UNIVERSITI TEKNOLOGI MARA

BIOINFORMATICS ANALYSIS OF HAPTOGLOBIN PROTEIN: AN APPROACH TOWARDS COLON CANCER BIOMARKERS IDENTIFICATION

NUR NADIA SURAYA BT SUDDIN

Dissertation submitted in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy (Hons.)

Faculty of Pharmacy

November 2009

ACKNOWLEDGEMENT

Thanks to Almighty Allah S.W.T for giving me patience and strength to complete this study promptly. I wish to express my sincere appreciation and grateful to my supervisor Mr Leow Chiuan Herng and my co-supervisor Mr Leow Chiuan Yee for their guidance, instruction, suggestion as well as information and time start from early stage of the project till the project finished successfully. Besides, I wish the deepest thankful to Dr Kalavathy Ramasamy as coordinator for this course. Last but not least, thanks to all my beloved family and my colleagues who always encourage and support me in order to complete this project efficiently and smoothly.

TABLE OF CONTENTS

ABSTARCT

Haptoglobin (Hp) is a hemoglobin-binding plasma protein consisting of two types of chains, called alpha (α) and beta (β), which originate from a common polypeptide. In humans, but not in other mammals, Hp has been shown to occur in two allelic forms, Hp1 and Hp2, which differ in the length of the a-chain. The longer a-chain (in Hp2) seems to have arisen by an internal duplication of a gene segment coding for almost the entire a-chain of Hp1.β chains of hp is more heavier than its α chains. The a-chain contains a complement control protein (CCP) domain and the b-chain contains a serine proteinase (SP) related domain. Hp plays a major role in binding to hemoglobin (Hb) and thereby inhibits its oxidative activity .It also acts as immunomodulatory and chaperone-like properties. It was reported that the Hp protein level increased during carcinogenesis and has been revealed to support angiogenesis. In this study, the identification of haptoglobin protein as a colon cancer biomarker was reported using *bioinformatics analysis* prediction approaches. Hp protein consist of 340 amino acids with the predicted value of both Mw and pI are 38415.7 Da and 6.26 respectively. Hp protein is a hydrophilic protein due to its GRAVY value shows the negative value,-0.312.It is considered stable in nature. Its aliphatic index is 83.98. From the analyzed data, this protein consists of 35% of α-helix, 28% of β-sheet, 24.5% of β-turns and 17.2% of coils loops. Hp protein is mainly found in an extracellular cell compared to intracellular. There are two domains found in the hp protein which are Sushi and Trypsin_Dom. Sushi domains are known to be involved in many recognition processes, including the binding of several complement factors to fragments C3b and C4b whereas Trypsin_Dom (trypsin family which is known as serine protease) usually involves in catalytic activity .From the multiple sequence alignment, it illustrated that human,orang utan, rabbit ,mouse, cattle and monkey share high sequence similarity. However, according to the phylogenetic tree, orang utan haptoglobin is the most similar towards human haptoglobin. As a conclusion, the information obtained from bioinformatics analysis of haptoglobin protein would then provide preliminary information for identification biomarker for colon cancer study.

CHAPTER 1

INTRODUCTION

1.1 Statement of problem

Colon cancer is a cancer that forms in the tissues of the colon (the longest part of the large intestine) (National Cancer Institute). There are two kinds of growths that occur in the colon: noncancerous growths, such as polyps and malignant or cancerous growths. Colon cancer usually begins with the growth of benign growths such as polyps. Most types of colorectal cancer are adenocarcinomas. This means that the cancer cells are formed from abnormal gland cells that line the inner surface of an organ. The prefix "adeno" means "gland." In colorectal cancer, the abnormal growth begins to form in the inner lining of the large bowel.

Colorectal cancer is the third most diagnosed cancer in the United States and Canada (after lung and breast in women, and lung and prostate in men). It is the third leading cause of cancer death in both men and women in the US. The American Cancer Society estimates that about 148,810 new cases and 49,960 deaths expected in 2008. (Colon Cancer Alliance).In Italy, the incidence of this disease can be estimated around 49/100 000.The sporadic forms are largely prevalent, while only 5to 10% is genetically inherited (Fearon *et al.*, 1990). Peterson *et al.* (1999) reported that among the hereditary forms, two syndromes accounting for about 3 to 6 % of