Ain Shams University
Faculty of Science
Department of Mathematics

On Solving the Gene Silencing Problem via Algorithmic Techniques

A thesis submitted in partial fulfillment of the requirements of the M.Sc. degree in Computer Science
By

Soha Ibrahim Soliman Ibrahim

B.Sc. Statistics & Computer Science

Supervised by

Prof. Dr. Fayed Fayek Mohamed Ghaleb
Ass. Prof. Emeritus of Mathematics,
Department of Mathematics,
Faculty of Science,
Ain Shams University

Dr. Mohammad Hashim Aly Abdel Rahman
Lecturer of Computer Science,
Department of Mathematics,
Faculty of Science,
Ain Shams University

Submitted to:
Faculty of Science,
Ain Shams University,
Cairo - Egypt.
2018.

AIN SHAMS UNIVERSITY

Author:	Soha Ibrahim Soliman Ibrahim
Title :	On Solving the Gene Silencing problem via Algorithmic Techniques
Division:	Computer Science
Department:	Department of Mathematics
Faculty:	Faculty of Science
Degree:	M.Sc.
Year:	2018

Contents

Acknowledgements

First of all, cordial thanks to **ALLAH**, from start to end, for enabling me to do and finish this study successfully.

I would like to express my deepest thanks and gratitude to ***Prof. Dr. Fayed Fayek M. Ghaleb***, *Prof. of Mathematics, Department of Mathematics, Faculty of Science, Ain Shams University*, for his kind supervision, valuable advice, continuous encouragement, support, and for his great efforts in reviewing the present manuscript.

Great appreciations and thanks are due to ***Dr. Mohammad H. Abdel Rahman***, *Lecturer of Computer Science, Department of Mathematics, Faculty of Science, Ain Shams University,* for suggesting and planning the present point of research, and for his valuable suggestions and fruitful advice throughout the present study, for his kind supervision, guidance and participation in reviewing the current manuscript.

Sincere thanks are due to the staff members and my colleagues in the *Department of Mathematics, Faculty of Science, Ain Shams University*, for their cooperation and encouragement during the present study.

All my love and respect are due to my dear parents for their kind support, encouragement, and always praying for me; and to my beloved husband and daughter for the patience, encouragement and missing me most of the time.

Abstract

This thesis introduces a new algorithm for solving one of the hottest problems in biology and medicine which is the gene silencing problem.

The new algorithm is called "Exogenous Gene Silencing using Hashing " (EGSH). This algorithm is specially designed to solve the exogenous silencing of a specific target gene, taking into consideration the possibility of both exact and partial matching between the target gene and small interfering RNA that produced by the new algorithm.

The theoretical analysis of the running time and memory complexity of the EGSH algorithm confirms that the EGSH algorithm achieves a remarkable speeding up for the running time and reducing of the memory space required comparable with other previously introduced algorithms.

Also, the experimental result obtained from implementing the EGSH algorithm (using Java Programming language) emphasizes the theoretical analysis results. It shows that for any specific Human gene, the program takes, in average, about three minutes running time consuming less one Gigabyte memory.

Summary

Some genetic disorder diseases are caused by mutant genes that synthesize harmful proteins causing the initiation and progression of such disorders.

One of the ways to avoid these diseases is to turn off the mutant genes to prevent them from producing the harmful proteins which is known as the Gene Silencing.

The Exogenous gene silencing means designing short interfering RNA sequences in laboratories called siRNA then injecting them into the cell to target a particular messenger RNA(mRNA) of the mutant gene and cause its degradation. The big challenge in this approach is that matching between the designed siRNA and target mRNA do not need to be a perfect matching.

In this thesis, a new algorithm Exogenous Gene Silencing using Hashing (**EGSH**) is introduced to solve the exogenous gene silencing problem. Although few previous algorithms were introduced to solve this problem but they focused on the totally matching between siRNA and target mRNA.

The new feature of the proposed algorithm is it solves both the total and the partial matching problem. The proposed algorithm uses the hashing technique that takes linear execution time and it also requires relatively small memory space compared to the previous algorithms.

This thesis is organized into six chapters:
Chapter 1: Explains the basic concepts of bioinformatics and why scientists use it, and gives a brief summary of algorithmic techniques, gene silencing, and the objective of the study.

Chapter 2: Is divided into two parts. The first part shows the importance of the computational tools like algorithmic techniques in solving some biological serious problems. The second part talks about the hash technique, it's definition,methodology, strategies, hash functions, it's characteristics and the advantage, and disadvantage of using the hash technique.

Chapter 3: Composed of two main parts. The first part illustrates some important biological terms that help to understand the problem of the thesis like Organism, Cell, Genome, Deoxyribonucleic Acid (DNA), Nucleotides, Gene, Ribonucleic Acid (RNA) and Proteins. The second part illustrates all the details of the gene silencing problem such as it's definition, the need for the gene silencing, mechanism of gene silencing, different types of gene silencing, finally the definition of RNA interference (RNAi) and how it works.

Chapter 4: Explains firstly the benefits of controlling a particular gene, this control can be done by one of two ways endogenous or exogenous gene silencing control. Secondly present some of the previous work that tries to solve the problem with it is advantages and disadvantages.

Chapter 5: Introduces in detail a new algorithm called EGSH for the exogenous gene silencing problem of a particular gene, also illustrates it is analysis and the experimental results of the new EGSH algorithm.

Chapter 6: The conclusion and future work; summarizes the results achieved in the thesis including the new algorithm advantages and limitation, and provide conclusion and future work in this field.

Publication

Soliman, S.I.; Ghaleb, F.F.M. and Abdel-Rahman, M.H. (2017): A Hashing Algorithm Improving the Exogenous Gene Silencing Problem. First International Conference of Faculty of Science, Ain Shams University, Hurghada, EGYPT.