Thesis Report

Prediction of nitrosocysteine sites using position and composition variant features

Submitted To

School Of System and Technology
in Partial Fulfillment of the Requirements
for the Degree of
MASTERS OF COMPUTER SCIENCE

Submitted By:

Aroosa Batool 15025050007
Session 2015-2017

Supervised by:
Dr. Yaser Daanial Khan

Co-Supervisor:
Dr. Nouman Rasool

FINAL APPROVAL

It is certified that the research work presented in this thesis entitled "Prediction of Nitrosocysteine Sites Using Position And Composition Variant Features" was conducted AroosaBatoolunder the supervision of Dr. YaserDaanial Khan at the University of Management and Technology, Lahore, Pakistan in April 2017 to fulfill the requirement of the degree of the MS (CS) Computer Science.

1. Supervisor

Dr. Yaser Daanial Khan,
HEC Approved Supervisor, _____
Director BS-CS program,
School of Science and Technology,
University of Management and Technology, Lahore
.

2. Director Graduate Studies

Dr. Shoaib Farooq
HEC Approved PhD Supervisor _____
Associate Professor,
School of Science and Technology,
University of Management and Technology, Lahore

3. Head of Department

Dr. Adnan Abid
Associate Professor _____
HEC Approved PhD Supervisor
Chairperson Department of Computer Science,
School of Systems and Technology,
University of Management and Technology, Lahore

4. Dean SST

Dr. Shaukat Iqbal,
Professor, _____
Chairperson Department of Informatics & Systems,
School of Systems and Technology,
University of Management and Technology, Lahore

DECLARATION

I ***Aroosa Batool*** ID #***15025050007***Session***2015-2017***hereby certify that this thesis is being submitted in partial fulfillment of the requirements for the ***MS***degree in***Computer Science***. This thesis is my original work, and the data/material presented herein has not been used for the acquisition of any other degree from any institution.

Signature: _____

Date: _____

AROOSA BATOOL

ACKNOWLEDGMENT

Alhamdulillah. First of all, all my gratitude is to Allah SWT, with His willing that gave me the opportunity to complete this thesis which is entitled as "Prediction Of Nitrosocysteine Sites Using Position And Composition Variant Features". I extend my heartfelt thanks to all the people who directly or indirectly helped me in the completion of my thesis.

From the beginning to the end, my project supervisor Dr. Yaser Daanial khan has shown immense patience and support, besides providing an incredible amount of guidance. Thank you, sir. I was highly encouraged by Dr. Nouman Rasool who was abundantly helpful and offered, support, invaluable assistance and guidance. There have been many more people involved in this venture, whose names may not have been mentioned above, but their help is undoubtedly acknowledged. I would like to thank all my friends especially those who helped me to complete my thesis work and for the wise idea throughout the project.

TABLE OF CONTENTS

`

LIST OF TABLE:

LISTOF FIGURES:

PREFACE

This thesis was prepared as a partial fulfillment of the requirements for acquiring the degree Master of Science in Computer Science, MS (CS), in the School of Systems and Technology, SST, located at the University of Management and Technology, UMT.

The thesis introduces a novel method based on sequence-derived features and effective feature selection techniques to identify S- nitrosocysteine residue in a polypeptide chain. This identification of SNO sites provides insights into drug discovery and disease progression as in this post-genomic age , we are facing the avalanche of biological sequences generation. So, it is important for both basic research and drug development to timely identify the PTM sites in proteins.

It is expect that the reader has a basic knowledge in the areas of chemistry and computer programming languages.

ABSTRACT

Protein S-nitrosylation, a significant posttranslational modification of protein, involves the addition of nitrogen oxide group to cysteine thiols to form S-nitrosocysteine. Growing evidence has suggested that S-nitrosylation plays a major role in numerous human diseases. So, it is highly anticipated for the intuition into biological research and drug discovery to develop such techniques for timely identification of S-nitrosylated proteins. The proposed system endeavors a novel strategy based on numerous intellectual computational method for the identification of S-nitrosocystiene site from a protein sequence. Statistical moments were used to extract the features and built a multilayer neural network model using Gradient Descending and Adaptive Learning Algorithm. The comparison results on the-state-of-the-art benchmark datasets have shown that this proposed scheme is very propitious, accurate and exceptionally effective for the prediction of S-nitrosocystiene in protein sequence.

Chapter 1

Introduction

CHAPTER 1

1. Introduction:

Proteins are long polymers composed of twenty different amino acids [1]. During protein translation process, all amino acids are coded by triplet codons, connected side by side with each other in the form of a polypeptide chain through peptide linkage [2]. After synthesis of a protein by ribosome, it gets enter into endoplasmic reticulum for further processing known as post translational modification (PTM). In PTM, various modifications occur in proteins altering the structure and biological activity of these macromolecules. One of these modifications is nitrosylation at tyrosine or cysteine residues. Nearly one thousands of proteins has been identified in various biological methods linked with S-nitrosylation [3].The cysteine residues within proteins underlie a variety of modifications because of the presence of nucleophilic thiol group. This thiol group is oxidized by nitric oxide with the help of specialized enzymes present in endoplasmic reticulum. The process of making of S-nitrosocysteine is done by the addition of nitroso group (-N=O) into the reduced cysteine chemically defined as S-nitrosylation [4].

S-nitrosylation is considered as one of the most important PTM which modulate biological activities and stability of proteins, cellular trafficking, apoptosis, circulation and muscle contractility [5,6, 7, 8].

The accurate determination of S-nitrosylation at cysteine residues is important for monitoring proper functioning of protein in a cell and also or development of drugs for fatal diseases[9,10,11].