GROUP-BASED QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP (G-QSAR) AND MOLECULAR DOCKING STUDIES OF BCL-XL, BCL-2 AND PI3K-Γ INHIBITORS FOR LUNG CANCER TREATMENT

BY

NADIA HANIS BINTI ABDUL SAMAT

A thesis submitted in fulfillment of the requirement for the degree of Master of Science
(Pharmaceutical Technology)

Kulliyyah of Pharmacy
International Islamic University Malaysia

JULY 2014

ABSTRACT

Lung cancer is one of the common cancers in Malaysia and there are many different types of drugs used in the treatment of lung cancer. Bcl-xL, Bcl-2 and PI3K-γ were found to play important role in lung cancer cell proliferation and development. Bcl-xL and Bcl-2 allow uncontrolled cell proliferation by inhibiting apoptosis while derangement of PI3K-γ pathway has been associated with uncontrolled cellular proliferation and survival. Thus these proteins become good targets for the development of targeted therapy in lung cancer. Various classes of compounds have been used successfully to targets Bcl-xL, Bcl-2 and PI3K-γ and in order to improve the biological activity of these compounds various QSAR methods have been employed. In this study, an advanced method in computational drug design, Group-based Quantity Structural Activity Relationship (G-QSAR) was performed using V-LIFE® on dataset of non-congeneric compounds with binding activity available in Binding Database website from various literatures and generated potential protein inhibitors. Combination of variable selection methods (simulated annealing, stepwise forward or stepwise forward backward) with model building method (multiple linear regression, K-nearest neighbour method or partial least square) has come out with three statistical significant models for each protein target which were used to generate new inhibitor candidates by computational method. G-QSAR analysis of Bcl-xL inhibitors has resulted in model SA-MLR (r^2=0.78, q^2=0.68), STP-MLR (r^2=0.80, q^2=0.70) and SA-kNN (q^2=0.75) which have been used to generate total number of 2520, 3600 and 5850 compounds respectively. Three best models obtained from Bcl-2 inhibitors G-QSAR analysis are STP-MLR (r^2=0.82, q^2=0.71), SA-MLR (r^2=0.83, q^2=0.75) and SA-kNN (q^2=0.65), which generated 4964, 1536 and 7576 compounds respectively. While G-QSAR analysis of PI3K-γ inhibitors come out with STP-MLR (r^2=0.72, q^2=0.67), SA-MLR (r^2=0.70, q^2=0.65) and STP-PLS (r^2=0.68, q^2=0.62), which generated 2040, 2244 and 2560 compounds respectively. Newly generated compounds were then analysed and validated by molecular docking into the available Bcl-xL, Bcl-2 and PI3K-γ crystal structures from Protein Databank using Glide® software. Six series of docking (docking of three datasets of G-QSAR models-based generated compounds and three docking of known target protein inhibitors) were performed for each protein target. High throughput virtual screening (HTVS) docking was conducted to filter large compound dataset, and this was followed by extra precision (XP) docking which is more detail and flexible than HTVS. Overall, XP docking results of Bcl-xL inhibitors showed common interaction with ASN136, ARG139, GLY138 and TYR101. Most of the Bcl-2 inhibitors interacted with Bcl-2 crystal structure at amino acid residues LEU96, ALA108, PHE112, PHE71 and TYR67. Lastly, majority of the PI3K-γ docked inhibitors were shown to interact with four common amino acids including TYR867, LYS833, VAL882 and TRP812. Non-covalent interactions such as hydrophobic, π-π interaction and hydrogen bond are the common types of interaction found in docking results of Bcl-xL, Bcl-2 and PI3K-γ inhibitors. Based on the docking results, the newly generated compounds interacted very well with target proteins as good as the known inhibitors in clinical trials.

خلاصة البحث

يعتبر سرطان الرئة واحداً من أكثر أنواع السرطانات شيوعاً في ماليزيا ويوجد العديد من الأدوية التي تستخدم لعلاج سرطان الرئة. وقد وجد أن Bcl-xl، Bcl-2 و PI3K-γ تلعب دوراً هاماً في تكاثر وتطور خلايا سرطان الرئة. يسبب Bcl-xl و Bcl-2 تكاثراً غير مضبوطاً للخلايا عن طريق تثبيط الموت المبرمج للخلية بينما لوحظ اضطراباً في نشاط **PI3K-γ** مترافقاً مع نكاثر وبقاء خلوي غير مضبوط. ولهذا تعتبر هذه البروتينات من الأهداف الهامة للعلاج الموجه لسرطان الرئة. في هذا البحث، تم استخدام طريقة متطورة من طرق تصميم الدواء عن طريق الكمبيوتر وهي علاقة البنية بالتأثير المعتمدة على الزمر الوظيفية باستخدام برنامج **V-life**حيث تم استخدام مجموعة غير متجانسة من المركبات ذات فعالية مثبطة للأنزيمات المدروسة. عدة طرق إحصائية وخوارزميات مختلفة تم استخدامها لانتقاء المتغيرات (محاكاة الصهر، الانتقاء المتدرج التقدمي أو التراجعي) وبناء النموذج نتج عنها ثلاثة نماذج مهمة إحصائياً لعلاقة البنية بالتأثير لكل أنزيم. النماذج المهمة لـــ Bcl-xl هي MLR-SA (r^2=0.78 ، q^2=0.68)، STP-MLR (r^2=0.80 ، q^2=0.70) و SA-kNN (q^2=0.75) وقد استخدمت هذه التماذج لبناء جزيئات جديدة ذات فعالية مثبطة محتملة. فيما يتعلق بعلاقة البنية بالتأثير ل Bcl-2 تم بناء ثلاث نماذج جزيئية وهي STP-MLR (r^2=0.72 ، q^2=0.67)، SA-MLR (r^2=0.70 ، q^2=0.65) و STP-PLS (r^2=0.68 ، q^2=0.62) وقد استخدمت هذه التماذج لتوليد 4969، 1536 و 7576 جزئية ذات فعالية محتلمة مثبطة. وبالمثل لـــ PI3K-γ فقد تم بناء ثلاث نماذج جزيئية مهمة وتم توليد 2040، 2244 و 2560 جزئية جديدة. وقد تم تقييم ودراسة الجزيئات المتولدة من النماذج التي تم بناؤها باستخدام طريقة التفاعلات الجزيئية مع البنى البلورية المتوافرة للأنزيمات الثلاث من قاعدة معلومات البروتينات باستخدام برنامج Glide. تم إنجاز ثلاث سلاسل من التفاعلات الجزيئية (ثلاث منها للمركبات المتولدة عن النماذج التي تم بناؤها وثلاث لمثبطات معروفة مسبقاً). تم استخدام تقنية المسح الافتراضي عالي الإنجاز (HTVS) لتقليل عدد المركبات الضخم ومن ثم استخدمت طريقة الدقة العالية (XP) والتي تعتبر أكثر مرونة ونفصيلاً من السابقة. أبدت التفاعلات الجزيئية لأنزيم Bcl-xl بطريقة XP ارتباطاً مع حمض الأسبارجين 136، الارجنين139 ، الغليسين 138 و التيروسين101 . في حين أن معظم مثبطات Bcl-2 أظهرت إرتباطاً مع الليوسين 96، آلانين 108، فنيل آلانين112 و تيروسين67. وأخيراً فإن أغلب مضادات PI3K-γ تفاعلت مع التيروسين 867، ليسين 833 ، فالين 882 و تربتوفان 812. أم التفاعلات التي تم ملاحظتها في التفاعلات غير التساهمية كالروابط الهدروجينية والتفاعلات الكارهة للماء. وبناءً على ذلك فإن المركبات التي تم توليدها من النماذج السابقة قد أرتبطت بالأنزيمات المدروسة بشكل مماثل لمركبات أخرى قيد الدراسات السريرية.

APPROVAL PAGE

I certify that I have supervised and read this study and that in my opinion, it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a thesis for the degree of Master of Science (Pharmaceutical Technology)

.....
Farahidah Mohamed
Supervisor

.....
Abubakar Danjuma Abdullahi
Co-Supervisor

I certify that I have read this study and that in my opinion, it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a thesis for the degree of Master of Science (Pharmaceutical Technology)

.....
Noraslinda Muhamad Bunnori
Internal Examiner

.....
Sharifuddin bin Md Zain
External Examiner

This thesis was submitted to the Department of Pharmaceutical Technology and is accepted as a fulfilment of the requirement for the degree of Master of Science (Pharmaceutical Technology)

.....
Juliana Md Jafri
Head, Dept of Pharmaceutical Technology

This thesis was submitted to the Kulliyah of Pharmacy and is accepted as a fulfilment of the requirement for the degree of Master of Science (Pharmaceutical Technology)

.....
Siti Hadijah Shamsudin
Dean, Kulliyyah of Pharmacy

DECLARATION

I hereby declare that this dissertation is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Nadia Hanis binti Abdul Samat

Signature …… Date ……..

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

GROUP-BASED QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP (G-QSAR) AND MOLECULAR DOCKING STUDIES OF BCL-XL, BCL-2 AND PI3K-Γ INHIBITORS FOR LUNG CANCER TREATMENT

Affirmed by Nadia Hanis binti Abdul Samat

…….. ……

Signature Date

ACKNOWLEDGEMENTS

In the name of Allah, The Most Beneficent, The Most Merciful. Alhamdulillah, with His blessings, finally, I manage to complete my thesis after the hard work, blood, sweat and tears spent during the whole study period.

First and foremost, I would like to express my deep and sincere gratitude to my parents, my sweethearts, Hj Abdul Samat and Hjh Salihah, as the main sponsor for this study, spiritual and financial. Without their sacrifice, I will not be able to complete this study, even to register as master student at this age. Secondly, I would like to present my humble appreciation to my supervisors, Asst. Prof Dr. Farahidah Mohamed and Dr Abubakar Danjuma Abdullahi for their great effort to guide me and sharing their wonderful knowledge trough out the research.

Also, I would like to thank my colleague, who always share a bundle of information, guide me in chemistry as well as motivate me to the end of this work. May Allah bless you and help you to finish your work as soon as possible.

I wish to thank my wonderful friends, Nur'izzati, Noorsyafawati, Nurul Hafizah, Sayyada, Aya and all my postgraduate friends who always be my bread and butter, be together with me with laughter and tears. Allah's bless will always with you guys and in sha allah you will success here and hereafter. Last, but not least, thank you to my wonderful siblings; Najwa Hanis, Nabila Hanis and Nabihah Hanis, not forgotten, to in laws, my cute niece and nephew, for their encouragement and moral support during the production of this dissertation.

Thank you Allah for Your wonderful gift,

Wassalam

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

α	Alpha
β	Beta
γ	Gamma
δ	Delta
π	Pi
nM	Nanomolar
μM	Micromolar
>	Larger/more than
<	Smaller/less than
≥	Larger or equal than
≤	Smaller or equal than

LIST OF EQUATIONS

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND OF THE STUDY

Lung cancer has been recorded as the most common cancer in the world for many years. It is the most common cancer in males worldwide with a very high rate especially in Eastern Asia, Southern Europe, Northern America and Central Eastern while it is the fourth most common cancer in females, worldwide (Jemal et al., 2011). Lung cancer was reported as the second common disease among males in Malaysia. According to National Cancer Registry Malaysia (NCR) there were 2,048 cases of lung cancer cases registered with NCR in 2006, comprising 1,445 males and 603 females (Zainal, Zainudin, & Nor Saleha, 2006). In fact, lung cancer is responsible for 19.8% of all death due to cancers (Liam et al., 2006).

The increase in lung cancer in Malaysia is attributed to increase in smoking habit among the citizens and it was found that majority of male lung cancer patients were smokers. A study has proved that in 1991-1999 period, 92% of male lung cancer patients were smokers (Chong-kin, 2002). It seems that smoking habit among Malaysians today is becoming worsened even though a lot of smoking prevention campaigns has been promoted throughout the years. Because of this close relationship between smoking habit and lung cancer, as number of smokers increase, the number of lung cancer cases also increases. Therefore, high provisions are required to overcome this problem including treating the cancer patients.

Nowadays, various treatments are available for lung cancer around the world. The treatment will be given depending on the types and also the severity of the cancer.

Lung cancer can be treated by chemotherapy, surgery and radiation or any combinations of these three types of treatments (Dillman et al., 1990). Chemotherapy is usually used as the first line treatment on lung cancer or as an additional treatment after surgery. In chemotherapy, cytotoxic drugs have been used to kill the cancer cells by intravenous or oral administration and combinations of drugs usually given in series of treatments within a period with breaks. Thus, the body can recover before receiving the next dosage. Examples of drugs which are commonly used to treat lung cancer are carboplatin, cisplatin, docetaxel, erlotinib and irinotecan (Rosti et al., 2006). These drugs can be used individually or by combining them to produce synergistic effect. However, there are still some disadvantages of these drugs such as toxic effect to human body and killing of normal cell apart from the cancerous cells. In some cases, cancer can reappear even after surgery and chemotherapy treatment and this is one of the weaknesses of current cancer treatment.

Developments in scientific research introduce various advanced methods of cancer treatment such as targeted cancer therapy, which has become a popular approach nowadays because of its high potential of effectiveness if compared to conventional therapy. Targeted cancer therapy can be explained by the use of drugs to target certain protein, which involves in cancer cell growth and proliferation in order to inhibit or activate them. Targeted therapy works in many ways including targeting proteins in cellular signal transduction pathways which govern basic cellular function and activity, and cancer cell death or apoptosis. Targeted cancer therapy might become more effective than chemotherapy because this type of treatment focuses on molecular and cellular changes that specific to the cancer and less harmful to the normal cells. There are several pathways where protein can be targeted to inhibit the cancer growth and kill the cancer such as alterations in proliferative signaling

(Hennessy et al., 2005), tumour-associated angiogenesis (Carmeliet, 2005) and cell-survival pathways (Kang & Reynolds, 2009). Inhibitors, such as small molecule have been shown to inhibit regulatory proteins involved in cancer cell growth (Hennessy et al., 2005) and the previous studies have indicated the high potential of small molecules for cancer treatment. Therefore, designing new potent inhibitors has become very important to tackle current situation in order to decrease number of death caused by lung cancer.

Development of cancer cells in the human body is associated with several molecular abnormalities such as evasion of apoptosis and disturbance of signaling pathways. Evasion of apoptosis can be caused by many factors such as mutation in the gene that control the anti-apoptotic proteins thus leading to over expression of the proteins and finally prevent apoptosis from occurring. Disturbance of signaling pathways can happen due to mutations, amplification of tyrosine kinases or mutation of the regulatory protein itself (Hennessy et al., 2005). Therefore, targeting the protein responsible for the abnormalities in both apoptosis and growth signaling pathway with certain drugs will be a good alternative lung cancer treatment.

This study attempts to design new potential small molecules inhibitors in order to target Bcl-2, Bcl-xL and PI3K-γ, which have been reported to be over expressed in lung cancer patients. These proteins were studied and analyzed by using group based-quantitative structural activity relationship (G-QSAR) and molecular docking approach.

1.2 RESEARCH OBJECTIVE

1. To identify the specific known cancer proteins important in cancer cell proliferation and development (from literature).
2. To generate new candidates small inhibitory molecules against Bcl-2, Bcl-xL and PI3Kγ.
3. To validate the G-QSAR models by molecular docking of newly generated compounds with available crystal structure of Bcl-2, Bcl-xL and PI3Kγ.

1.3 RESEARCH HYPOTHESIS

1. B-cell lymphoma-2 (Bcl-2), B-cell lymphoma extra large (Bcl-xL) and Phospoinositide-3-kinase gamma (PI3Kγ) are important proteins involved in cancer survival and proliferation, which can be targeted for lung cancer therapies.
2. New potential inhibitors for Bcl-2, Bcl-xL and PI3Kγ can be generated based on G-QSAR models obtained from G-QSAR analysis of available Bcl-2, Bcl-xL and PI3Kγ inhibitor dataset.
3. Molecular docking of newly generated compounds into available crystal structures in PDB (Protein Databank) (http://www.rcsb.org/) website may validate the G-QSAR models of Bcl-2, Bcl-xL and PI3Kγ inhibitors.

CHAPTER TWO

LITERATURE REVIEW

2.1 LUNG CANCER

Lung cancer is uncontrolled cell growth in the tissues of the lungs or the cells lining the airways. It occurs when normal lung cells become cancer cells, usually after series of mutations, and begin to divide out of control.

2.1.1 Small Cell Lung Cancer and Non-Small Cell Lung Cancer

Lung cancer can be categorized into two groups, which are small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). SCLC accounts for 15 to 20 percent of all lung cancer cases. It is characterized as an aggressive cancer with rapid doubling time and early metastatic behavior. Majority of SCLC patients do not manifest the symptoms until shortly before diagnosis. SCLC can be classified as limited or extensive stage SCLC. In limited stage SCLC, the tumour is found on ipsilateral side of the chest but up to 60 to 70 percent of SCLC are extensive stage and usually spread to other regions of the body at the time of diagnosis (Rosti et al., 2006).

NSCLC comprises 80-85 percent of all lung cancer cases and can be further divided into adenocarcinoma, squamous cell carcinoma and large cell carcinoma. Adenocarcinoma accounts for 40 percent of all lung cancer cases and it is the commonest cancer among non-smokers. It develops at the outer region of the lungs in the cells that produce mucus and other substances (Sato, Shames, & Gazdar, 2007). Squamous cell carcinoma comprises 25-30 percent of lung cancer cases and is highly associated with smoking. Squamous cell carcinoma usually develops at the lining of