybridizations among several DNA strings. The free ends of single stranded DNA also known as ‘sticky ends’, bind the building blocks to each other. Therefore, the well designed DNA strands enabled more efficient assembles for building huge structures that have specific criteria, like the 2d-crystal [39].

Several building blocks that are connected to each other via sticky ends are called ‘DNA tiles’. The specific tiles are assembled into larger structures through hybridizations among their respective sticky ends. The DNA tiles are used to encode information for self-assembly computations. The accuracy of the self-assembly computations relies highly on the DNA strands applied to design the tiles [40].