UNIVERSITI TEKNOLOGI MARA

IN SILICO IDENTIFICATION OF *MYCOBACTERIUM TUBERCULOSIS* RV2969C MEMBRANE PROTEIN: AN APPROACH TOWARDS VACCINOLOGY

NURUL SYAHIRAH BINTI MOHAMAD SARIB

Dissertation submitted in partial fulfilment of the requirement of the Bachelor of Pharmacy (Hons.)

Faculty of Pharmacy

October 2008

ACKNOWLEDGEMENTS

Greatest thanks to the Almighty God; I have finally finished the research project entitled "*In silico* Identification of *Mycobacterium tuberculosis* Rv2969c Membrane Protein: An approach towards vaccinology."

In preparation of this research project, I have benefited from the guidance and advice from my supervisor, Mr. Leow Chiuan Yee and co-supervisor, Mr. Leow Chiuan Herng. Not to forget, Mrs. Samsiah Kandar who helps in reviewing the research projects. I am gratefully acknowledging their contributors in providing criticism at various stage of the research.

I would like to acknowledge my family and friends who have provided invaluable support throughout the writing process.

Finally, I have an enduring appreciation for the Dean of Faculty of Pharmacy, Prof. Dr. Abu Bakar bin Abdul Majeed, the co-ordinator of this course (Research and Instrumentation), Dr. Zainol and all lecturers whose suggestions provide insight.

TABLE OF CONTENTS

ABSTRACT

Mycobacterium tuberculosis (TB), the major causative agent of tuberculosis, is responsible for the death of 2 million annually and the infection of other 9 million people around the world. BCG vaccine which was introduced in the early 90's, initially proved to be successful in reducing mortality from tuberculosis in about 90% of vaccinated children. However, the vaccine had been found to have little effect on pulmonary tuberculosis. In this study, the bioinformatics tools were used to predict the sequence that is most likely to be the vaccine candidates. The gene sequence of the Mycobacterium tuberculosis was retrieved from NCBI website in FASTA sequence. By using the sequence, the amino acids composition was predicted to have alanine (14.1%) and followed closely by valine (11.8%). The GRAVY of this sequence is 0.097. The motif of the sequence cannot be predicted by using ScanProsite tool, therefore FingerPRINTScan tool was used. As a result, the Poaallergen seem to have almost similar sequence. The PSORTb program to predict bacterial protein subcellular localization prediction was used. The subcellular protein was found to localize at the cytoplasmic region with final score of 8.87. The The PSIPRED protein structure prediction server was used to predict the protein secondary structure. The most abundant structure found is the loop. The antigen epitopes of the protein sequence was predicted by using ProPred Web tool. The sequence that was found abundantly is VVFAVVLVF. Therefore, the sequence of VVFAVVLVF is most likely suitable to be the promiscuous binder of vaccine candidates.

CHAPTER 1

INTRODUCTION

1.0 Background of study

Mycobacterium tuberculosis (TB), the major causative agent of tuberculosis, is responsible for the death of 2 million annually and the infection of other 9 million people around the world. Years ago, it was reported that the number of TB cases in Malaysia for the years 2000, 2001 and 2002 are 15,643, 14, 820 and 14,389 respectively with the mean mortality rate of 6.2/100,000 population (Aziah, 1998). Meanwhile, TB is a well recognised occupational hazard for healthcare workers (HCWs) Recently, a report of 25 HCWs working in 11 general hospitals in Malaysia were infected with TB in 2004 (Tan and Kamarulzaman, 2005). This disease's prevalence has increased over the past few years due to co-infection with the Human Immunodeficiency Virus (HIV) and the appearance of multi-drug and extensive-drug resistant strains; which, added to the financial and practical limitations of directly observed therapy strategies (DOTS) (Kaufmann, 2000), has increased the need for a much deeper understanding of this microorganism's biology for developing new cost-effective therapies and vaccines.

Recently, vaccine development has taken advantage of the genome sequence of pathogenic bacteria (Rappuoli and Covacci, 2003). With the advent of whole-genome sequencing and advances in bioinformatics, this approach can now mine the