Chapter IX
Evolutionary Multi-Objective Optimization for DNA Sequence Design

Soo-Yong Shin
Seoul National University, Korea

In-Hee Lee
Seoul National University, Korea

Byoung-Tak Zhang
Seoul National University, Korea

ABSTRACT

Finding reliable and efficient DNA sequences is one of the most important tasks for successful DNA-related experiments such as DNA computing, DNA nano-assembly, DNA microarrays and polymerase chain reaction. Sequence design involves a number of heterogeneous and conflicting design criteria. Also, it is proven as a class of NP problems. These suggest that multi-objective evolutionary algorithms (MOEAs) are actually good candidates for DNA sequence optimization. In addition, the characteristics of MOEAs including simple addition/deletion of objectives and easy incorporation of various existing tools and human knowledge into the final decision process could increase the reliability of final DNA sequence set. In this chapter, we review multi-objective evolutionary approaches to DNA sequence design. In particular, we analyze the performance of ε-multi-objective evolutionary algorithms on three DNA sequence design problems and validate the results by showing superior performance to previous techniques.

INTRODUCTION

The genomic revolution including the Human Genome Project has been producing a huge amount of data in rapid succession, which are too large or complex to solve with traditional methods. Therefore, a variety of techniques have been applied to understand these data. Among them, evolutionary computation has been highlighted as one of the most successful techniques due to its rapid search ability in a very large and complex problem space and reasonable solution quality (Fogel & Corne, 2002).

Recently, researchers have found lots of biological problems naturally having more than one conflicting objective or constraint to satisfy. This multi-objective nature of biological problems and the success of evolutionary computation have encouraged the usage of multi-objective evolutionary algorithms for bioinformatics such as cancer classification (Deb & Reddy, 2003), protein structure prediction (Cutello, Narzisi, & Nicosia, 2006; Ray, Zydallis, & Lamont, 2002), DNA sequence/probe design (Lee, Wu, Shiue, & Liang, 2006; Rachlin, Ding, Cantor, & Kasif, 2005a; Shin, Lee, Kim, & Zhang, 2005a; Shin, Lee, & Zhang, 2006), gene regulatory networks inference (Spieth, Streichert, Speer, & Zell, 2005), peptide binding motifs discovery (Rajapakse, Schmidt, & Brusic, 2006), protein-ligand docking problems (Oduguwa, Tiwari, Fiorentino, & Roy, 2006), and medicine (Lahanas, 2004; Lahanas, Baltas, & Zamboglou, 2003).

Among the various biological problems, we review the multi-objective evolutionary optimization approaches to DNA sequence design in this chapter. The DNA sequence design problem is the most basic and important task for biological applications which require DNA sequences. Therefore, it involves many applications, including DNA microarray design, DNA computing, and DNA nanotechnology. Previous works have found that sequence design involves a number of heterogeneous and conflicting design criteria with many local optima and little gradient information (Shin, 2005). This supports that multi-objective evolutionary algorithms are actually a good approach for DNA sequence optimization.

Here we report on the evaluation results of MOEAs on the three representative DNA sequence design problems: orthogonal DNA sequence design for DNA computing, probe design for DNA microarrays, and primer design for multiplex polymerase chain reaction. We first formulate each problem as a multi-objective optimization task. Then, ε-multi-objective evolutionary algorithm is applied to the chosen problems. Finally, the performance of the evolutionary multi-objective approach is analyzed.

BACKGROUND

In this section, we will briefly introduce the background information for DNA sequence design. First, the abstract criteria for good DNA sequence and their mathematical definitions will be shown. Then, the three chosen target problems such as DNA computing sequence design, microarray probe design, and multiplex PCR primer design will be explained shortly. These applications belong to the most popular DNA sequence optimization problems and cover a wide variety of similar problems.

Criteria of Good DNA Sequence

In biology experiments which handle DNA sequences, the hybridization between a DNA sequence and its base-pairing complementary (also known as Watson-Crick complementary) sequence is the most important factor to retrieve and process the information stored in DNA sequences, since the rest of experimental steps are based on the perfect hybridization. For this reason, we desire a set of DNA sequences should form a stable duplex (double stranded DNA) with their complements. Also, we require crosshybridiza-