Prediction of Nitrotyrosine Sites Based On Composition and Position Based Features

By

Ahmad Waseem (15026050008)

A Dissertation Submitted For the Degree of Master in Computer Science in the Department of Computer Science, School of Systems and Technology, University of Management and Technology.

May, 2018

Supervised by:

Dr. Yaseer Daniyaal Khan

FINAL APPROVAL

It is certified that the research work presented in this thesis entitled "Prediction of Nitrotyrosine Sites Based on Composition and Position Based Features" was conducted by Ahmad Waseem Ghauri under the supervision of Dr. Yaser Daanial Khan at University of Management and Technology, Lahore, Pakistan in May 2018 to fulfil the requirements of the degree in M.S/M.Phil in Computer Science.

- **Supervisor**

  Dr. Yaser Daanial Khan

  Associate Professor, Chairperson Department of Computer Science, School of Systems and Technology, Department of Computer Science,

  University of Management and Technology _____

- Director Graduate Studies

  Dr. Shoaib Farooq

  Associate Professor, Director Graduate Studies, School of Systems and Technology, Department of Computer Science,

  University of Management and Technology _____

DECLARATION

I hereby **Ahmad Waseem Ghauri** with Student ID **15026050008** declare in this declaration that it's my own original work. Where collaborations with other researchers are involved, or materials generated by other researchers are included, the parties and/or materials are acknowledged or are explicatory referenced as appropriate.

This work is being submitted to the University of Management and Technology in partial fulfilment of the requirements of the degree of Master of Science (M.S) in Computer Science.

This thesis has not been submitted to any other University or institution for any degree or examination.

Date	Student Signature
Date	Supervisor Signature

ABSTRACT

Closely related to causes of various diseases such as rheumatoid arthritis, septic shock, and coeliac disease; tyrosine nitration is considered as one of the most important post-translational modification in proteins. Inside a cell, such modifications occur accurately by the action of sophisticated cellular machinery. This task is accomplished by specific enzymes present in endoplasmic reticulum. The identification of potential tyrosine residues in a protein primary sequence which can be nitrated is a challenging task. To counter the prevailing, laborious and time-consuming experimental approaches, here we introduce a novel computational model. Based on experimentally verified tyrosine nitration sites, they are transformed to their feature vectors. An adaptive training algorithm is then used to train a back propagation neural network for prediction purposes. To objectively measure the accuracy of the proposed model, rigorous verification and validation tests are carried out which led to a promising accuracy of 88%, a sensitivity of 85% and a specificity of 89.18% and Mathew correlation coefficient of 0.627. We believe that this computational model may provide foundation for further investigation and can be used deal with the other PTM sites in proteins.

ACKNOWLEDGEMENTS

Firstly, I'd like to thank my supervisor, Dr. Yaser Daanial Khan for his for his patience, motivation, enthusiasm, and on-going support throughout my M.S study and research.

I'd would also like to thank Dr. Nauman Rasool for his encouragement and insightful comments.

Last but not the least, my parents deserve credit for always being there for me and providing their wise counsel.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

1.1 Problem Definition

Given the post genomic age we currently live in, the protein sequences generated in this era has been exponentially increased [1]. Nitrotyrosine is considered a marker for cell damage and inflammation and is linked to wide range of human pathological conditions [2] and is detected in a large number of diseases, such as lung cancer, cardiovascular disease, asthma, Alzheimer's disease, rheumatoid arthritis, septic shock and coeliac disease [3]. The determination of nitrotyrosine residues in a protein is not an easy and inexpensive task, but time consuming involving advanced technologies. The lack of experimentally verified susceptible sites has made it difficult to examine global biophysical and evolutionary trends of nitrotyrosine in a protein [4].

1.2 Research Objectives

To counter the prevailing, laborious and time-consuming experimental approaches for the detection of tyrosine nitration sites, our research objective is to introduce a novel computational model which is based on experimentally verified tyrosine nitration sites. Our aim is to predict nitrotyrosine sites which can provide us insights of its impact at the proteome level. In order to predict tyrosine residues which are prone to nitrosylation, many computational models have been proposed with varying sensitivies and accuracies. The prediction of nitrotyrosine sites in proteins is under active research by the community and considered basis for detection of pathological conditions and drug development [5].