International Journal of Innovative
Computing, Information and Control
Volume **5**, Number **8**, August **2009**

ICIC International ©2009 ISSN 1349-4198
pp. **2249–2255**

DNA SEQUENCE SETS DESIGN BY PARTICLE SWARM OPTIMIZATION ALGORITHM

Qiang Zhang, Rui Zhang and Bin Wang

Liaoning Key Lab of Intelligent Information Processing
Dalian University
Dalian, 116622, P. R. China
zhangq@dlu.edu.cn

Received January 2008; revised July 2008

Abstract. *Designing DNA sequence sets is a fundamental issue in the fields of nanotechnology and nanocomputing. Not only quality but also quantity of DNA coding sequences affect the reliability of DNA computing. For this reason, many researchers have paid their attentions to find more and better DNA sequences in DNA computing. In this paper we present particle swarm optimization algorithm (PSO) for the design of DNA sequence sets, namely sets of equal-length sequence over the nucleotides alphabet A,C,G,T that satisfy H-distance constraint. In our computational experiments, we succeed in generating better sequences sets. We give some practical values which satisfy H-distance constraint, and it has some directions for the theoretical lower bound in DNA computing.*
Keywords: DNA sequence sets, Particle swarm optimization algorithm, H-distance

1. **Introduction.** The design of DNA sequence sets, or sets of short DNA strands that satisfy combinational constraints, is motivated by the tasks of storing information in DNA strands which are used for computation or as molecular bar-codes in chemical libraries [1,2]. The sequence design is an approach of the control, which aims to design DNA sequences satisfying some constraints to avoid such unexpected molecular reactions. Since expected or unexpected reactions depend on the applications or the purposes, usually several representative constraints are considered as below mentioned. Another requirement for DNA sequence sets is that the sets should be large. This is because designed DNA sequences are used as elemental components of computation; the size of sequences is considered the size of the computational resource [3,4]. So good DNA sequence sets design is important in order to minimize errors due to non-specific hybridization between distinct sequence and their complements, and also important to obtain a higher information density, and larger sets of sequences for large-scale application. In this paper, our purpose is to design large sets of sequences that satisfy H-distance constraint.

In the sequence design, many types of constraints are considered. Among them, combinatorial and thermodynamic constraints are well studied. Most common measures of combinatorial constraints are the Hamming Distance constraint (HD) and the Reverse Complement Hamming Distance constraint (RC). Also, the H-distance, which is stricter than the HD and RC, is well considered. Based on these constraints, many researchers have discussed the sequence set design. Arita applied techniques from the coding theory to design sequence sets [5]. Since Hamming distance measure has been discussed in the coding theory for a long time, in order to utilize the results, they proposed the template method that projected a set of sequences satisfying HD from the coding theory into a set of sequences satisfying all the three measures based on hamming distance. Tulpan proposed a Stochastic Local Search (SLS) method for the sequence sets design [6-8]. They