



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-2، Bcl-x1 و PI3K- γ تلعب دوراً هاماً في تكاثر وتطور خلايا سرطان الرئة. يسبب Bcl-2 و Bcl-x1 تكاثراً غير مضبوطاً للخلايا عن طريق تثبيط الموت المبرمج للخلية بينما لوحظ اضطراباً في نشاط PI3K- γ مترافقاً مع نكاث 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 لـ Bcl-x1 هي SA-kNN ($q^2=0.75$) و STP-MLR ($r^2=0.80$ ، $q^2=0.68$)، MLR-SA ($r^2=0.78$) استخدمت هذه النماذج لبناء جزيئات جديدة ذات فعالية مثبطة محتملة. فيما يتعلق بعلاقة البنية بالتأثير لـ 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 النماذج لتوليد 4969، 1536 و 7576 جزيئية ذات فعالية محتملة مثبطة. وبالمثل لـ PI3K- γ فقد تم بناء ثلاث نماذج جزيئية مهمة وتم توليد 2040، 2244 و 2560 جزيئية جديدة. وقد تم تقييم ودراسة الجزيئات المتولدة من النماذج التي تم بناؤها باستخدام طريقة التفاعلات الجزيئية مع البنى البلورية المتوفرة 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-x1 بطريقة XP ارتباطاً مع حمض الأسبارجين 136، الأرجينين 139، الغليسين 138 و التيروسين 101. في حين أن معظم مثبطات Bcl-2 أظهرت ارتباطاً مع الليوسين 96، آلانين 108، فينيل آلانين 112 و تيروسين 67. 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 الهيدروجينية والتفاعلات الكارهة للماء. وبناءً على ذلك فإن المركبات التي تم توليدها من النماذج السابقة قد ارتبطت 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, Kulliyah 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**

Copyright ©2014 by Nadia Hanis binti Abdul Samat. All rights reserved.

**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**

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder except as provided below.

1. Any material contained in or derived from this unpublished research may be used by others in their writing with due acknowledgement.
2. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purposes.
3. The IIUM library will have the right to make, store in a retrieved system and supply copies of this unpublished research if requested by other universities and research libraries.

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 through 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

Abstract	ii
Abstract in Arabic	iii
Approval Page.....	iv
Declaration Page	v
Copyright Page.....	vi
Acknowledgements	vii
List of Tables	xii
List of Figures	xiv
List of Symbols	xviii
List of Equations	xix
CHAPTER ONE: INTRODUCTION.....	1
1.1 Background of the Study.....	1
1.2 Research Objective.....	4
1.3 Research Hypothesis	4
CHAPTER TWO: LITERATURE REVIEW	5
2.1 Lung Cancer	5
2.1.1 Small Cell Lung Cancer and Non-Small Cell Lung Cancer.....	5
2.1.2 Causes and Symptoms of Lung Cancer	6
2.1.3 Conventional Treatment of Lung Cancer	7
2.1.4 Targeted Cancer Therapy	8
2.2 Cell Regulation of Apoptosis.....	8
2.2.1 Evasion of apoptosis by lung cancer	9
2.3 Cell Signal Transduction- PI3K/AKT Pathway.....	10
2.3.1 Phosphoinositide-3-Kinase Pathway Disturbance.....	10
2.4 B-CELL LYMPHOMA-2 (BCL-2)	11
2.5 B-CELL LYMPHOMA-Extra Large (BCL-XL).....	14
2.6 Phosphatidylinositide 3-KINASE-GAMMA (PI3K- γ).....	18
2.7 Quantitative Structural Activity Relationship (QSAR) And Group- Based Quantitative Structural Activity Relationship (G-QSAR).....	21
2.7.1 Sphere Exclusion for Data Selection	23
2.7.2 Simulated Annealing for Variable Selection	24
2.7.3 Stepwise Regression for Variable Selection.....	25
2.7.4 Multiple Linear Regressions for Model Building.....	27
2.7.5 K-Nearest Neighbor; Model Building Method.....	27
2.7.6 Partial Least Square Regression for Model Building	28
2.8 Molecular Docking.....	28
2.8.1 HTVS and XP Docking	30
CHAPTER THREE: GROUP-BASED QUANTITATIVE STRUCTURAL ACTIVITY RELATIONSHIP	32
3.1 GQSAR	32
3.2 G-QSAR Methodology	32
3.2.1 Data mining	32

3.2.2	Compounds' Fragmentation	34
3.2.2.1	Bcl-xL Compounds' Fragmentation	34
3.2.2.2	Bcl-2 Compounds' Fragmentation.....	36
3.2.2.3	Pi3k- γ Compounds' Fragmentation.....	38
3.2.3	Molecular Descriptor Calculation	40
3.2.4	Creation of Training and Test Sets	40
3.2.5	Variable Selection and Model Building	41
3.2.6	Model Evaluation and Validation.....	42
3.3	Results and Discussion.....	43
3.3.1	G-QSAR for Bcl-xL	43
3.3.1.1	Statistical Analysis of the Models	43
3.3.1.2	Model SA-MLR.....	47
3.3.1.3	Model STP-MLR	49
3.3.1.4	Model SA-kNN.....	51
3.3.1.5	Description of the Models' Descriptors.....	52
3.3.1.5.1	Electrotopological State Index	52
3.3.1.5.2	Electrotopological State Number	60
3.3.1.5.2	Chi (Retention Index).....	63
3.3.1.5.3	Polar Surface Area	63
3.3.1.5.4	Cluster, Path Cluster and Chain Path Count	64
3.3.1.5.5	Element Count.....	66
3.3.1.5.6	Keir shape descriptor	67
3.3.1.5.7	Octanol water partition coefficient	68
3.3.1.6	Conclusion	69
3.3.2	G-QSAR for Bcl-2.....	69
3.3.2.1	Statistical Analysis of the Models	69
3.3.2.2	Model STP-MLR	73
3.3.2.3	Model SA-MLR.....	76
3.3.2.4	Model SA-kNN.....	78
3.3.2.5	Descriptions of Model's Descriptor.....	78
3.3.2.5.1	Electrotopological State Index	79
3.3.2.5.2	Electrotopological State Number	83
3.3.2.5.3	Chi and Chi Chain.....	87
3.3.2.5.4	Path Count and Path Cluster	88
3.3.2.5.5	Element Count.....	89
3.3.2.5.6	Keir Shape Descriptor	90
3.3.2.5.7	Polarizability	90
3.3.2.5.8	Hydrophobicity	91
3.3.2.6	Conclusion	91
3.3.3	G-QSAR for Pi3k- γ	92
3.3.3.1.1	Statistical Analysis of the Models	92
3.3.3.1.2	Model STP-MLR	95
3.3.3.1.3	Model SA-MLR.....	97
3.3.3.1.4	Model STP-PLS	99
3.3.3.1.5	Descriptions of Model's Descriptor.....	101
3.3.3.1.5.1	Electrotopological State Index	101
3.3.3.1.5.2	Electrotopological State Number	104
3.3.3.1.5.3	Chi, Chi Cluster and Chiv Chain	106
3.3.3.1.5.4	Path Count.....	107

3.3.1.5.5 Element Count, H-Donor Count and Rotatable Bonds Count.....	107
3.3.1.5.6 Keir Shape Descriptor.....	108
3.3.1.5.7 Polar Surface Area.....	109
3.3.1.5.8 Hydrophobicity.....	109
3.3.1.6 Conclusion.....	110
3.4 G-QSAR models Compound Generation.....	110
CHAPTER FOUR: MOLECULAR DOCKING.....	112
4.1 Molecular docking Methodology.....	112
4.1.1 Protein Preparation.....	112
4.1.2 Ligand Preparation.....	113
4.1.3 Receptor Grid Generation.....	114
4.1.4 Ligand Docking.....	114
4.2 Result and Discussion.....	118
4.2.1 Molecular Docking of Bcl-xL Inhibitors.....	118
4.2.1.1 HTVS Docking of Bcl-xL Inhibitors.....	119
4.2.1.2 XP Docking of Bcl-xL Inhibitors.....	120
4.2.1.3 Conclusion of Bcl-xL Inhibitors Docking.....	123
4.2.2 Molecular Docking of Bcl-2 Inhibitors.....	123
4.2.2.1 HTVS Docking of Bcl-2 Inhibitors.....	124
4.2.2.2 XP Docking of Bcl-2 Inhibitors.....	125
4.2.2.3 Conclusion of Bcl-2 Inhibitors Docking.....	128
4.2.3 Molecular Docking of PI3K- γ Inhibitors.....	129
4.2.3.1 HTVS Docking of PI3K- γ Inhibitors.....	129
4.2.3.2 XP Docking of PI3K- γ Inhibitors.....	130
4.2.3.3 Conclusion of PI3K- γ Inhibitors Docking.....	134
CHAPTER FIVE: DISCUSSION.....	135
5.1 G-QSAR Contributing Descriptor and Molecular Docking of Newly Generated bcl-xL Inhibitors.....	136
5.2 G-QSAR and Molecular Docking of Bcl-2 Inhibitors.....	138
5.3 G-QSAR and Molecular Docking of PI3K- γ Inhibitors.....	140
CHAPTER SIX: CONCLUSION.....	143
6.1 Suggestions of Future Study.....	145
REFERENCES.....	146
APPENDIX.....	158
Table A1.....	158
Table A2.....	160
Table B1.....	161
Table B2.....	163
Table C1.....	164
Table C2.....	167
Table D1.....	167
Table D2.....	169
Table E1.....	169

Table E2	171
Table F1	171
Table F2	173
Table G1	173
Table G2	180
Table H1	181
Table H2	187
Table I1	188
Table I2	194
Table J1	195
Table J2	196
Table J3	197
Table J4	198
Table K1	199
Table K2	200
Table K3	201
Table K4	202
Table L1	203
Table L2	204
Table L3	207
Table L4	208

LIST OF TABLES

<u>Table No.</u>	<u>Page No.</u>
3.1. Number of calculated descriptors	40
3.2. Sphere exclusion value	41
3.3. G-QSAR models' created by combination of variable selection method and model building method.	42
3.4. Unicolumn statistic results for all Bcl-xL inhibitors G-QSAR models.	44
3.5. Statistical analysis of Bcl-xL inhibitors G-QSAR models.	45
3.6. Contributing Descriptors for SA-kNN model with the range of regression coefficient.	52
3.7. Unicolumn statistic result for all three Bcl-2 G-QSAR models.	72
3.8. Statistical parameter of all three Bcl-2 inhibitors G-QSAR models.	73
3.9. Contributing Descriptors for SA-kNN model with the range of regression coefficient.	78
3.10. Unicolumn statistic result for all three PI3K- γ inhibitors G-QSAR models.	93
3.11. Statistical parameters for all three PI3K- γ inhibitors G-QSAR models.	93
3.12. Total number of compound generated (virtually) based on Bcl-xL inhibitors G-QSAR models.	111
3.13. Total number of compound generated (virtually) based on Bcl-2 inhibitors G-QSAR models.	111
3.14. Total number of compound generated (virtually) based on PI3K- γ inhibitors G-QSAR models.	111
4.1. Number of selected compounds from HTVS docking for XP docking into Bcl-xL.	119
4.2. Number of selected compounds from HTVS docking for XP docking for Bcl-2	124

4.3. Number of selected compounds from HTVS docking for XP docking for PI3K- γ .

130

LIST OF FIGURES

<u>Figure No.</u>	<u>Page No.</u>
2.1. Crystal structure of Bcl-2 protein (Petros et al., 2004).	12
2.2. Crystal structure of Bcl-xL (Muchmore et al., 2012).	15
2.3. Overall structure of PI3K γ with bound ATP (Walker et al., 1999).	18
3.1. Fragmentation pattern of Bcl-xL inhibitors.	35
3.2. Fragmentation pattern of Bcl-2 inhibitors.	37
3.3. Fragmentation pattern for PI3K- γ inhibitors	39
3.4. Scatter plot of actual versus predicted activity for all Bcl-xL G-QSAR models.	46
3.5. Percentage of contribution of descriptors for SA-MLR model (Bcl-xL inhibitors).	48
3.6. Percentage of contribution of descriptors for STP-MLR model (Bcl-xL inhibitors).	50
3.7. Compound 291.	54
3.8. Compound 138.	54
3.9. Compound 298.	54
3.10. Compound 165.	56
3.11. Compound 270.	57
3.12. Compound 303.	57
3.13. Compound 239.	57
3.14. Compound 302.	57
3.15. Compound 213.	58
3.16. Compound 109.	59
3.17. Compound 117.	59
3.18. Compound 237.	61

3.19.	Compound 151.	61
3.20.	Compound 117.	62
3.21.	Compound 106.	63
3.22.	Compound 303.	64
3.23.	Compound 252.	66
3.24.	Compound 228.	67
3.25.	Compound 134.	67
3.26.	Scatter plot of actual versus predicted activity for all Bcl-2 G-QSAR models.	71
3.27.	Percentage of contribution of descriptors for STP-MLR model (Bcl-2 inhibitors).	75
3.28.	Percentage of contribution of descriptors for SA-MLR model (Bcl-2 inhibitors).	77
3.29.	Compound 77.	80
3.30.	Compound 123.	80
3.31.	Compound 133.	81
3.32.	Compound 90.	82
3.33.	Compound 121.	83
3.34.	Compound 78.	84
3.35.	Compound 143.	84
3.36.	Compound 123.	85
3.37.	Compound 75.	85
3.38.	Compound 78.	86
3.39.	Compound 54.	86
3.40.	Compound 28.	87
3.41.	Compound 102.	88
3.42.	Compound 70.	89

3.43.	Scatter plot of actual versus predicted for all PI3K- γ G-QSAR models.	94
3.44.	Percentage of contribution of descriptors for STP-MLR model (PI3K- γ inhibitors).	96
3.45.	Percentage of contribution of descriptors for SA-MLR model (PI3K- γ inhibitors).	98
3.46.	Percentage of contribution of descriptors for SA-MLR model (PI3K- γ inhibitors).	100
3.47.	Compound 204.	102
3.48.	Compound 102.	102
3.49.	Compound 149.	103
3.50.	Compound 149.	104
3.51.	Compound 491.	105
3.52.	Compound 422.	106
3.53.	Compound 112.	108
4.1.	HTVS docking of ligands into Bcl-xL protein.	115
4.2.	HTVS docking of ligands series into Bcl-2 protein.	116
4.3.	HTVS docking of ligands series into PI3K- γ protein.	116
4.4.	XP docking of ligand series for Bcl-xL protein.	117
4.5.	XP docking of ligand series for Bcl-2 protein.	118
4.6.	XP docking of ligand series for PI3K- γ protein.	118
4.7.	BMS1-BCLXL on Bcl-xL crystal structure (PDB ID 2YXJ).	121
4.8.	BMS2-BCLXL on Bcl-xL crystal structure (PDB ID 2YXJ).	121
4.9.	BMS3-BCLXL on Bcl-xL crystal structure (PDB ID 2YXJ).	122
4.10.	ABT737 on Bcl-xL crystal structure (PDB ID 2YXJ).	122
4.11.	ORIBCLXL on Bcl-xL crystal structure (PDB ID 2YXJ).	123
4.12.	BMS1-BCL2 on Bcl-2 crystal structure (PDB ID 4AQ3).	126
4.13.	BMS2-BCL2 on Bcl-2 crystal structure (PDB ID 4AQ3).	126

4.14.	BMS3-BCL2 on Bcl-2 crystal structure (PDB ID 4AQ3).	127
4.15.	ABT737 on Bcl-2 crystal structure (PDB ID 4AQ3).	127
4.16.	ORIBCL2 on Bcl-2 crystal structure (PDB ID 4AQ3).	128
4.17.	NATBCL2 on Bcl-2 crystal structure (PDB ID 4AQ3).	128
4.18.	BMS1-PI3K γ on PI3K- γ crystal structure (PDB ID 3L08).	131
4.19.	BMS2-PI3K γ on PI3K- γ crystal structure (PDB ID 3L08).	132
4.20.	BMS3-PI3K γ on PI3K- γ crystal structure (PDB ID 3L08).	132
4.21.	NVP-BGT226 on PI3K- γ crystal structure (PDB ID 3L08).	133
4.22.	ORIP3K γ on PI3K- γ crystal structure (PDB ID 3L08).	133
4.23.	NAT PI3K γ on PI3K- γ crystal structure (PDB ID 3L08).	134

LIST OF SYMBOLS

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

LIST OF EQUATIONS

Equation 1 Sphere exclusion	24
Equation 2 Multiple linear regression	27
Equation 3 Model's equation of SA-MLR for Bcl-xL	47
Equation 4 Model's equation of STP-MLR for Bcl-xL	49
Equation 5 Electrotopological state indice	53
Equation 6 First kappa shape index	67
Equation 7 Lipid and aqueous partition coefficient	68
Equation 8 LogP of lipid and aqueous partition coefficient	68
Equation 9 Model's equation of STP-MLR for Bcl-2	73
Equation 10 Model's equation of SA-MLR for Bcl-2	76
Equation 11 PolarizabilityAHC (atomic hybrid component)	90
Equation 12 Model's equation of STP-MLR for PI3K- γ	95
Equation 13 Model's equation of SA-MLR for PI3K- γ	97
Equation 14 Model's equation of STP-PLS for PI3K- γ	99

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 PI3K γ .
3. To validate the G-QSAR models by molecular docking of newly generated compounds with available crystal structure of Bcl-2, Bcl-xL and PI3K γ .

1.3 RESEARCH HYPOTHESIS

1. B-cell lymphoma-2 (Bcl-2), B-cell lymphoma extra large (Bcl-xL) and Phosphoinositide-3-kinase gamma (PI3K γ) 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 PI3K γ can be generated based on G-QSAR models obtained from G-QSAR analysis of available Bcl-2, Bcl-xL and PI3K γ 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 PI3K γ 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