UNIVERSITI TEKNOLOGI MARA

PREDICTION OF NOVEL DOPING AGENT THROUGH THE INTEGRATION OF CHEMICAL AND BIOLOGICAL DATA USING *IN SILICO* METHOD

NURUL AIN BINTI MOHD ROSMAN

Dissertation submitted in partial fulfilment of the requirements for the
BACHELOR OF PHARMACY (Hons.)

Faculty of Pharmacy

JULY 2016

Acknowledgement

First of all, I would like to thank God for giving me the opportunity to learn more on this course and giving me the strength to complete my research project. I would also like to express my appreciation to Dr Fazlin binti Mohd Fauzi for her guidance and patience in helping me improve my knowledge. Without her valuable assistance, this work would not have been completed. I would also like to thank my parents for everything and always praying for my success. Thank you also to my group members, Nurul Amalina and Muhammad Artif for helping me out and encouraged me to complete this project and sharing information to improve my understanding. Lastly, I would like to thank my family, and my close friends for their advice, moral support when I feel like giving up and encouraging me to do better so that I can be a good student. Without their help, I don't think I can get as far as I am today.

TABLE OF CONTENT

CHAPTER ONE (INTRODUCTION)

CHAPTER TWO (LITERATURE REVIEW)

CHAPTER THREE (MATERIALS AND METHODS)

ABSTRACT

The identification of novel doping agent is not quick enough to prevent the misused of the potential doping agent among athletes. Existing techniques e.g. HPLC are costly and time consuming. Hence, other alternative methods are needed to handle this problem and one such method is *in silico* method. In this study, an *in silico* method that integrates chemical and biological data was used to predict potential doping agents. The *in silico* method, also known as *in silico* target prediction, first analyse patterns of protein-ligand binding from chemical and biological data through the use of machine learning algorithm. The deductive model built can then be used to predict potential protein targets for a novel compound, given its chemical structure. The models built were trained on compounds binding to protein targets known to produce doping effects e.g. androgen receptor which increases muscle mass obtained from ChEMBL database. This study employed three molecular descriptors (MACCS Keys, ECFP_4 and FCFP_4) and two machine learning algorithms (Decision Tree and Naive Bayes Classifier) to build the predictive model. Two validations were performed on the models which are internal and external validation. Sensitivity was used as a performance measure. The performance of the models were also analysed at a certain cut off scores and ranks. The internal validation showed that the combination of FCFP_4 and Decision Tree model performed best with a sensitivity value of 0.94 , when a cut off of rank 5 was applied. This model then proceeded to the external where compounds from WADA Monitoring List were subjected to testing. It was found that mitragynine and nicotine were predicted to bind to κ-opioid receptor which is supported by scientific literature. This proved that the model was able to predict the protein target for compound outside of the training set. Hence, *in silico* method can be used to predict potential doping agents in a timely and efficient manner and consequently prevent the widespread of misused of doping agent.

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

The identification of novel doping agent is not quick enough to prevent the misuse of the potential doping agent.

1.2 Hypothesis of the Study

Novel doping agents can be predicted through the use of *in silico* methods that incorporate chemical and biological data.

1.3 Objective of the Study

To build an *in silico* target prediction that can analyse the patterns of protein-ligand binding of doping agents and consequently predict novel doping agents based on its chemical structure.

1.4 Significant and Limitation of Study

Significant of this study is the novel doping agent can be predicted before the compound is misuse and abuse by the user. Limitation of this study is that the prediction of novel doping is based on the training set used. Hence, the accuracy of prediction beyond the training space cannot be guaranteed.