

University of Tehran

University College of Science

School of Mathematics, Statistics and Computer Science

Finding Maximal Pairs With Bounded Gap

by

Pooya Davoodi

Under Supervision of

Dr. Abbas Nowzari

And

Dr. Mahdi Sadeghi

A thesis submitted to the graduate studies office

in partial fulfillment of the requirements for the degree of M.Sc. in

Computer Science

February 2007

Abstract

DNA sequence analysis such as motif finding (discovery of conserved patterns) and finding repeated sequences, is one of the important problems in molecular biology because of its many practical applications like understanding the gene regulation. Experimental methods for solving this problem are very tedious and time consuming with high cost and also there are large amounts of genome sequence data and gene expression micro-array data, therefore it seems that the best way is to solve the problem computationally. Most of these types of problems are considered as NP-Complete problems in computer science and there are various techniques to attack them such as heuristic algorithms.

In this thesis, different models for representing motifs and some algorithms for finding motifs (WINNOWER, Gibbs Sampler and CONSENSUS) are studied and some methods for testing these algorithms are given. An algorithm for finding maximal pairs with bounded gap is discussed. All of these algorithms are considered and finally their benefits, faults and difficulties are compared.

Keywords: DNA, Motif, Signal, Promoter, Consensus Sequence, Profile, Relative Entropy, Suffix Tries, AVL Tries, Maximal Pairs.