MOLECULAR MODELING AND COMPUTER AIDED MOLECULAR DESIGN FOR OPTIMAL DRUG DESIGN OF NEW HIGHLY POTENTIAL ANTI-TUBERCULOSIS AGENTS AND ANTI-CANCER AGENTS

PHARIT KAMSRI

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DOCTOR OF PHILOSOPHY
MAJOR IN CHEMISTRY
FACULTY OF SCIENCES
UBON RATCHATHANI UNIVERSITY
ACADEMIC YEAR 2015

UBON RATCHATHANI UNIVERSITY

THESIS APPROVAL

DOCTOR OF PHILOSOPHY

MAJOR IN CHEMISTRY FACULTY OF SCIENCE

TITLE MOLECULAR MODELING AND COMPUTER AIDED MOLECULAR DESIGN FOR OPTIMAL DRUG DESIGN OF NEW HIGHLY POTENTIAL ANTI-TUBERCULOSIS AGENTS AND ANTI-CANCER AGENTS

AUTHOR MR. PHARIT KAMSRI

EXAMINATION COMMITTEE

PROF. DR. SUPA HANNONGBUA	CHAIRPERSON
ASSOC. PROF. DR. PORNPAN PUNGPO	MEMBER
ASST. PROF. DR. CHAN INNTAM	MEMBER

ADVISOR .....

(ASSOC. PROF. DR. PORNPAN PUNGPO)

....

(ASSOC. PROF. DR. UTITH INPRASIT) (ASSOC. PROF. DR. ARIYAPORN PONGRAT)

DEAN, FACULTY OF SCIENCE VICE PRESIDENT FOR ACADEMIC AFFAIRS

ACADEMIC YEAR 2015

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Associate Professor Dr. Pornpan Pungpo, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject. I would never have reached my goal without her. I am heartily thankful to Professor Dr. Peter Wolschann, Professor Dr. Stephan Irle, Professor Dr. Weiliang Zhu and Professor Dr. Wolfgang Sippl, who provide me for fruitful suggestions and discussions on various aspects of my work. I am deeply appreciated Professor Dr. Supa Hannongbua for her helpful guidance and suggestion throughout the entire study. Furthermore, I would like to thank Assistant Professor Dr. Chan Inntam, my thesis committee for his helpful suggestion. Special thanks to Assistant Professor Dr. Patcheenart Saparpakorn and Dr. Auradee Punkvang for valuable assistance and continuing supports in various way.

I would like to thank the scholarship from The Royal Golden Jubilee Ph.D (RGJ-Ph.D.) Program (PHD0004/2554), for financial support. The Thailand Research Fund (DBG5380006 and DBG5680003), Higher Education Research Promotion, the National Research Council of Thailand and ASEA-UNINET are gratefully acknowledged for research supports. Department of Chemistry, Faculty of Science, Ubon Ratchathani University, Thailand, Department of Chemistry, Faculty of science, Kasetsart University, Thailand, Vienna Scientific Cluster of University of Vienna, Austria, Quantum chemistry group and Institute of Transformative Bio-Molecules (WPI-ITbM), Nagoya University, Japan, Drug Discovery and Design Centre (DDDC), Shanghai Institute of Meteria Medica (SIMM), Chinese Academy of Science, China, Institute of Pharmacy, Martin Luther University of Halle-Wittenberg, Germany and the High Performance Computing Center of the National Electronics and Computer Technology (NECTEC), Thailand are also gratefully acknowledged for computational resource and software facilities.

My utmost gratitude must be extended to my family members, who always love and understand whatever I am. They always encourage and unconditional support me throughout in the entire study.

Pharit Kamsri

(Pharit Kamsri)

Researcher

บทคัดย่อ

เรื่อง : การจำลองแบบและการออกแบบโมเลกุลด้วยการคำนวณเพื่อการออกแบบสารออกฤทธิ์ตัวใหม่ที่มีศักยภาพสูงในการต้านโรควัณโรคและสารออกฤทธิ์ต้านโรคมะเร็ง

ผู้วิจัย : พฤทธิ์ คำศรี

ชื่อปริญญา : ปรัชญาดุษฎีบัณฑิต

สาขาวิชา : เคมี

อาจารย์ที่ปรึกษา: รองศาสตราจารย์ ดร.พรพรรณ พึ่งโพธิ์

คำสำคัญ : วัณโรค, มะเร็ง, คิวเอสเออาร์, เอ็มดีซิมมูเลชั่น, การคัดสรรเสมือนจริง

การจำลองแบบและการออกแบบโมเลกุลด้วยการคำนวณกำลังเป็นเครื่องมือสำคัญ ในการค้นหาและปรับเปลี่ยนสารยับยั้งได้อย่างรวดเร็วและมีประสิทธิภาพ ในงานวิจัยนี้ระเบียบวิธีการจำลองแบบและการออกแบบโมเลกุลด้วยการคำนวณถูกประยุกต์ใช้เพื่อที่จะทำให้เข้าใจถึงข้อมูลทางโครงสร้างและพัฒนาเป็นสารต้านโรควัณโรคและสารออกฤทธิ์ต้านโรคมะเร็งชนิดใหม่ที่มีประสิทธิภาพสูง เอนไซม์ตัวแรกที่เลือกมาใช้ในการพัฒนายารักษาโรควัณโรค เอนไซม์อีโนอิลเอซีพีรีดักเตสหรือเอนไซม์ InhA ของเชื้อวัณโรคเป็นเอนไซม์เป้าหมายในการออกฤทธิ์ยับยั้งของตัวยาหลักในการรักษาโรควัณโรคอย่างยาไอโซไนอาซิด การดื้อต่อยาไอโซไนอาซิดอย่างรุนแรง เกี่ยวเนื่องกับการกลายพันธุ์ของเอนไซม์ InhA และเอนไซม์คะตะเลสเปอร์ออกซิเดสหรือเอนไซม์ katG จากปัญหาการดื้อยาของตัวยาไอโซไนอาซิดซึ่งเกิดจาก katG สารอนุพันธ์ไดฟีนิลอีเทอร์และสารอนุพันธ์เบนโซฟลูแรน ไพโรลิดีน ไพราโซลถูกพัฒนาขึ้นเพื่อใช้เป็นสารยับยั้งเอนไซม์ InhA โดยตรงเพื่อที่จะให้บรรลุเป้าหมายในการเพิ่มประสิทธิภาพในการยับยั้งเอนไซม์ InhA และการยับยั้งเชื้อวัณโรค ระเบียบวิธีคิวเอสเออาร์และการจำลองพลวัติเชิงโมเลกุลถูกประยุกต์ใช้เพื่อศึกษาข้อมูลที่มีความสำคัญ การยับยั้งการส่งสัญญาณระหว่างแบคทีเรียกับโฮสต์คือแนวคิดใหม่สำหรับการออกแบบยารักษาโรควัณโรคของเอนไซม์เป้าหมายตัวที่สองที่เลือกของยารักษาโรควัณโรค เอนไซม์เซอรีน/ทรีโอนีนไคเนส จี หรือเอนไซม์ PknG ซึ่งเป็นเอนไซม์ที่เกี่ยวข้องกับกระบวนการส่งผ่านสัญญาณถูกบ่งชี้เป็นเอนไซม์เป้าหมายที่มีศักยภาพในการพัฒนายารักษาวัณโรค การจำลองพลวัติเชิงโมเลกุลร่วมกับระเบียบวิธีคิวเอสเออาร์ถูกใช้ในการศึกษาความต้องการทางโครงสร้างของสารอนุพันธ์เบนโซ-ไทโอฟีนเพื่อออกแบบสารยับยั้งเอนไซม์ PknG ชนิดใหม่ที่มีศักยภาพในการยับยั้งสูง ผลการวิเคราะห์ร่วมกันระหว่างการจำลองพลวัติเชิงโมเลกุลและระเบียบวิธีคิวเอสเออาร์ให้ข้อมูลทางโครงสร้างที่เป็น

ประโยชน์ในระดับโมเลกุลซึ่งเป็นแนวทางที่สำคัญในการออกแบบสารยับยั้งเอนไซม์ PknG ที่มีศักยภาพสูงเพื่อใช้เป็นยาต้านวัณโรค นอกจากนี้การคัดสรรเสมือนจริงเชิงโครงสร้างถูกประยุกต์ใช้ในการค้นหาสารโครงสร้างสารชนิดใหม่ของสารยับยั้งเอนไซม์ InhA และ PknG ตัวใหม่เพื่อใช้เป็นยาต้านโรควัณโรค เพื่อที่จะพัฒนาสารต้านโรคมะเร็งที่มีศักยภาพสูงและมีความเป็นพิษที่ต่ำ สารอนุพันธ์เอซาแนพโธควิโนนที่ต่อกับวงพิวโรลจึงถูกพัฒนาขึ้น การจำลองพลวัติเชิงโมเลกุลร่วมกับระเบียบวิธีคิวเอสเออาร์จึงถูกประยุกต์ใช้กับสารอนุพันธ์เอซาแนพโธควิโนนเพื่อที่จะศึกษาลักษณะทางโครงสร้างที่สำคัญ การวางตัวและอันตรกิริยาในการจับกับดีเอ็นเอ ดังนั้นผลจากการศึกษาคือความรู้ที่ให้ลักษณะที่สำคัญทางโครงสร้างและแนวทางที่เป็นประโยชน์สำหรับการปรับเปลี่ยนโครงสร้าง นำไปสู่การออกแบบสารอนุพันธ์เอซาแนพโธควิโนนที่ต่อกับวงพิวโรลชนิดใหม่ที่มีประสิทธิภาพในการยับยั้งสูง การศึกษานี้ได้นำเสนอสารยับยั้งที่ได้จากออกแบบใหม่ ซึ่งมีค่ากัมมันต-ภาพในการยับยั้งจากการทำนายสูงกว่าสารต้นแบบและโครงสร้างสารใหม่ของสารต้านโรควัณโรคและสารออกฤทธิ์ต้านโรคมะเร็งได้เป็นผลสำเร็จ

ABSTRACT

TITLE : MOLECULAR MODELING AND COMPUTER-AIDED MOLECULAR DESIGN FOR OPTIMAL DRUG DESIGN OF NEW HIGH POTENTIAL ANTI-TUBERCULOSIS AGENTS AND ANTI-CANCER AGENTS
AUTHOR : PHARIT KAMSRI
DEGREE : DOCTOR OF PHILOSOPHY
MAJOR : CHEMISTRY
ADVISOR : ASSOCIATE PROFESSOR PORNPAN PUNGPO, PhD
KEYWORDS : TUBERCULOSIS, CANCER, QSAR, MD SIMULATION, VIRTUAL SCREENING

Molecular modeling and computer-aided molecular design approaches is becoming an essential tool in assisting fast and cost-efficient lead discovery and optimization. In the present study, molecular modeling and computer-aided molecular design approaches were applied to understand the molecular basis for developing new and more potent anti-tuberculosis (TB) and anti-cancer agents. The first target for anti-TB agents, enoyl-ACP reductase (InhA) of *M. tuberculosis*, has been shown to be the primary target of the isoniazid, frontline drugs. The high levels of INH resistance arise from the mutations in InhA and catalase-peroxidase (KatG) enzymes. Because of the INH resistance associated with KatG mutations, diphenyl ether and benzofuran pyrrolidine pyrazole derivatives have been developed as the direct InhA inhibitors. To achieve the structural basis to improve InhA and antimycobacterial activity, QSAR and molecular dynamics (MD) simulations were applied to elucidate beneficial information. Inhibition of bacterial and host cell signaling is a novel drug discovery concept in the second selected target of anti-TB agents. Serine/threonine protein kinase G (PknG), an enzyme in signal transduction pathways, has been identified as a promising target. MD simulations combined with 3D-QSAR studies were used to investigate the structural requirements of benzothiophene derivatives to rational design new potent PknG inhibitors. The integrated results from MD simulations and QSAR approaches provided useful structural information at the molecular level, a powerful

guideline for designing effective PknG inhibitors as anti-TB agents. Moreover, a structure-based virtual screening approach was applied to identify novel scaffolds of InhA and PknG inhibitors as anti-TB agents. In an attempt to develop highly effective anti-cancer agents showing lower toxicity levels, azanaphthoquinone annelated pyrrole analogues have been developed. MD simulations and QSAR studies were applied on azanaphthoquinone derivatives to evaluate their key structural features, binding mode, and binding interactions in the DNA duplex. Accordingly, the results were informative, providing key features and beneficial guidelines for further modification leading to the design of new and more potent azanaphthoquinone annelated pyrrole compounds. Newly designed compounds with higher predicted activities compared with the parent compounds and novel scaffolds of anti-TB agents and anti-cancer agents were proposed in this study.

CONTENTS

LIST OF TABLES

LIST OF TABLES (CONTINUED)

LIST OF FIGURES

LIST OF FIGURES (CONTINUED)

LIST OF FIGURES (CONTINUED)

LIST OF FIGURES (CONTINUED)

LIST OF FIGURES (CONTINUED)

LIST OF FIGURES (CONTINUED)

LIST OF FIGURES (CONTINUED)

LIST OF FIGURES (CONTINUED)

LIST OF FIGURES (CONTINUED)

ABBREVIATIONS

ABBREVIATIONS FULL WORD

Å	Angstrom
β	Beta
μm	Micrometre
μM	Micromolar
ACP	Acyl carrier protein
Ala	Alanine
Amk	Amikacin
ASP	Astex statistical potential
BFE	Binding free energy
C-terminal	Carboxy -terminal
CADD	Computer aided drug design
CAMD	Computer aided molecular design
CASTp	Computed atlas of surface topography of proteins
CDC	Disease control and prevention
Cfz	Clofazimine
Cfx	Ciprofloxacin
Cln	Cilastatin
Clr	Clarithromycin
Clv	Clavulanate
Cm	Capreomycin
CoA	Acyl-coenzyme A
CoMFA	Comparative molecular field analysis
CoMSIA	Comparative molecular similarity indices analysis
Cs	Cycloserine
Dcs	Cycloserine
DNA	Deoxyribonucleic acid
DOT	Directly observed treatment
DOTS	Direct observed therapy, short course

ABBREVIATIONS (CONTINUDED)

ABBREVIATIONS FULL WORD

E	Ethambutol
ETH	Ethionamide
ETH-NAD	Ethionamide- nicotinamide-adenine dinucleotide
FAS	Fatty acid synthase
FAS-I	Type I fatty acid synthase
FAS-II	Type II fatty acid synthase
FDA	Food and Drug Administration
FEP	Free-energy perturbation
fs	Femtosecond
G_0 phase	Gap 0 phase
G_1 phase	Gap 1 phase
G_2 phase	Gap 2 phase
GA	Genetic algorithm
GAFF	General AMBER force field
GC-rich	Guanine-cytosine-rich
Glide	Grid-based ligand docking with energetics
Gly	glycine
Gfx	Gatifloxacin
GOLD	Genetic optimisation for ligand docking
GTF	Gaussian type function
GTO	Gaussian type orbital
HBCs	High TB burden countries
H	Isoniazid
h	Hour
HIV	Human Immunodeficiency Virus
HTS	High throughput screening
HTVS	High throughput virtual screening

ABBREVIATIONS (CONTINUDED)

ABBREVIATIONS	FULL WORD
IC_{50}	The inhibitory concentration of compound require for 50 % inhibition of enzyme/organism
Ile	Isoleucine
InhA	Enoyl-ACP reductase
Ipm	Imipenem
INH-NAD	Isonicotinic-acetyl-nicotinamide-adenine dinucleotide
INH	Isoniazid
K	Kelvin
KatG	Catalase peroxidase
kcal/mol	kilocalories per mole
kg	kilogram
K_i	Inhibitory constant
Km	Kanamycin
LBDD	Ligand based drug design
LCAO	Linear combination of atomic orbitals
Leu	Leucine
Lfx	Levofloxacin
LGA	Lamarckian genetic algorithm
LIE	Linear interaction energy
log P	The logarithm of its partition coefficient between n-octanol and water
LOO	Leave-one-out
LTBI	Latent TB infection
Lys	Lysine
Lzd	Linezolid
M. bovis BCG	*Mycobacterium bovis* Bacillus Calmette–Guérin
Mbp	Million base pairs
MC	Monte Carlo
MD	Molecular dynamics

ABBREVIATIONS (CONTINUDED)

ABBREVIATIONS	FULL WORD
MDR-TB	Multidrug resistance tuberculosis
Met	Methionine
Mfx	Moxifloxacin
mg	Milligram
MIC	Minimum inhibitory concentration
MIC_{50}	Minimum Inhibitory Concentration required to inhibit the growth of 50% of organisms
MIC_{90}	Minimum Inhibitory Concentration required to inhibit the growth of 90% of organisms
MM	Molecular mechanics
MM-GBSA	Molecular Mechanics Generalized Born Surface Area
MM-PBSA	Molecular Mechanics Poisson–Boltzmann Surface Area
MO	Molecular orbital
MOE	Molecular Operating Environment
M phases	Mitosis phases
M. tuberculosis	*Mycobacterium tuberculosis*
M. tuberculosis H37Ra	*Mycobacterium tuberculosis* strain H37Ra
M. tuberculosis H37Rv	*Mycobacterium tuberculosis* strain H37Rv
M. smegmatis	*Mycobacterium smegmatis*
N-terminal	Amino-terminal
Na^+	Sodium ion
NADH	Nicotinamide adenine dinucleotide
NAD^+	Nicotinamide adenine dinucleotide
NCI	National Cancer Institute Database
NCL	Nested Chemical Library™
nM	Nanomolar
NMR	Nuclear magnetic resonance
Non-RD kinase	Non-arginine-aspartate kinas
ns	nanosecond

ABBREVIATIONS (CONTINUDED)

ABBREVIATIONS	FULL WORD
Ofx	Ofloxacin
Pas	*para*-Aminosalicylic Acid
PDB	Protein data bank
PDT	Photodynamic therapy
pH	
Phe	Phenylalanine
pKa	Acid dissociation constant
PknG	Serine/theonine kinase G
PLP	Piecewise linear potential
PLS	Partial least squares
PME	Particle Mesh Ewald
PP2C	Protein phosphatase type 2C
PRESS	Prediction Error Sum of Squares
Pro	Proline
ps	Picosecond
PTK	Protein tyrosine kinase
Pto	Protionamide
PTPs	Protein tyrosine phosphatases
q^2	Cross-validated correlation coefficient, predictive ability
QSAR	Quantitative Structure-Activity Relationship
R	Rifampicin
r^2	Non-cross-validated correlation coefficient
RESP	Restrained electrostatic potential
RMSD	Root mean square deviations
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
S	Streptomycin
S-phase	Synthesis phase
SA	Simulated Annealing

ABBREVIATIONS (CONTINUDED)

ABBREVIATIONS	FULL WORD
SASA	Solvent accessible surface area
SBDD	Structure based drug design
SBVS	Structure based virtual screening
SEE	Standard Error of Estimates
Ser	Serine
SP	Standard precision
S_{PRESS}	The Standard of Error of Prediction
Stm	Streptomycin
STO	Slater type orbital
STPKs	Serine/threonine protein kinases
TB	Tuberculosis
TDR-TB	Totally drug resistant tuberculosis
Thr	Threonine
Thz	Thioacetazone
TI	Thermodynamic Integration
TPR	Tetratricopeptide repeat
Trd	Terizidone
Tyr	Tyrosine
US$	United States dollar
Val	Valine
Vim	Viomycin
VS	Virtual Screening
WHO	World Health Organization
XDR-TB	Extensively drug resistant tuberculosis
XP	Extra precision
XXDR-TB	Extremely drug resistant tuberculosis
Y158-in	Tyrosine 158 in conformation
Y158-out	Tyrosine 158 out conformation
Z	Pyrazinamide

CHAPTER 1
INTRODUCTION

1.1 Drug discovery and design

Drugs are essential for the prevention and treatment of disease. To find new drugs entry to market, drug researchers discover new drugs through new insights into a disease process. The initial drug research, often occurring in academic and clinical research and from the commercial sector, generates data to develop a hypothesis that the inhibition or activation of a protein or pathway will result in a therapeutic effect in a disease state (Hughes et al., 2010). The entire drug discovery process with the tentative timeline is presented in Figure 1.1. Drug discovery process started from target identification and validation to find essential biological process to block disease cause. Ideally the drug candidate should be binds to a single target only (Martis et al., 2012). In drug target validation process, validation techniques help drug researchers to modulation of a desired target in disease patients through the use of whole animal models (Hughes et al., 2010; Martis et al., 2012). Following the target validation process, hit identification and lead discovery steps are developed via screening strategies. Potential lead compounds obtained from lead identification process may be assumed to have met the initial goals of the lead optimization to find new candidate compounds. Then, candidate compounds were subjected to preclinical studies is to investigate the safety of the newly molecule in drug development process. The next process after preclinical studies is the clinical studies, actual testing of the newly molecule in the human volunteers to study the safety and efficacy. This process consists of three phases, phase 1, phase 2 and phase 3 studies. Three of these steps are carried out after the drug has been launched in to the drug market. The Food and Drug Administration (FDA) issues an approval to the company, based on their data compiled from the clinical trials, the drug can be launched in the market. Overall process took times approximately 12 years and cost more than US$800 million on average (DiMasi et al., 2003). The Therefore, the process of drug discovery and development is challenging, complex process, time consuming, expensive, and

requires consideration of many aspects (Mandal et al., 2009). Several new technologies have hence been developed and applied in drug research and development to shorten the research timeline and to reduce the cost of research (Tang et al., 2006; Jorgensen, 2004).

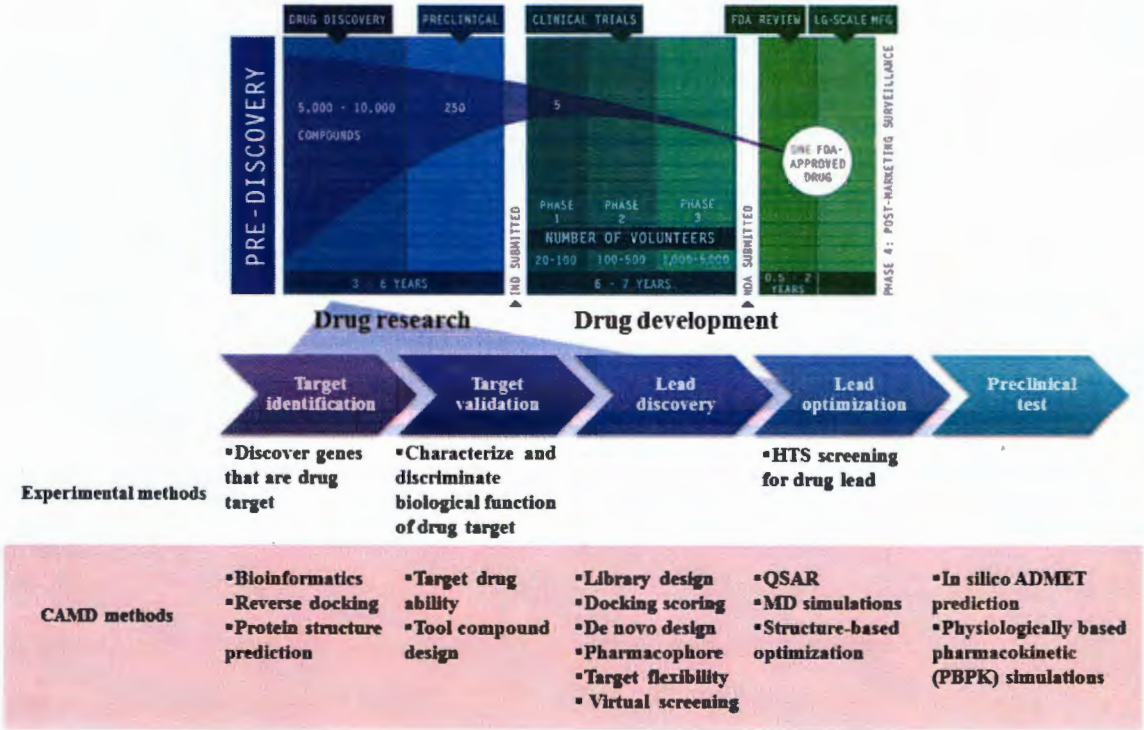


Figure 1.1 Drug discovery pipeline vs. CAMD tools

Resource: Tang et al. (2006); PhRMA, (2016)

Computer aided molecular design (CAMD) or computer aided drug design (CADD) approaches are new technologies to increase the efficiency of the drug discovery process in the past decade. Several CAMD approaches have been developed for drug research process. Particular emphasis is placed on virtual screening, molecular docking, pharmacophore, de novo design, quantitative structure–activity relationship (QSAR), evaluation of druglikeness, and advanced methods for determining protein-ligand binding affinity (Ooms, 2000; Jorgensen, 2004; Tang et al., 2006; Kapetanovic, 2008; Talele, Khedkar and Rigby, 2010). New drug compounds that discovered or optimized via CAMD approaches were reported such as Captopril, Dorzolamide, Saquinavir, Zanamivir, Oseltamivir, Aliskiren, Boceprevir, Nolatrexed,

TMI-005, LY-517717, Rupintrivir and NVP-AUY922 (Talele, Khedkar and Rigby, 2010; Kubinyi, 2006(c)). Therefore, CAMD method has even been used in the drug discovery pipeline, from target identification to lead discovery, from lead optimization to preclinical or clinical trials for a certain target. Two types of disease, tuberculosis (infectious disease) and cancer (non-infectious disease) were selected to develop new and more potent inhibitors in this study.

1.2 Tuberculosis

1.2.1 Tuberculosis and situation

Tuberculosis (TB) is a communicable infectious disease usually caused by pathogenic *Mycobacterium tuberculosis* (*M. tuberculosis*). *M. tuberculosis* is non-motile, nonsporulating, weakly gram-positive, acid-fast bacilli and aerobic that appears microscopically as straight or slightly curved rods, 1 to 4 μm in length and 0.3 to 0.6 μm wide as shown in Figure 1.2 (Sakamoto, 2012). To establish an infection, the pathogenic bacterium *M. tuberculosis* host to host transmission is transmitted between TB patients to new infected TB patient *via* aerosol droplets that contain mycobacteria as shown in Figure 1.3. *M. tuberculosis* pathogen reaches lung airways and is rapidly phagocytosed by alveolar macrophages and dendritic cells in host cell (Gengenbacher and Kaufmann, 2012; Nunes-Alves et al., 2014). The TB infected host cell induces a localized proinflamatory response that attracts macrophages and other leukocytes leads to build up a granuloma, which can contain *M. tuberculosis* (the hallmark tissue reaction of TB). Dendritic granulomas also phagocytose *M. tuberculosis* and then migrate to regional lymph nodes to present mycobacterial antigens to lymphocytes (Sakamoto, 2012; Tang Yam and Chen, 2016). The site of the TB infection can often be recognized as a granulomatous lesion.

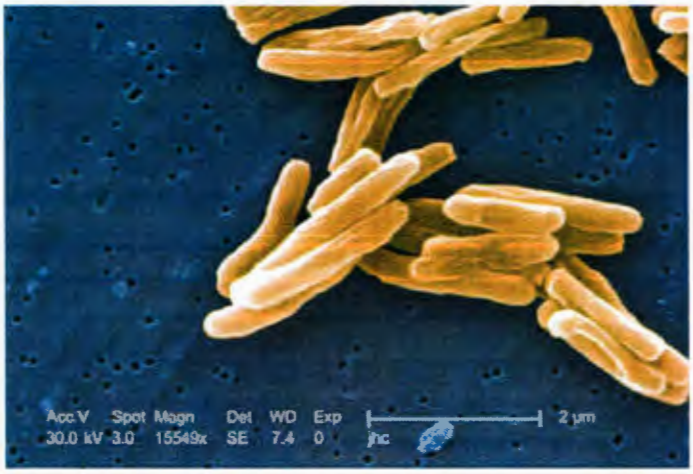


Figure 1.2 ***M. tuberculosis*** **scanning electron micrograph**

Resource: Carr (2005)

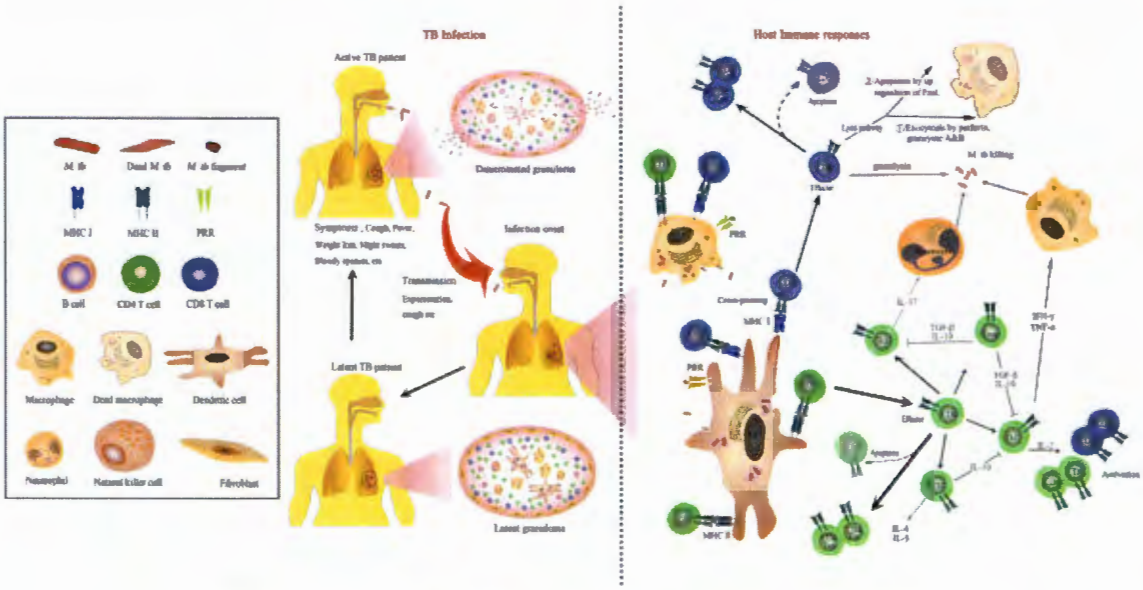


Figure 1.3 TB infection and host immune responses

Resource: Tang Yam and Chen (2016)

TB is a major global health problem. Every year, the reports of statistical TB situation from World Health Organization (WHO) were released. This disease became to leasing cause of death from infection disease worldwide in 2014 (WHO, 2015(c)). The numbers of new patient are slightly increasing every year. There were an estimated 9.6 million new TB cases with 5.4 million among men, 3.2 million among women and 1.0 million among children. The 58% patients of new TB cases in

2014 were in the South-East Asia and Western Pacific regions followed by African Region (28 %). Small numbers of new TB patient were Eastern Mediterranean Region (8%), the European Region (3%) and the Region of the Americas (3%). TB killed 1.5 million people, which approximately 890,000 were men, 480,000 were women and 140,000 were children. Thailand was reported as 18th ranging of the 22 high TB burden countries (HBCs). From, 67,726,000 population in Thailand, total new and relapse TB patients are 67,722 and total cases notified are 71,618. Among 34,394 new cases of TB case notifications 2014, 119 cases aged less than 15 years. The ratio of patient between male and female are 2.5.

1.2.2 Current treatment against tuberculosis

Current treatment of TB is based on drugs that are more than 40 years old (Zumla, Nahid and Cole, 2013; Global Alliance for TB Drug Development, 2008; Lienhardt Vernon and Raviglione, 2010).Two classification of TB drugs, first-line and second-line drugs based on their efficiency and toxicity were summarized in Table 1.1. There are 10 approved TB drug from FDA. The TB drugs are used in differing combinations in different circumstances. The first line TB drug includes the four key drugs, isoniazid, rifampicin, pyrazinamide and ethambutol that make up the initial phase of the ideal short-course chemotherapy regimen. Isoniazid was discovered in 1952. This compound is the highest bactericidal against replicating tubercle bacilli through inhibition of mycolic acid synthesis. Therefore, this drug is mostly used for anti-TB drug. Rifampicin, the first line drug found in 1963 inhibits the essential rpoB gene product β-subunit of DNA dependent RNA polymerase activity. This drug has excellent bactericidal activity on actively dividing as well as slowly metabolizing tubercle bacilli, and this property was instrumental in shortening the duration of treatment. In 1954, pyrazinamide was found to be kill tuberculosis. This compound is synthetic analogue of nicotinamide. Pyrazinamide is only weakly anti-bacterial agent against *M. tuberculosis*, but this compound had potent sterilizing activity in the relative acidic intracellular environment of macrophages via inhibition of translation and trans-translation at S1 component of 30S ribosomal subunit (Shi et.al, 2011). This drug plays a unique role in shortening the duration of tuberculosis chemotherapy. Ethambutol, the first line drug was found in 1961. This compound inhibits arabinosyl transferases involved arabinogalactan and lipoarabinomannan biosynthesis pathway in

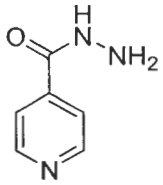
mycobacterium cell-wall biosynthesis. The attractive target of ethambutol is arabinosyl transferases.

The second line drug, para-amino salicylic acid found as anti-tuberculosis agent in 1948 was reported as dihydropteroate synthase inhibitor in folate biosynthesis. This drug is now mostly used as a second-line drug (Chakraborty et al., 2013). Ethionamide, a structural analog of isoniazid, an important second-line drug used for the TB treatment inhibits enoyl-ACP reductase function in mycolic acid biosynthesis pathway; the attractive target of tuberculosis drug was found in 1961. Cycloserine, an amino acid derivative was reported as second line anti-TB drug that found in 1955. Cycloserine works as an antibiotic by inhibiting peptidoglycan synthesis in cell-wall biosynthesis of *M. tuberculosis*. The attractive target of this drug is d-alanine racemase and ligase (Wargel et al., 1971; Bruning et al., 2011; Prosser and de Carvalho, 2013). Ofloxacin, fluoroquinolone compound is a DNA gyrase and DNA topoisomerase inhibitors that is responsible for supercoiling of DNA found in 1980 (Fillion et al., 2013; Li et al., 2014). This action inhibits DNA replication process in (Bryskier, 1993, Drlica et al., 2003(a), 2009(b), Fàbrega et al., 2009, Aldred et al., 2014). This compound is weakly bactericidal agents. Protein synthesis inhibitor, capreomycin, the cyclic polypeptide found in 1963 inhibits protein synthesis by binding to the interbridge B2a between 30S and 50S ribosomal subunits (Johansen et al., 2006, Sirgel et al., 2011). Kanamycin, an aminoglycoside compound was found in 1957. This compound was used as the second line tuberculosis drug inhibited at 16S rRNA in the 30S ribosomal subunit (Salian et al., 2012). Amikacin found in 1972 is aminoglycoside that inhibits protein synthesis by binding tightly to the conserved A site of 16S rRNA in the 30S ribosomal subunit (Sirgel et al., 2011).Streptomycin, an antibiotic purified from *Streptomyces griseus* was discovered in 1944 that is used for tuberculosis treatment and sensitive gram-positive bacteria. Streptomycin is a protein synthesis inhibitor. It binds to the S12 and 16S rRNA components of 30S ribosomal subunit (Schatz, Bugie, and Waksman, 1944; WHO, 1991(a)).

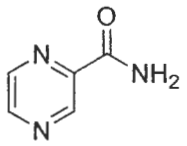
Table 1.1 Main tuberculosis drugs in clinical use and their targets

Drug (year of discovery)	Target	Effect
First-line drugs		
Isoniazid (1952)	Enoyl-ACP reductase	Inhibits mycolic acid synthesis
Rifampicin (1963)	RNA polymerase, beta subunit	Inhibits transcription
Pyrazinamide (1954)	S1 component of 30S ribosomal subunit	Inhibits translation and *trans*-translation, acidifies cytoplasm
Ethambutol (1961)	Arabinosyl transferases	Inhibits arabinogalactan biosynthesis
Second-line drugs		
Para-amino salicylic acid 1948)	Dihydropteroate synthase	Inhibits folate biosynthesi
Ethionamide (1961)	Enoyl-ACP reductase	Inhibits mycolic acid synthesis
Cycloserine (1955)	d-alanine racemase and ligase	Inhibits peptidoglycan synthesis
Ofloxacin (1980)	DNA gyrase and DNA topoisomerase	Inhibits DNA supercoiling
Capreomycin (1963)	Interbridge B2a between 30S and 50S ribosomal subunits	Inhibits protein synthesis
Kanamycin (1957)	30S ribosomal subunit	Inhibits protein synthesis
Amikacin (1972)	30S ribosomal subunit	Inhibits protein synthesis
Streptomycin (1944)	S12 and 16S rRNA components of 30S ribosomal subunit	Inhibits protein synthesis

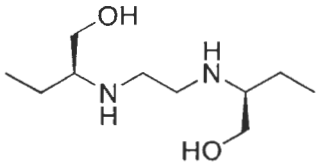
Resource: Zumla, Nahid and Cole (2013)



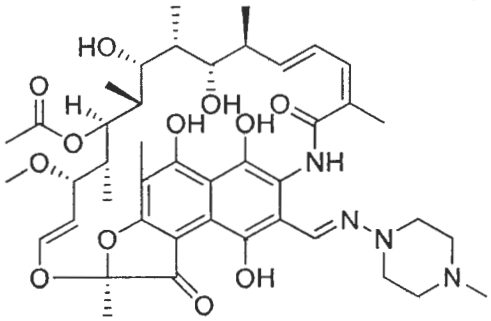
Isoniazid



Pyrazinamide

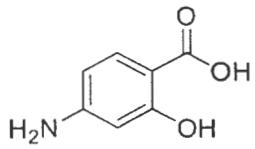


Ethambutol

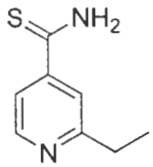


Rifampicin

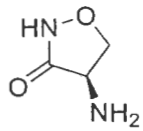
Figure 1.4 First-line anti-TB agents



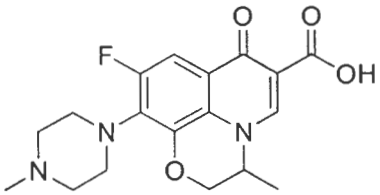
Para-amino salicylic acid



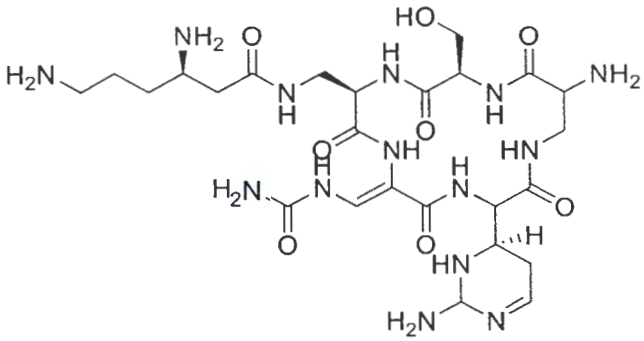
Ethionamide



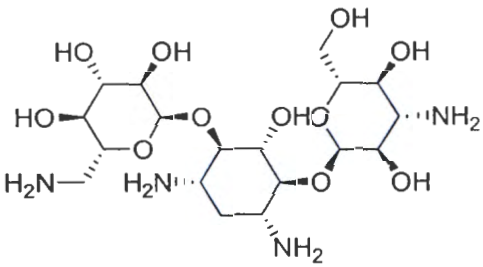
Cycloserine



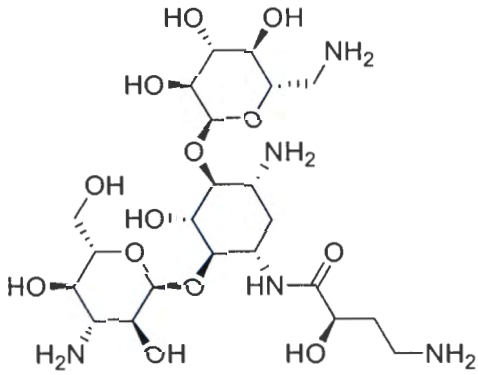
Ofloxacin



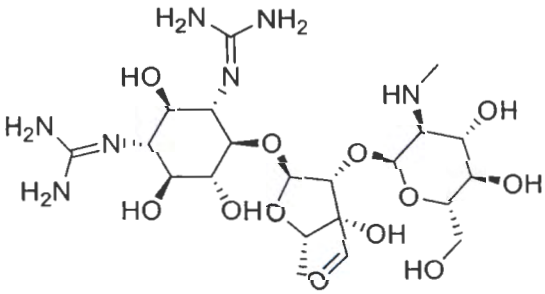
Capreomycin



Kanamycin



Amikacin



Streptomycin

Figure 1.5 Second-line anti-TB agents

1.2.3 Tuberculosis treatment regimen

The WHO guidelines for the programmatic management of drug-resistant tuberculosis classified available anti-TB drugs in five groups based on evidence of efficacy, potency, drug class and experience of use as shown in Table 1.2 (Caminero et al., 2010). Many classes of antibiotics have been approved for the treatment of TB, the long treatment regimens required for cure can result in poor patient compliance and in the rapid emergence of drug-resistant strains. That emergence has made the administration of monotherapy for TB obsolete. The WHO now recommends the Direct Observed Therapy, Short Course (DOTS) programme, a multidrug cocktail consisting of four first-line drugs (isoniazid, rifampicin, pyrazinamide and ethambutol) administered for 2 months, followed by administration of isoniazid and rifampicin for an additional 4 months under direct observation by a healthcare worker. The DOTS programme has proven to be effective for treating TB and for minimizing the emergence of more drug-resistant strains, but it still imposes major burdens on patients and healthcare workers, given the long treatment regimens and the necessity to track drug cocktails. From the treatment of TB guidelines, the standard treatment regimens for TB, DOTS and the stop TB Strategy were recommended as shown in Table 1.3.

Table 1.2 TB drug classification and daily dose

Group	Classed Used	Drug (Abbreviation)	Daily dose
First-line anti-TB drugs			
Group 1	Oral	Isoniazid (H) Rifampicin (R) Pyrazinamide (Z) Ethambutol (E)	5 mg/kg 10 mg/kg 30 mg/kg 15–25 mg/kg
Second-line anti-TB drugs			
Group 2	Injectable aminoglycosides	Streptomycin (S/Stm) Kanamycin (Km) Amikacin (Amk)	15 mg/kg 15 mg/kg 15 mg/kg
	Injectable polypeptides	Capreomycin (Cm) Viomycin (Vim)	15 mg/kg
First-line anti-TB drugs			
Group 3	Oral and injectable fluoroquinolones	Ciprofloxacin (Cfx) Levofloxacin (Lfx) Moxifloxacin (Mfx) Ofloxacin (Ofx) Gatifloxacin (Gfx)	500-750 mg 15 mg/kg 7·5–10 mg/kg 15 mg/kg 400 mg
Group 4	Oral	Para-aminosalicylic acid (Pas) Cycloserine (Dcs) Terizidone (Trd) Ethionamide (ETH) Prothionamide (Pto) Thioacetazone (Thz) Linezolid (Lzd)	150 mg/kg 15 mg/kg 15 mg/kg 15 mg/kg 15 mg/kg 150 mg 600 mg
Third-line anti-TB drugs			
Group 5		Clofazimine (Cfz) Linezolid (Lzd) Amoxicillin plus clavulanate (Amx/Clv) Imipenem plus cilastatin (Ipm/Cln) Clarithromycin (Clr)	100 mg 600 mg 875/125 mg every 12 h 500 mg/12 h

Resource: Caminero et al. (2010)

Table 1.3 WHO-recommended treatment regimens

Treatment category	Patients	Tuberculosis treatment*	
		Initial phase	Continuation phase
I	New cases of smear-positive pulmonary tuberculosis or severe extrapulmonary tuberculosis or severe smear-negative pulmonary tuberculosis or severe concomitant HIV disease	2 months $H_3R_3Z_3E_3$ or 2 months $H_3R_3Z_3S_3$ 2 months HRZE or 2 months HRZS	4 months H_3R_3 4 months HR 6 months HE
II	Previously treated smear-positive pulmonary tuberculosis; relapse; treatment failure; treatment after default	2 months $H_3R_3Z_3E_3S_3$/ 1 month $H_3R_3Z_3E_3$ 2 months HRZES/ 1 month HRZE	5 months $H_3R_3E_3$ 5 months HRE
III	New cases of smear-negative pulmonary tuberculosis or with less severe forms of extrapulmonary tuberculosis	2 months $H_3R_3Z_3E_3$ 2 months HRZE	4 months H_3R_3 4 months HR 6 months HE

*Subscript after letters refers to the number of doses per week; daily has no subscript.

Resource: WHO (2010(b))

1.2.4 Tuberculosis treatment problems

In the TB treatment process, TB patients must take drugs from 6 months to 2 years or longer. The mycobacteria that cause TB can develop resistance to the antimicrobial drugs used to cure the disease. The efficiencies of anti-TB drug were decreased based on TB drug resistant. Moreover, the other problem of TB treatment is co-infection with human immunodeficiency virus (HIV). The treatment of TB becomes more complicated.

1.2.4.1 Multidrug resistant tuberculosis (MDR-TB)

Multidrug resistant tuberculosis, MDR-TB (defined as resistance against the two most potent anti-TB drugs, namely, at least Rifampicin and Isoniazid) is a worldwide public health problem. This problem means standard first-line tuberculosis drugs failed to treat TB. Globally, an estimated 3.3% (480,000) of new cases and 20% (300,000) of previously treated cases have MDR-TB. An estimated 190,000 people died of MDR-TB. In Thailand, 2% (1,100) of new TB cases and 19% (1,100) of retreatment TB cases with MDR-TB were reported.

1.2.4.2 Extensively drug resistant tuberculosis (XDR-TB)

Extensively drug-resistant tuberculosis (XDR-TB) is defined as TB caused by *M. tuberculosis* with resistance to at least two first line TB drugs, isoniazid and rifampicin, any fluoroquinolone, and at least one of three injectable second-line aminoglycoside drugs (amikacin, capreomycin, or kanamycin) (Jassal and Bishai, 2009). The term and definition of XDR-TB were first developed by the US Centers for Disease Control and Prevention (CDC) (Raviglione and Smith, 2007). Globally reported in 2014, an estimated 9.7% of people with MDR-TB have XDR-TB (WHO, 2015(c)). In case of Thailand, 9,580 XDR-TB patients were reported in 2014. These data included 13% (4,370) of new TB cases and 38% (2,209) of retreatment TB cases.

1.2.4.3 Totally drug resistant tuberculosis (TDR-TB)

Totally drug resistant TB (TDR-TB) or extremely drug resistant TB (XXDR-TB) is a generic term for TB strains that are resistant to a wider range of TB drugs than XDR-TB strains. TDR-TB has been identified in three countries; India, Iran, and Italy. Isolated cases were reported in Italy that had resistance to all first-line anti-TB drugs and second-line anti-TB drugs.

1.2.4.4 Co-infection of TB with HIV

Patients with HIV-associated tuberculosis have an increased recurrence rate, which results from re-infection rather than relapse. In 2014, an estimated 1.2 million (12%) of the 9.6 million people who developed TB worldwide were HIV-positive.

1.2.5 Drug discovery and development for TB drugs

Development of new drugs for tuberculosis is lengthy, expensive, and risky, and the expected revenues are too small to justify commercial investment. The focus of TB drug development has always been on regimens rather than single drugs (Fox, Ellard and Mitchison, 1999; Ma et al., 2010). To achieve global control of this epidemic, there is an urgent need for new TB drugs, which can: (1) shorten treatment duration; (2) target MDR or XDR strains; (3) simplify treatment by reducing the daily pill burden; (4) lower dosing frequency (for example, a once-weekly regimen); and (5) be co-administered with HIV medications (Koul et al., 2011). Figure 1.5 shows some of the major milestones in the discovery and development of TB drugs and regimens. The identification and development of novel TB drug combinations are essential to

address the challenges associated with the present treatments for tuberculosis (Ginsberg and Spigelman, 2007). An ideal drug combination should consist of at least three TB drugs that are active against MDR and XDR tuberculosis, and have potent, synergistic, and complementary activities against various types of *M. tuberculosis*.

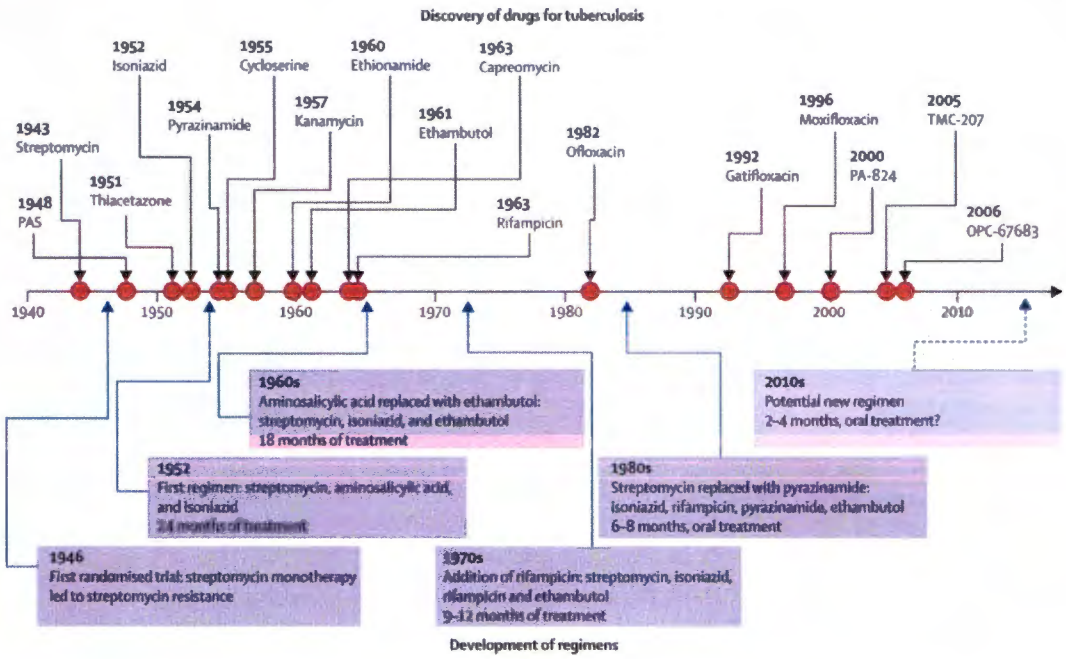


Figure 1.6 History of drug discovery and development of treatment regimens for tuberculosis

Resources: Ma et al. (2010)

The "ideal" drug targets are considered. TB drug target that essential for growth and/or survival under the various conditions encountered during *M. tuberculosis* infection in humans (Koul et al., 2011; Ioerger et al., 2013; Xiong et al., 2013; Zumla, Nahid and Cole, 2013; Kana et al., 2014). Moreover, the target should be vulnerable to inhibition under these conditions. TB drug targets that have a low tolerance for mutation would limit the emergence of drug-resistant variants. These targets would serve as ideal chokepoints, inhibition of which would lead to rapid cell death due to the buildup of toxic metabolites. A drug target that is intrinsically essential for the bacteria in both active and latent disease is extremely desirable and would allow for the treatment of both of these clinical presentations. Furthermore, those drug targets or metabolic pathways that are required for prolonged survival of *M.*

tuberculosis during latent TB infection (LTBI) may be similar to those required for persistence of tubercle bacilli in the face of prolonged TB chemotherapy (Boshoff and Barry, 2006). The GC-rich (65.6%) 4.4-Mbp genome of *M. tuberculosis* is one of the biggest among the bacteria and encodes about 4,000 predicted proteins. The important enzymes that important in *M. tuberculosis* were validated as potential drug targets to develop new TB drugs as summarized in Table 1.4 (Xiong et al., 2013).

Table 1.4 Summary of validated drug targets in *M. tuberculosis*

Mechanism/Pathway	Target	Rv number	Human Homologue	Drug/ Inhibitor*
Cell wall biosynthesis				
Peptidoglycan biosynthesis	AlR	Rv3423c	None	D-cycloserine
	Ddl	Rv2981c	None	D-cycloserine
	DapA	Rv2753c	None	
	DapB	Rv2773c	None	
	LdtA	Rv0116c	None	
	LdtB	Rv2518c	None	
Arabinogalactan biosynthesis	DprE1	Rv3790	None	
	DprE2	Rv3791	None	
	AftA	Rv3792	None	
	AftB	Rv3805c	None	
	RmlC	Rv3465	None	
	RmlD	Rv3266c	None	
	EmbA	Rv3794	None	
	EmbB	Rv3795	None	
Mycolic acid biosynthesis				OPC-67683, PA-824
	KasA	Rv2245	None	
	KasB	Rv2246	None	
	FabH	Rv0533c	None	
	FabG1/MabA	Rv1483	None	
	InhA	Rv1484	None	Isoniazid, Ethionamide
	PcaA	Rv0470c	None	
	CmaA1	Rv3392c	None	
	CmaA2	Rv0503c	None	

* The listed drugs or inhibitors are launched or in clinical studies

Resource: Xiong et al. (2013)

Table 1.4 Summary of validated drug targets in *M. tuberculosis* (continued)

Mechanism/Pathway	Target	Rv number	Human Homologue	Drug/ Inhibitor*
	MmpL3	Rv0206c	None	SQ109
	AccD4	Rv3799c	None	
	Pks13	Rv3800c	None	
	FadD32	Rv3801c	None	
	FadD13	Rv3089	None	
	EmbC	Rv3793	None	Ethambutol
	GpgS	Rv1208	None	
	GlmU	Rv1018c	None	
RNA Transcription / DNA Replication	RNA Polymerase		None	Rifampicin, Rifabutin, Rifapentine
	GyrA	Rv0006	None	Moxifloxacin, Ofloxacin, Levofloxacin, Ciprofloxacin
	GyrB	Rv0005	None	
	DnaB	Rv0058	None	
	CarD	Rv3583c	None	
	NrdF2	Rv3048c	Yes	
	ThyX	Rv2754c	None	
	LigA	Rv3014c	None	
Amino acid / Protein biosynthesis	Glutamine Synthetase	Rv2220	Yes	
	HisG	Rv2121c	None	
	TrpC	Rv1611	None	
	TrpD	Rv2192c	None	
	ArgC	Rv1652	None	
	IlvB1	Rv3003c	None	
	IlvB2	Rv3470c	None	
	MapA	Rv0734	Yes	
	MapB	Rv2861c	Yes	
	Def	Rv0429c	Yes	
	Chorismate Mutase	Rv1885c	None	

* The listed drugs or inhibitors are launched or in clinical studies

Resource: Xiong et al. (2013)

Table 1.4 Summary of validated drug targets in *M. tuberculosis* (continued)

Mechanism/Pathway	Target	Rv number	Human Homologue	Drug/ Inhibitor*
	Ribosome		Yes	amikacin, kanamycin, pyrazinamide, capreomycin, streptomycin, linezolid, AZD-5847, PNU-100480
Cofactor biosynthesis				
Biotin/Vitamin H	BioA	Rv1568	None	
Pantothenate/Vitamin B5	PanC	Rv3602c	None	
Riboflavin/Vitamin B2	RibH	Rv1416	None	
Menaquinone/Vitamin K2	MenA	Rv0534c	None	
	MenB	Rv0548c	None	
Coenzyme A	CoaA	Rv1092c	None	
Folate	DfrA	Rv2763c	Yes	
	FolB	Rv3607c	None	
MEP Pathway	DXS	Rv2682	None	
	DXR/IspC	Rv2870c	None	
	IspD	Rv3582c	None	
	IspE	Rv1011	None	
	IspF	Rv3581c	None	
Shikimate pathway	AroG	Rv2178c	None	
	AroD	Rv2537c	None	
	AroE	Rv2552c	None	
	AroK	Rv2539c	None	
	AroF	Rv2540c	None	
ATP synthesis	ATP Synthase		Yes	
Signal transduction	DosS/ DevS	Rv3132c	None	
	DosR/DevR	Rv3133c	None	
	PtpB	Rv0153c	Yes	
	PknB	Rv0014c	Yes	
	PknG	Rv0410c	Yes	
Cytochrome P450 systems	CYP51	Rv0764c	Yes	
	CYP121	Rv2276	Yes	

* The listed drugs or inhibitors are launched or in clinical studies

Resource: Xiong et al. (2013)

Table 1.4 Summary of validated drug targets in *M. tuberculosis* (continued)

Mechanism/Pathway	Target	Rv number	Human Homologue	Drug/ Inhibitor*
Mycothiol biosynthesis	MshA	Rv0486	None	
	MshC	Rv2130c	None	
Other important targets	MbtI	Rv2386c	None	
	BlaC	Rv2068c	None	
	CysH/APSR	Rv2392	None	
	CanB	Rv3588c	None	
	GlgE	Rv1327c	None	
	Isocitrate Lyase	Rv0467	None	
	AhpC	Rv2428	None	
	AhpD	Rv2429	None	
	ClpP1	Rv2461	Yes	
	ClpP2	Rv2460c	Yes	
	FtsZ	Rv2150c	None	
	Rel		None	

* The listed drugs or inhibitors are launched or in clinical studies.

Resource: Xiong et al. (2013)

1.2.6 Enoyl-ACP reductase (InhA)

1.2.6.1 Mechanism of InhA

There are two types of fatty acid synthetase systems present in *M. tuberculosis* and other mycobacteria, named type I and type II fatty acid synthase (FAS) systems (Figure 1.6). Both systems are indispensable for synthesis of mycolic acids, the mycobacterium cell wall (Takayama, Wang and Besra, 2005). The product of FAS-I, acyl-coenzyme A (CoA) (C16–C26), is extended by FAS-II to form the precursor of mycolic acids (C60–C90) (Barry et al., 1998). Enoyl-ACP reductase, InhA of *M. tuberculosis* catalyzes the NADH-specific reduction of 2-*trans*-enoyl-ACP in the elongation cycle of the FAS II pathway as shown the mechanism in Figure 1.6. Mycolic acids are very long chain (C74–C90) a-alkyl b-hydroxy fatty acids covalently linked to arabino-galactan. C16 substrate, a product of FAS I, is utilized as a starting point for further elongation by FAS II cycle. Therefore, InhA is specific for substrate with aliphatic chains C16 and longer. Mycobacterial InhA molecule is a homotetramer in aqueous solution, and its crystal structure in complex with NAD^+ and a fatty acyl

substrate reveals that the substrate binds in a general U-shaped conformation (Rozwarski et al. 1999(b)). The residues crucial for *trans*-enoyl reduction catalysis, Tyr158 and Lys165, are situated at the active site. Tyr158 and Lys165 residues promote deprotonation of the 20-nicotinamide hydroxyl group. InhA facilitates reduction by a catalytic mechanism in which a hydride is transferred to the C3 carbon of the C2–C3 double bond. The C1 carbonyl oxygen accepts a proton from the Tyr 158 hydroxyl, which forms an enolate anion (Parikh et al., 1999). The fatty acyl chain of the bound substrate is completely surrounded by hydrophobic residues, the majority of which are located on the substrate binding loop (Met103, Phe149, Tyr158, Lys165, Thr196, Met199, Leu207 and Ile215). Tyr158 is conserved in the bacterial FabIs and plant enoyl-ACP reductases; interaction with Tyr158 is likely to be a key feature of fatty acyl substrate binding common to other enoyl-ACP reductases (Parikh et al., 1999). The impaired action of InhA therefore leads to the loss cell integrity and consequently to the cell death. Since InhA plays a crucial role in the fatty acids elongation cycle it was validated as a promising target of novel antitubercular compounds (Lu et al., 2011; Tonge, Kisker and Slayde, 2007).

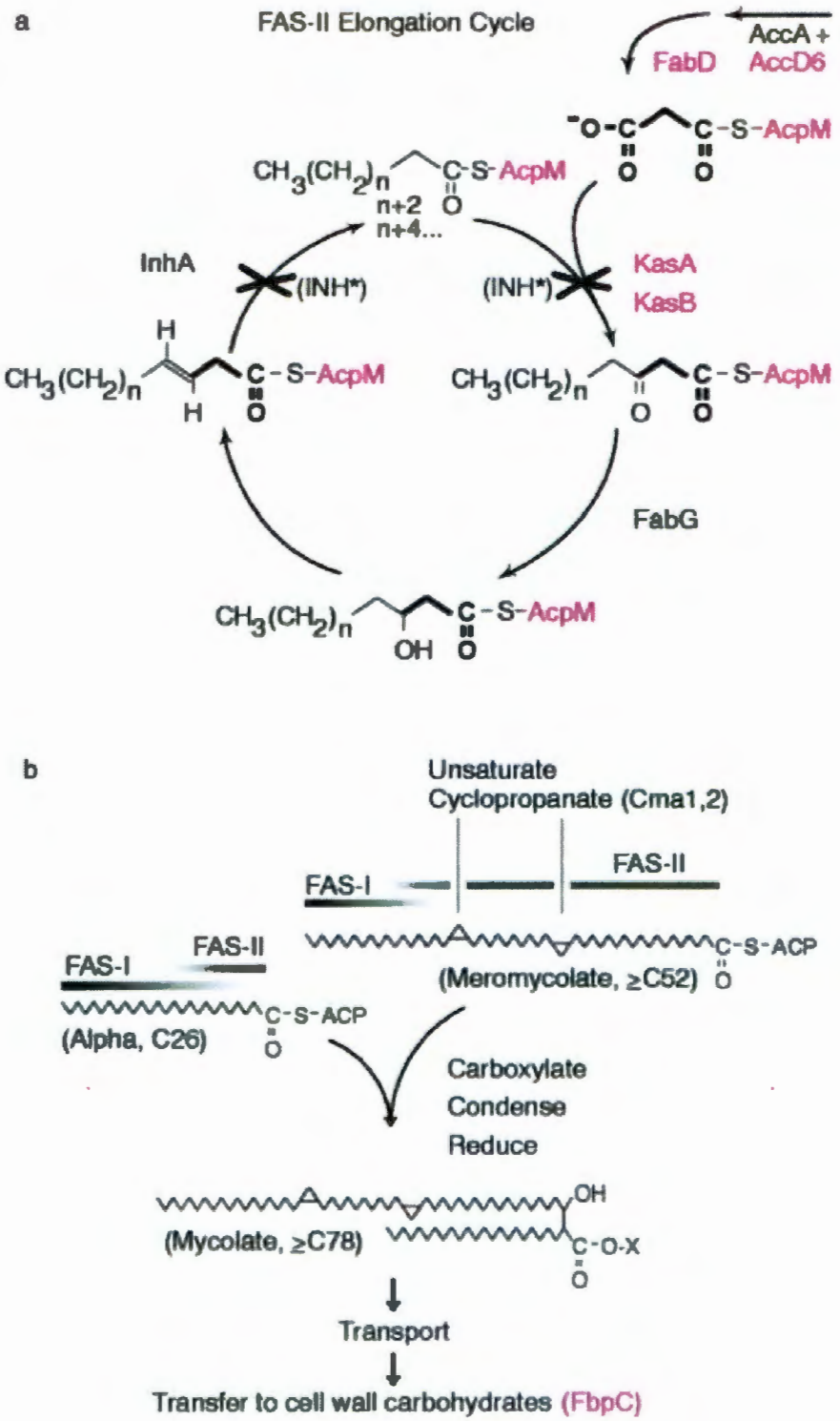


Figure 1.7 Schematic diagrams of the biosynthetic pathways involved in lipid metabolism in *M. tuberculosis*

Resource: Wilson et al. (1999)

1.2.6.2 Indirect InhA inhibitors

There are two available anti-TB drugs that reported as indirect InhA inhibitors. Isoniazid (INH) (Figure 1.7) is a well-known as first line anti-TB drug. This drug is pro-drug. It has been used in clinical treatment since 1952. The second-line anti-TB drug, ethionamide (ETH) as shown the chemical structure in Figure 1.7 (Dessen, et al., 1995) was also reported as the indirect InhA inhibitors. Although ETH is the analogue of INH, the activation pathways for these two drugs are different. INH is activated by catalase peroxidase (KatG) to generate the reactive acyl radical of INH. Then, reactive radical covalently formed between NAD^+ to generate the INH-NAD adduct, the active form of INH (Zhang, et al., 1992, Rozwarski, et al., 1998(a)) as shown in Figure 1.7. On the other hand, the flavin-dependent monooxygenase enzyme, ethA-encoded mono-oxygenase activates ETH to yield the ETH-NAD adduct (DeBarber, et al., 2000, Vannelli, 2002). These two adducts subsequently inhibit InhA and NADH-dependent enoyl-ACP reductase effectively to block the mycolic acid biosynthesis and leading to cell lysis (Vilcheze et al., 2005; Kumar and Shaik, 2011). The study of *M. tuberculosis* clinical isolates carrying ethA, inhA, and katG mutations indicated that katG mutation cannot change the level of ETH resistance, but do confer INH resistance, particularly mutation at codon 315. In contrast, ethA and inhA structural genes mutations were relevant to relatively high levels of ETH resistance (Morlock, et al., 2003). Dominant mutations in InhA or recessive mutations in ndh, which encode a type II NADH dehydrogenase (NdhII), could mediate the cross-resistance to INH and ETH. The M. smegmatis strain with the ndh mutation had high-level resistance to INH and ETH, while the mutants in M. bovis BCG showed a lower resistance to INH and ETH. The mechanism of resistance was mediated by increasing the NADH cellular concentration and then competitively inhibiting INH-NAD or ETH-NAD adduct binding of InhA (Vilcheze, et al., 2005)

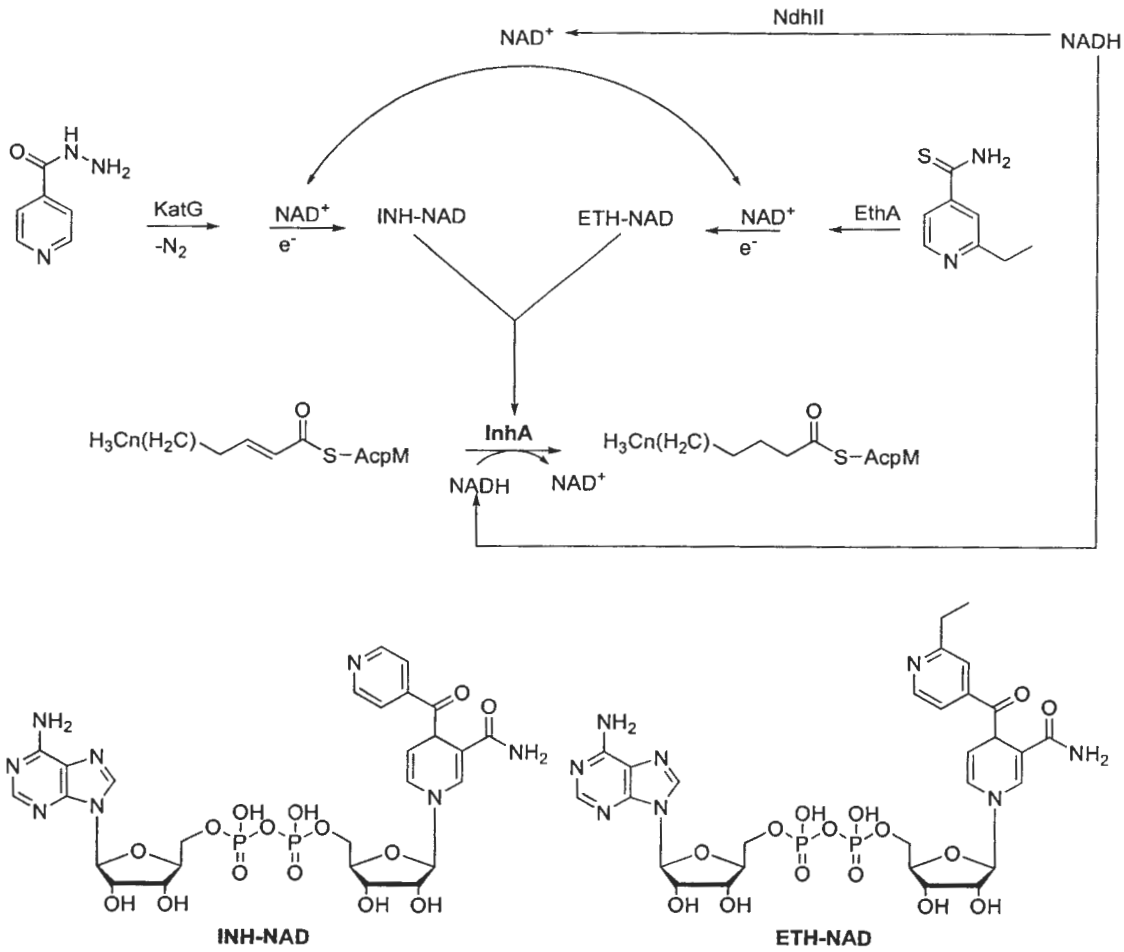


Figure 1.8 Formation of INH-NAD and ETH-NAD adducts

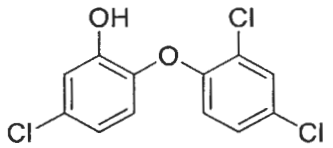
Resource: Vilcheze et al. (2005)

1.2.6.3 Direct InhA inhibitors

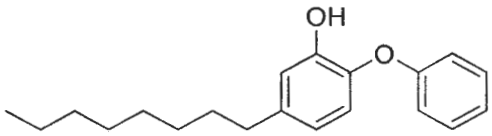
The specificity is determined by a loop of the binding region of InhA, called the substrate-binding loop, which has been shown to be flexible (Rozwarski et al., 1999(b); Kuo et al., 2003). Superposition of the crystal structure of ecFabI with InhA demonstrates that there is a significant difference between these two enzymes with respect to the location of their substrate-binding loops. In InhA, the loop (residues 194 - 220) creates a substance-binding crevice (18 Å) with more depth than loop (residues 192 - 209) of ecFabI (10 Å). The intrinsic specificity observed in the substrate-binding loop is consistent with the size and shape of the conserved hydrophobic pocket adjacent to the active site of InhA (Lu, Huang and You, 2011).

1) InhA inhibitors with Tyr158-in conformation

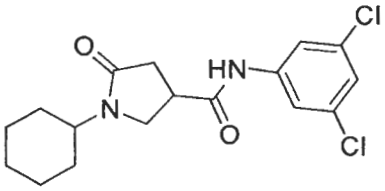
Triclosan was reported as the first direct InhA inhibitors at the acyl substrate-binding pocket (Parikh, Xiao and Tonge, 2000). This structure was modified to obtained potent triclosan inhibitors (Freundlich et al., 2009; Stec et al., 2014). The first generation of alkyl substituted diphenyl ethers (structure of most active compound depicted at Figure 1.8) was prepared to improve affinity towards InhA (Sullivan et al., 2006). The biological activity results against InhA and *M. tuberculosis* suggest that 5-octyl-2- phenoxyphenol (Figure 1.8) is the most potent compound with a K_i value of 1.1nM for InhA. Then, this series were continuously modified (am Ende et al, 2008; Pan et al., 2014). Pyrrolidine carboxamide derivatives were discovered through high-throughput screening co-workers also performed a high throughput screening of a library of 30,000 compounds and were then structurally modified (He et al., 2006). Modification of various substitutions at phenyl ring of pyrrolidine carboxamides resulted in compounds with significant activity. Unfortunately, pyrrolidine carboxamide classes with good enzyme inhibitory do not exhibit ideal activity against *M. tuberculosis* H37Rv with MIC above 125 μM. The results suggest that pyrrolidine carboxamides may exhibit poor membrane permeability. Similar high-throughput experimental design published by He and co-workers (He et al., 2006) led to arylamide derivertives a novel direct InhA inhibitors (He et al., 2007). Based on the structures modifications, aryl amide derivatives display high InhA inhibitory activity but poor membrane permeability or are easily extruded by efflux pumps (He et al., 2007; Chollet et al., 2015). The natural product, pyridomycin was reported as InhA inhibitors in 2012 (Hartkoorn et al., 2012). 2-(4-Oxoquinazolin-3(4*H*)-yl)acetamide derivatives and benzo[d]oxazol-2(3*H*)-one derivatives were identified from virtual screening followed by biological evaluation (Kumar et al., 2013; Pedgaonkar et al, 2014(a); 2014(b)). Cell-based high-throughput screen of 2.2 million compounds using *M. tuberculosis* H37Ra resulted in ~8000 hits were reported. 4-Hydroxy-2-pyridone presented as an attractive hit series based on the in vitro pharmacokinetic properties and moderate potency against *M. tuberculosis* (MIC_{50} = 1.54 μM) (Manjunatha and Smith, 2015; Ng et al., 2015). Rhodanine derivatives were developed as InhA inhibitors (Slepikas et al., 2016).



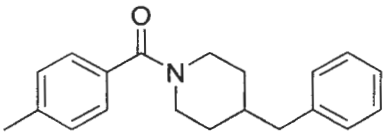
Triclosan



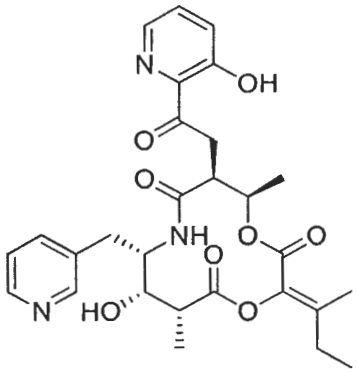
Diphenyl ether



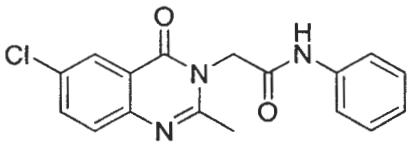
Pyrrolidine carboxamide



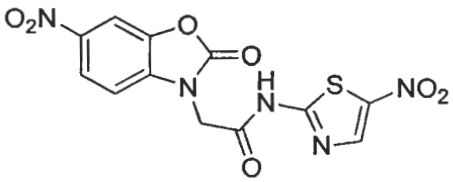
Arylamide



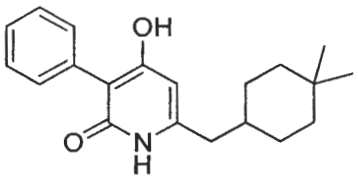
Pyridomycin



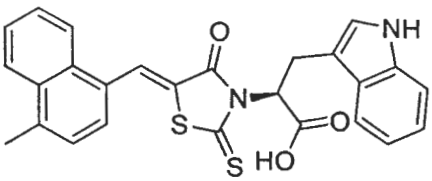
2-(4-Oxoquinazolin-3(4*H*)-yl)acetamide



Benzo[d]oxazol-2(3*H*)-one



4-Hydroxy-2-pyridone

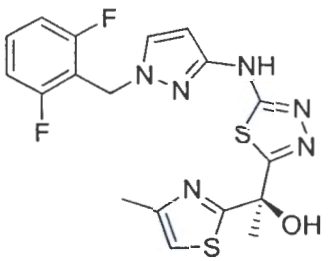


Rhodanine

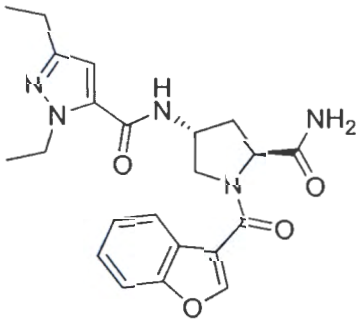
Figure 1.9 Structure of Y158-in InhA inhibitors

2) InhA inhibitors with Tyr158-out conformation

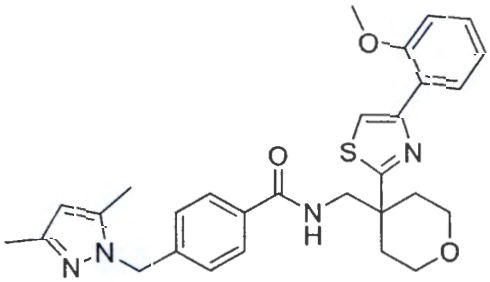
Methyl triazole and it derivatives were reported as direct InhA inhibitors with a novel mechanism of InhA inhibition, unreported "Y158-out" inhibitor-bound conformation (Shirude et al., 2013; Sink et al., 2015). Benzofuran pyrrolidine pyrazole derivatives were found as Y158-out InhA inhibitors in 2014. These compounds were identified based on the Encoded Library Technology (Encinas et al, 2014). A tetrahydropyran scaffold as Y158-out InhA inhibitors was identified by high-throughput screening against InhA with the GlaxoSmithKline compound collection. The modification compounds in these showed good biological activity with improved antimycobacterial activity and low cytotoxicity (Pajk et al., 2016). *N*-Benzyl-4-((heteroaryl)methyl)benzamide derivatives were identified by high-throughput screening against InhA (Guardia, et al., 2016).



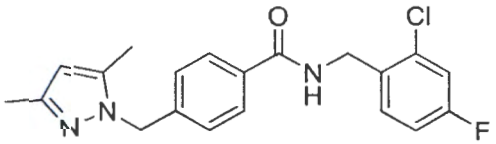
Methyl triazole



Benzofuran pyrrolidine pyrazole



Tetrahydropyran



N-Benzyl-4-((heteroaryl)methyl)benzamide

Figure 1.10 Structure of Y158-out InhA inhibitors

1.2.7 Serine/Theonine Kinase G (PknG)

1.2.7.1 Mechanism of PknG

Due to the success of kinase inhibitors in the treatment of non-infectious human diseases, *M. tuberculosis* signal transduction pathway has become a prime target for the development of novel TB therapeutics. In *M. tuberculosis*, signal transduction is co-mediated by five main families of kinases and phosphatases. The first class is a histidine kinase and a response regulator. The second encodes the "eukaryotic-like" serine/threonine protein kinases (STPKs). The third contains a sole Ser/Thr phosphatase belonging to the protein phosphatase type 2C (PP2C) family. The fourth class contains a pair of protein tyrosine phosphatases (PTPs); and fifth, *M. tuberculosis* protein tyrosine kinase (PTK) (Chao et al., 2010). STPKs have been shown to be important virulence factors in various pathogenic bacteria (Cozzone, 2005). The *M. tuberculosis* genome contains eleven STPKs (namely PknA-PknL, excluding PknC) two of which are soluble proteins (PknG and PknK) and the rest are transmembrane kinases (Cole et al., 1998; Av-Gay and Everett, 2000). The Diagrammatic representation of the domain organization of serine/threonine protein kinases found in *M. tuberculosis* displays in Figure 1.10. Phylogenetic analysis based on neighbor-joining method revealed clustering of mycobacterial STPKs into five different clades as shown in Table 1.5 (Narayan et al., 2007). Nine of these kinases were predicted to be localized to the cell membrane due to the presence of a putative transmembrane domain as shown in Figure 1.10.

Table 1.5 Different clades of mycobacterial STPKs

Clade	Serine/threonine protein kinases	Description of clade
I	PknA, PknB, PknL, ABL group	The group of transmembrane sensor kinases from different mycobacterial species
I	PknH, PknE, PknD, HED group	The group of integral membrane receptor as well as cytoplasmic kinases
III	PknF PknI, PknJ, FIJ group	The group of kinases that cluster along with PknF, PknI and PknJ
IV	PknK	The soluble kinase that cluster along with PknK
V	PknG	The soluble kinases that cluster along with PknG

Resource: Narayan et al. (2007)

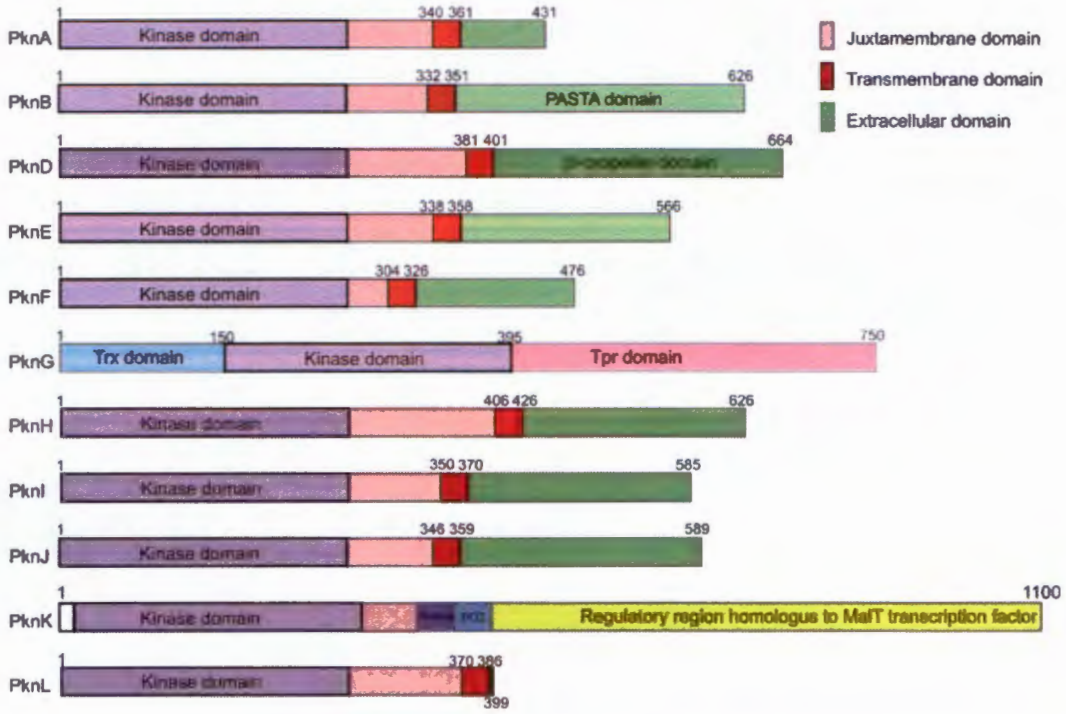


Figure 1.11 Schematic representation of *M. tuberculosis* STPKs

Resource: Chakraborti et al. (2011)

Serine/threonine protein kinase G (PknG) is one of two soluble STPKs (PknG and PknK) in *M. tuberculosis*, containing an *N*-terminal rubredoxin domain (Scherr et al., 2007), a central kinase domain, and a C-terminal tetratricopeptide repeat (TPR) domain (Av-Gay and Everet, 2000; Scherr et al., 2007). PknG is classified as a non-RD kinase. This enzyme has been shown to play a role in the survival of the pathogen in host macrophages by modulating phagosome-lysosome fusion (lysosomal transfer) after macrophages phagocytose mycobacterium as shown in Figure 1.11 (Walburger et al., 2004; Nguyen and Pieters, 2005; Warner and Mizrahi, 2007; Chakraborti et al., 2011). Most microbes and nonpathogenic mycobacteria quickly find themselves in lysosomes in the host cell, where they are killed. By contrast, *M. tuberculosis* stays within phagosomes; the bacterium releases PknG to block phagosome-lysosome fusion. More importantly, identification and validation of this protein as novel drug targets for TB therapeutics have been explored for screening and designing specific inhibitors against *M. tuberculosis* PknG (Av-Gay and Everett, 2000; Alber, 2009).

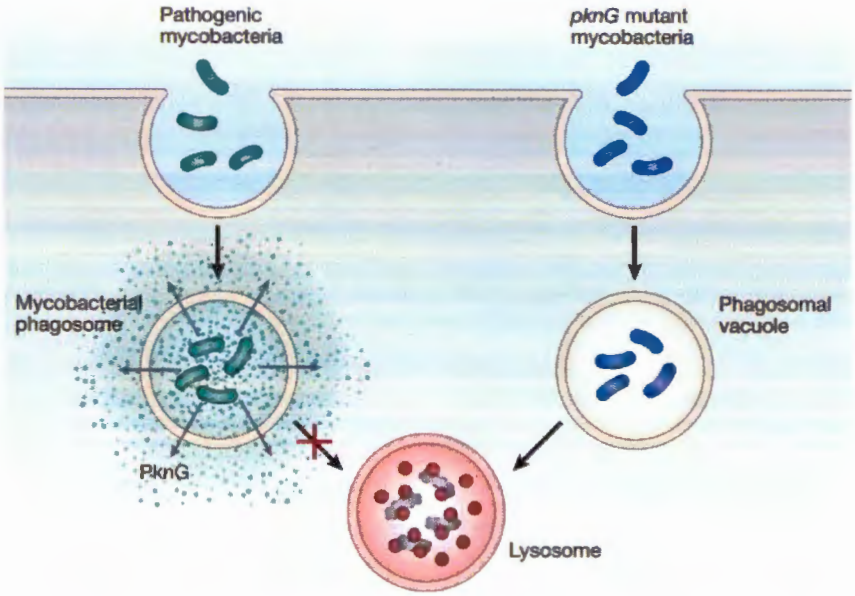
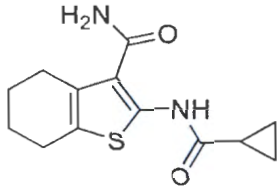


Figure 1.12 PknG affects the intracellular traffic of *M. tuberculosis* in macrophages

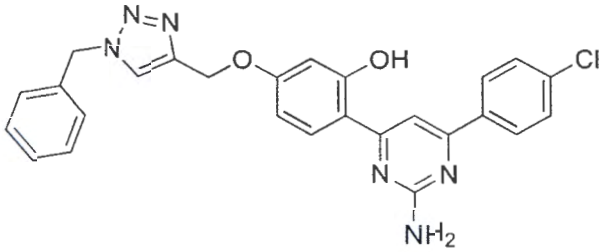
Resource: Warner and Mizrahi (2007)

1.2.7.2 PknG inhibitors

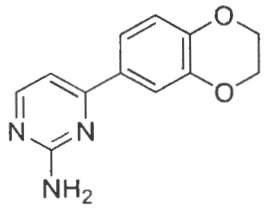
PknG is attractive target for anti-TB drugs research thought inhibitor screening. Using combined screening and medicinal chemistry strategies, AX20017, a tetrahydrobenzothiophene compound was identified as the first inhibitor that specifically inhibits PknG kinase activity in vitro (Walburger et al., 2004). This inhibitor showed low inhibitory activity in other bacterial serine/threonine kinases and human kinases (Walburger et al., 2004; Scherr et al., 2007). The Nested Chemical Library™ (NCL) technology and pharmacophore modelling were applied to find the hit PknG inhibitors and structural modification of tetrahydrobenzothiophene derivatives (Hegymegi-Barakonyi et al., 2007; Banhegyi, 2008; Székely et al., 2008; Sipos et al., 2015). Triazolylmethoxy 2-aminopyrimidine derivatives were designed, synthesized and inhibitory activity against PknG evaluated. The obtained results showed that this series displayed moderate activity against PknG (53 ± 0.61 % inhibition at 100 μm) (Anand et al., 2012). The compound FM00174 that identified as PknG inhibitors from virtual screening method showed 78 % inhibition respectively against PknG at 25μM (Singh et al., 2015).



Tetrahydrobenzothiophene



Triazolylmethoxy 2-aminopyrimidine



FM 00174

Figure 1.13 Structure of PknG inhibitors

1.3 Cancer

Cancer (a noncommunicable disease) is a group of diseases characterized by the uncontrolled growth (division beyond the normal limits), invasion (intrusion on and destruction of adjacent tissues), and sometimes metastasis (spread to other locations in the body via lymph or blood) of abnormal cells. It is characterized by alterations in the expression of multiple genes, leading to dysregulation of the normal cellular program for cell division and cell differentiation. Cancers are usually classified as simple (or benign) and malignant. Benign cancers tend to remain localized, are often surrounded by a capsule and rarely give rise to serious effects. Cancers are classified according to the embryological origin of the tissues (Table 1.6) (Calman and Paul, 1978; King and Robins, 2006; Ruddon, 2007).

Table 1.6 Classification of the commoner types of cancers

Tissue	Normal cells involved	Benign cancer	Malignant cancer
Connective tissue and muscle	Fibrocyte Fat cell Osteocyte Muscle Muscle, smooth Muscle, striped	Fibroma Myxoma Lipoma Osteoma Myoma Leiomyoma Rhabdomyoma	Fibrosarcoma Myxosarcoma Liposarcoma Osteosarcoma Myosarcoma Leiomyosarcoma Rhabdomyosarcoma
Vascular endothelium		Haemangioma Lymphangioma	Haem angiosarcoma
Epithelium	Squamous and transitional Glandular	Papilloma Adenoma	Carcinoma (squamous, basal-cell) Glandular carcinoma
Neural	Glial Nerve Ganglionic Melanoblast	Glioma neuroma Melanoma	Sympathicoblastoma (neuroblastoma) Malignant melanoma Ocular melanoma
Haemopoietic	Reticulum cell Plasma cell Leucocytes	Lymphoma	Lymphosarcoma Reticulosarcoma Myeloma Leukaemia
Embryonal		Teratoma	Teratocarcinoma Chorioncarcinoma

Resource: Calman and Paul (1978)

1.3.1 Principles of cell cycle and cancer

The cell cycle, the process by which cells progress and divide, lies at the heart of cancer. In normal cells, the cell cycle is controlled by a complex series of signaling pathways by which a cell grows, replicates its DNA and divides. Cell division consists of two consecutive processes, mainly characterized by DNA replication and segregation of replicated chromosomes into two separate cells. Originally, cell division was divided into two stages: mitosis (M), i.e. the process of

nuclear division; and interphase, the interlude between two M phases (Figure 1.13). Stages of mitosis include prophase, metaphase, anaphase and telophase. Under the microscope, interphase cells simply grow in size, but different techniques revealed that the interphase includes G_1, S and G_2 phases (Norbury and Nurse 1992). Replication of DNA occurs in a specific part of the interphase called S phase. S phase is preceded by a gap called G_1 during which the cell is preparing for DNA synthesis and is followed by a gap called G_2 during which the cell prepares for mitosis. G_1, S, G_2 and M phases are the traditional subdivisions of the standard cell cycle (Figure 1.13). Cells in G_1 can, before commitment to DNA replication, enter a resting state called G_0. Cells in G_0 account for the major part of the non-growing, non-proliferating cells in the human body (Vermeulen, Bockstaele and Berneman, 2003). In cancer, as a result of genetic mutations, this regulatory process malfunctions, resulting in uncontrolled cell proliferation.

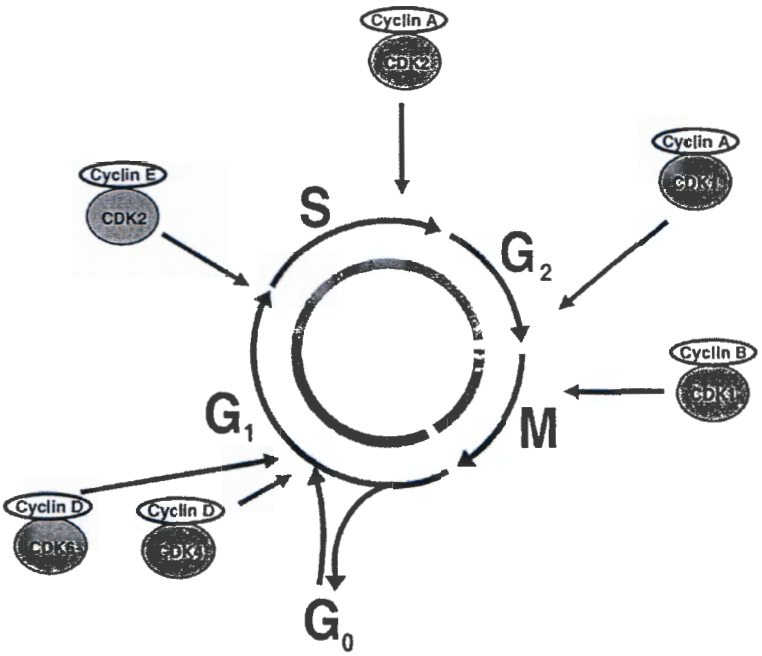


Figure 1.14 Simplified model of the mammalian cell cycle

Resource: Vermeulen Van Bockstaele and Berneman (2003)

1.3.2 Steps of cancer development

Cancer development and progression is a complex process that involves a host of functional and genetic abnormalities. This can include epigenetic modifications as well as the development of genomic mutations and other insults that can lead to altered gene expression and overall cell function. Generally, cancer develops in 3 main steps, initiation, progression and metastatic, respectively. It can take a long time for cancer to develop because several steps and several genetic mutations are usually required. Inquiries into the molecular mechanisms behind malignant transformation and metastatic progression, is the basis for the development of many new diagnostic and therapeutic strategies.

The first cell to exhibit growth disinhibition has entered a process known as tumor initiation. This initial change may be caused by carcinogens, such as chemicals, smoking or exposure to radiation, but often the cause is unknown and may be a random. Initiation of malignant transformation of normal cells by a carcinogenic agent involves a permanent, heritable change in the gene expression of the transformed cell. Tumor-initiating agents most likely act by interacting with DNA to induce mutations, gene rearrangements, or gene amplification events that produce a genotypically altered cell (Ruddon, 2007). The initiation and progression of tumors can either involve loss of tumor suppressor function or induction of oncogene function. The development of cancer exhibits several noteworthy phenomena. The first obvious behavior is the lack of normal constraint on cell proliferation. Cancer cells do not exhibit normal contact inhibition, in which cells proliferate until they reach a finite density, determined in part by the availability of certain growth factors. Transformed cells are often noted to survive in the absence of the growth factors that are normally required by their untransformed ancestors. This failure to undergo apoptosis during a state of deprivation has been postulated to contribute to the growth and survival of metastatic cells in ectopic sites. Instead of responding to the signals that cause normal cells to cease proliferation and enter the Go phase of the cell cycle, cancer cells continue to grow beyond the normal density limit. New phenotypes which portend lower rates of apoptosis, faster rates of division, lower metabolic requirements, increased ability to recruit neo-vasculature, and metastatic competency gain a selection advantage and will ultimately assume a more dominant proportion of the tumor burden. This process of

clonal selection continues as the disease progresses. As cancer cells divide, they can invade surrounding tissue. Figure 1.14 is a summary of cancer development and progression.

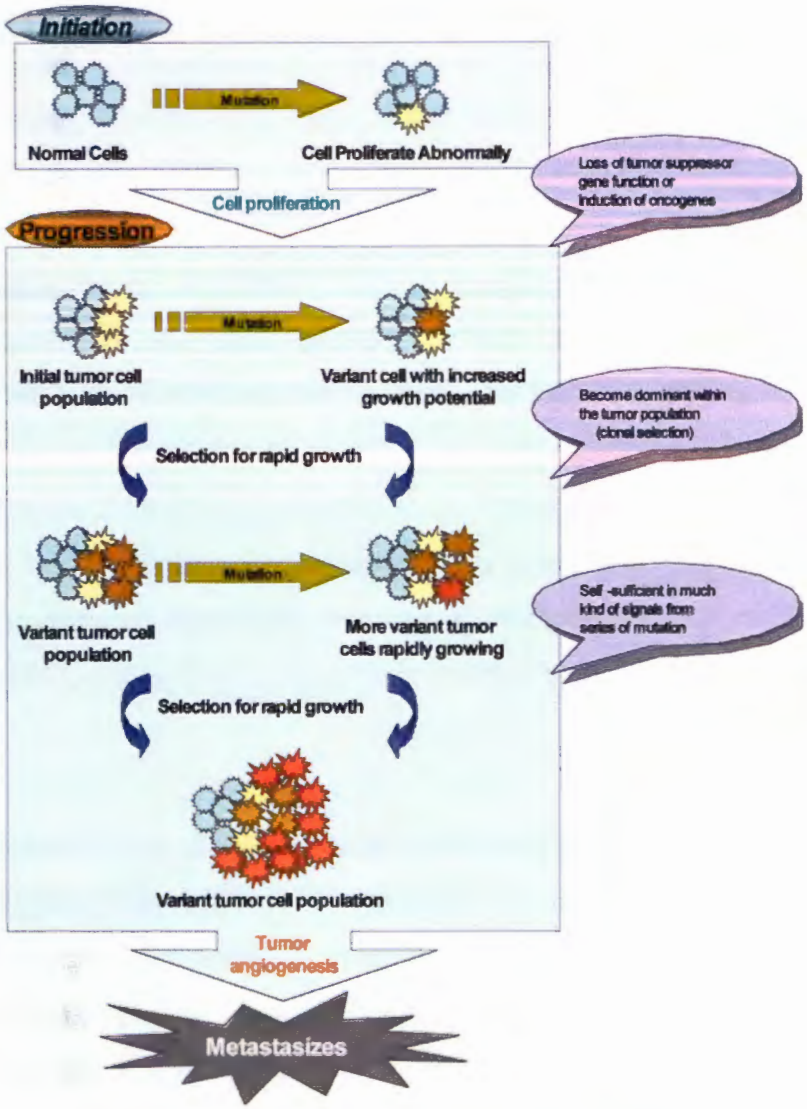


Figure 1.15 Cancer development and progression

Resource: Collino (2011)

The aims of the classification system developed for cancer staging are (1) to aid oncologists in planning treatment; (2) to provide categories for estimating prognosis and evaluating results of treatment; and (3) to facilitate exchange of

information (Rubin, 1973). The staging categories listed below represent a useful generalization (Ruddon, 2007).

Stage I: Primary tumor is limited to the organ of origin. There is no evidence of nodal or vascular spread. The tumor can usually be removed by surgical resection. Long-term survival is from 70% to 90%.

Stage II : Primary tumor has spread into surrounding tissue and lymph nodes immediately draining the area of the tumor (''first-station'' lymph nodes). The tumor is operable, but because of local spread, it may not be completely resectable. Survival is 45% to 55%.

Stage III: Primary tumor is large, with fixation to deeper structures. First-station lymph nodes are involved; they may be more than 3 cm in diameter and fixed to underlying tissues. The tumor is not usually resectable, and part of the tumor mass is left behind. Survival is 15% to 25%.

Stage IV: Extensive primary tumor (may be more than 10 cm in diameter) is present. It has invaded underlying or surrounding tissues. Extensive lymph node involvement has occurred, and there is evidence of distant metastases beyond the tissue of origin of the primary tumor. Survival is under 5% (Ruddon, 2007).

1.3.3 Cancer situations

From the world cancer statistical report in 2012, there were 14.6 million (14,067,900) new cancer cases and 8.2 million (8,201,600) cancer deaths. The 5 most frequent cancers that ranked from total number of cases are lung (13.0%), breast (11.9%), colorectum (9.7%), prostate (7.8%), and stomach (6.8%) cancers, respectively. The most frequent cancer in men is lung cancer (16.8%, 1,241,601 from 7,410,376 incidence cases). For women, the most frequent cancer is breast (25.1%, 1,671,149 from 6,657,518 incidence cases). In Thailand, there were 62,764 new cancer cases and 48,100 cancer deaths in 2012. The 5 most frequent cancers that ranked from total number of cases are liver (16.5%), lung (15.8%), breast (11.0%), colorectum (9.3%) and cervix uteri (6.6%) cancer respectively. The most frequent cancer in men is liver cancer (22.5%, 14,739 from 62,764 incidence cases). For women, the most frequent cancer is breast (22.4%, 13,653 from 61,037 incidence cases) (GLOBOCAN2012, 2012).

1.3.4 Cancer treatments

Cancer therapy is based on surgery and radiotherapy, which are, when possible, rather successful regional interventions, and on systemic chemotherapy. Approximately 50% of cancer patients are not cured by systemic chemotherapy and obtain only a prolonged survival. Surgery is a common cancer treatment option. Radiotherapy also known as radiation cancer treatment can also be used to treat many forms of cancer. However, the side effects of radiotherapy are different for each patient and depend on the type of cancer, location, doses, and your patient health. Approximately half of cancer patients are not cured by these treatments and may obtain only a prolonged survival or no benefit at all (Avendano and Menendez, 2007(a)). Photodynamic therapy (PDT) is cancer treatment method minimally invasive therapeutic procedure that can exert a selective cytotoxic activity toward malignant cells. The procedure of PDT process involves administration of a photosensitizing agent followed by irradiation at a wavelength corresponding to an absorbance band of the sensitizer. In the presence of oxygen, a series of events lead to direct tumor cell death, damage to the microvasculature, and induction of a local inflammatory reaction (Pass, 1993; Dougherty et al., 1998; Dolmans, Fukumura and Jain, 2003; Brown, Brown and Walker, 2004; Castano, Demidova and Hamblin, 2004; Hopper, 2000; Agostinis et al., 2011). PDT is suitable to treat solid tumours (bladder cancer, lung cancer and in malignant diseases of the skin and upper aerodigestive tract) (Schuitmaker et al., 1996). Activating the immune system for cancer immunotherapy therapeutic benefit in cancer has long been a goal in immunology and oncology (Rosenberg, Yang and Restifo, 2004; Mellman, Coukos and Dranoff, 2011).

1.3.5 Cancer drugs

The aim of modern cancer chemotherapy of most chemotherapeutic drugs in clinical used is to kill malignant tumor cells by inhibiting some of the mechanisms implied in cellular division. Anti-cancer drugs were classified based on the role of drug mechanisms of action as shown in Table 1.7 (Avendano and Menendez, 2007(a); 2015(b)).

Table 1.7 Classifications of anti-cancer drugs

Drug classes	Description
1	Antimetabolites that interfere with nucleic acid biosynthesis
2	Anticancer drugs that modulate hormone action
3	Anticancer drugs acting via radical species
4	Anticancer drugs acting via DNA alkylating agents
5	Anticancer drugs acting via DNA minor groove
6	Anticancer drugs acting via DNA intercalation
7	Epigenetic therapy
8	Anticancer drugs targeting tubulin and microtubules
9	Drugs that inhibit signaling pathways for tumor cell growth and proliferation
10	Non-biological approaches to targeted cancer chemotherapy
11	Biological therapy of cancer

Resource: Avendano and Menendez (2008(a)); (2015(b))

1.3.6 DNA intercalating agents

1.3.6.1 Mechanism of DNA intercalating agents

DNA has been identified as a primary target for anticancer drugs, which can change DNA conformation and inhibit duplication or transcription of DNA in cell cycle, and is considered one of the most promising biological receptors for the development of chemotherapeutic agents (Li et al., 2012; Rescifina et al., 2014). They are many anticancer drugs in clinical therapy that interacted with DNA via non-covalent bond interactions (pi-stacking, hydrophobic, ionic, hydrogen bonding, and van der Waals of the small molecule with nucleic acid bases). There are few major modes for reversible binding of anti-cancer drug with double-helical DNA, DNA intercalation and DNA groove binding (Chaires, 1998; Palchaudhuri and Hergenrother, 2007; Nakamoto, Tsuboi and Strahan, 2008; Ihmels and Thomas, 2011; Rescifina et al., 2014). The binding mode of anti-cancer with DNA can be classified as (i) electrostatic attractions with the anionic sugarphosphate backbone of DNA, (ii) interactions with the DNA major groove, (iii) interactions with the DNA minor groove, (iv) intercalation between base pairs via the DNA major groove, (v) intercalation between base pairs via the DNA minor groove, and (vi) threading intercalation mode as shown in Figure 1.15 (Ihmels and Thomas, 2012; Rescifina et

al., 2014). The binding mechanism of DNA intercalating agents interacted with DNA was studied (Rizzo, Sacchi and Menozzi, 1989; Lei, Wang and Wu; 2012). The five step kinetic model proposed based on the experimental data studied by Rizzo (Rizzo, Sacchi and Menozzi, 1989) and simulations results provided by Le (Le, Wang and Wu; 2012) includes a parallel arrangement of step 1 (one off-pathway weak bound step) and step 2 (on-pathway weak bound step, on-pathway minor groove binding), followed by on-pathway intercalation step 3 (an opening and flipping of base), followed by another parallel arrangement of step 4 (flipping back of one base) and 5 (either conformational rearrangement of the drug–DNA complex or redistribution of bound drug to preferred sites without dissociation).

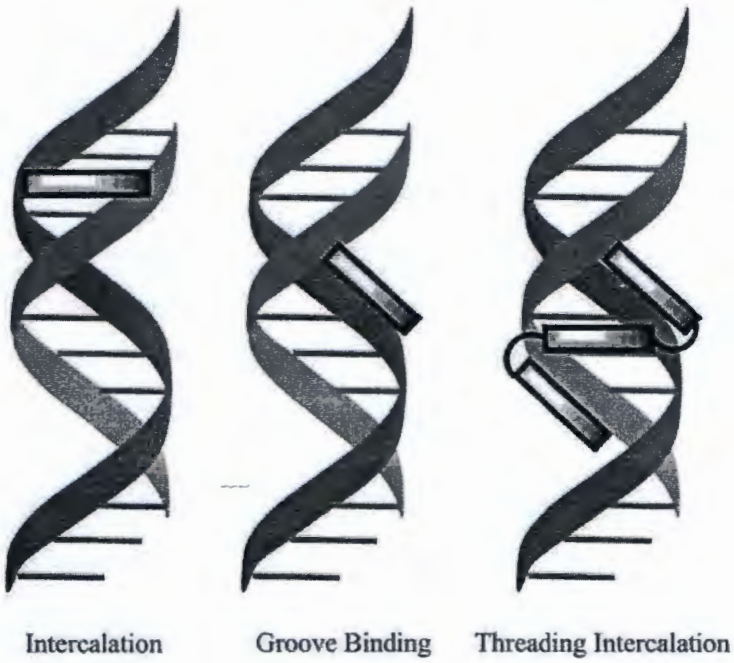


Figure 1.16 DNA-drug binding mode

Resource: Rescifina et al. (2014)

1.3.6.2 Classifications of intercalating agents

Intercalating anti-cancer drug can be defined as the process by which compounds containing planar aromatic or heteroaromatic ring systems are inserted between adjacent base pairs perpendicularly to the axis of the helix DNA. Most of intercalating drugs contain three or four fused rings. Intercalation of a drug

molecule into DNA is only the first step in a series of events that eventually lead to its biological effects. Structural changes induced in DNA by intercalation lead to interference with recognition and function of DNA-associated proteins such as polymerases, transcription factors, DNA repair systems, and, especially, topoisomerases. The anti-cancer drugs that interacted with DNA duplex via DNA intercalation process were summarized in Table 1.8 (Avendano and Menendez, 2007(a); 2015(b)).

Table 1.8 Classifications of intercalating anti-cancer drugs

Mechanism	Class	Drug or clinical trial inhibitor
Monofunctional intercalating agents	Ellipticine and its analogs	Celiptium® Datelliptium®
	Actinomycin	Actinomycin D (dactinomycin, Cosmegen®)
	Fused quinolines	TAS-103
	Naphthalimides	Mitonafide Amonafide
	Chartreusin and Elsamicin A	Chartreusin IST-622 Elsamicin A
	Other monofunctional intercalating agents	Acridines Anthracyclines
Bifunctional intercalating agents		Ditercalinium Elinafide (LU 79553) Echinomycin

Research: Avendano and Menendez (2007(a)); (2015(b))

Table 1.8 Classifications of intercalating anti-cancer drugs (continued)

Mechanism	Class	Drug or clinical trial inhibitor
Indirect DNA damage by DNA topoisomerase inhibitors	Camptothecin as Topoisomerase I inhibitors	Camptothecin (CPT)
		Topotecan (Hycampin®)
		Irinotecan (Camptosar®)
		elomotecan (BN80927)
		diflomotecan (BN80915)
		S39625
		Silatecan (AR67)85
		Karenitecin (BNP1350)
		Rubitecan
		9-aminocamptothecin
		Lurtotecan
		Exatecan (DX-8951f)
		DRF1042,94
		Belotecan (Camtobell®)
		Gimatecan
		Namitecan (ST1968)
		CMMD-Gly
		TLC388 (Lipotecan®)
	Non-camptothecin as topoisomerase I inhibitors	Staurosporine
		UCN-01
		Rebeccamycin
		NSC-655649 (BMY-27557-14)
		NB-506
		Edotecarin (J-107088)
		NSC-725776
		NSC-724998
		ARC-111 (topovale)
		Genz-644282
		Lamellarin D

Research: Avendano and Menendez (2007(a)); (2015(b))

Table 1.8 Classifications of intercalating anti-cancer drugs (continued)

Mechanism	Class	Drug or clinical trial inhibitor
	Acridine derivatives as topoisomerase II poisons inhibitors	Amsacrine (m-AMSA, Amsidyl®, Amsidine®, Amerkin®) Asulacrine DACA (XR5000) KW-2170 PD-115934
	Anthracyclines and related compounds topoisomerase II poisons inhibitors	Doxorubicin (Adriamycin®, Rubex®) Daunomycin (Cerubidine®) Idarubicin (Idamycin®, 4-demethoxydaunorubicin) Nogalamycin Mitoxantrone (Novantrone®) Pixantrone (Pixuvri®)
	Etoposide and its analogs as non-intercalating topoisomerase II poisons inhibitors	Podophyllotoxin (PPT, podofilox) Etoposide (VP-16-213) Teniposide (Vumon®) Etoposide phosphate (Etopophos®) TOP-53 F14512 Tafluposide
	Salvicine as non-intercalating topoisomerase II poisons inhibitors	Salvicine

Research: Avendano and Menendez (2008(a)); (2015(b))

Table 1.8 Classifications of intercalating anti-cancer drugs (continued)

Mechanism	Class	Drug or clinical trial inhibitor
	Inhibitors of the binding of topoisomerase II to DNA	Aclarubicin (aclacinomycin A, Aclacin®) Merbarone Dexrazoxane hydrochloride (Totec®, Savene®) Sobuzoxane (MST-16, Perazolin®)
Telomerase inhibitors and other anticancer approaches targeting telomeres	G-quadruplex ligands	Ethidium bromide Dibenzo[*bj*](1,7)phenanthroline 115405 Telomestatin Quarfloxin (quarfloxacin, CX-3543)
	Inhibitors of telomerase reverse transcriptase	BIBR1532 (Sirong®)
	Inhibitors of the RNA domain template	Imetelstat (GRN-163 L)
DNA repair inhibitors		Ecteinascidin 743 (trabectedin, ET-743, Yondelis®) Nemorubicin

Research: Avendano and Menendez (2007(a)); (2015(b))

1.3.6.3 Azanaphthoquinone annelated pyrrole derivatives

The anthracyclines analogues and mitoxanthrone, a DNA intercalating compounds have been known as a key class of compounds for cancer chemotherapy (Fox and Smith, 1990; Thuston and Lobo, 1998; Edan, Morrissey and Le Page, 2004; Vuimo et al., 2006). However, the major drawback of using these agents is the cytotoxicity (mostly blood, bone marrow and heart) (Frishman et al., 1997; Gonsette, 2007). For these reasons many efforts are focusing on the development of new core structures of DNA intercalating agents which show lower organ toxicity. In the search for new core structures of DNA intercalating agents with

lower cardiotoxicity, the development of aza-bioisosteric chemotypes were applied (Krapcho et al., 1994; Shchekotikhin et al., 2009). Based on the aza-bioisosteric chemotypes approach, an aza-anthracene-9,10-dione, BBR 2778 (Pixantrone) was reported an DNA intercalatine agent with less cardiotoxicity (Engert et al., 2006; Cavalletti et al., 2007; Tomillero and Moral, 2009; Mukherji and Pettengell, 2009). Based on this finding, aza and diaza bioisosteric anthracene-9,10-dione were developed as anti-cancer agents (Krapcho et al., 1995(a); 1998(b); Burckhardt et al., 1998; Sissi and Palumbo, 2004; Antonini et al., 2008). Azanaphthoquinone annelated pyrrole core structures as shown in Figure 1.16 were developed by H. Spreitzer (Spreitzer et al., 2001). With the continuous effort to develop novel DNA intercalating agents, compounds based on the azanaphthoquinone annelated pyrrole scaffold were developed as anti-cancer agents (Shanab et al., 2007(a); 2010(b); 2011(c); Pongprom et al; 2009(a); 2010(b); 2012). The evaluation for cytotoxic activity against different human cancer cell lines shows the promising activities of azanaphthoquinone annelated pyrrole derivatives. Moreover, lead compounds of this derivative show better antiproliferative effects than paclitaxel and doxorubicin on multidrug resistant cell lines (Shanab et al., 2007(a)). However, the major drawback of these compounds is the easy metabolic cleavage of the oxime group (Pongprom et al; 2009(a)). To overcome this disadvantage of azanaphthoquinone annelated pyrrole derivatives, compounds containing a piperidinyl carbinol instead of the oxime group were developed (Pongprom et al; 2009(a)). The replacement of the oxime group with a piperidinyl carbinol could increase the stability of compounds from the metabolic cleavage. However, the series of synthesized compounds containing a piperidinyl carbinol display only moderate activity (Pongprom et al; 2009(a)).

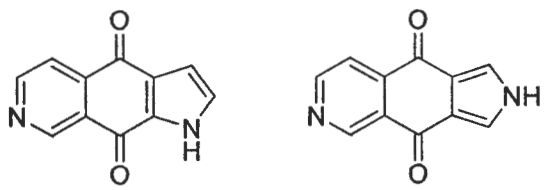


Figure 1.17 Structure of azanaphthoquinone annelated pyrrole core structures

Resource: Spreitzer et al. (2001)

1.4 Objectives

In the present work, molecular modeling and computer aided drug design approaches have been applied to elucidate both anti-tuberculosis agents targeting the InhA and PknG inhibitors; and cancer in the class of DNA intercalating agents with the aim

1.4.1 To gain insight into the structural requirement to improve the inhibitory activity against InhA and PknG enzymes using 3D-QSAR CoMSIA method

1.4.2 To elucidate the potential anti-TB binding modes and important inhibitor-enzyme interactions of direct InhA and PknG inhibitors using MD simulation calculations and binding free energy calculations

1.4.3 To design new and highly potent InhA and PknG inhibitors as anti-TB agents based on the combination results from 3D-QSAR CoMSIA model and MD simulations

1.4.4 To discovery new analogues of InhA and PknG inhibitors as anti-TB agents thought structure based virtual screening approach

1.4.5 To determine the structural requirement of intercalating agent in a series of azanaphthoquinone anulated pyrrole derivatives as anti-cancer agents using 3D-QSAR CoMSIA method

1.4.6 To gain insight into the potential binding modes and main interactions of azanaphthoquinone anulated pyrrole derivatives on DNA duplex using MD simulation and binding free energy calculations

1.4.7 To design novel azanaphthoquinone anulated pyrrole derivatives as anti-cancer agents with the high predicted inhibitory activities based on the integrated results from 3D-QSAR CoMSIA model and MD simulations

CHAPTER 2
COMPUTER AIDED DRUG DESIGN STRATEGY

To find of new potential anti-TB agents and anti-cancer agents, the potential CAMD approaches were applied. The theories and methods of drug design methods were summarized in this chapter.

2.1 Theory of drug design and discovery methods

The explosive growth of computer power has led to the development of large-scale simulation for rational design new potent inhibitors (Sagui and Darden, 1999). Characteristic length-scales currently associated with varying levels of description in biomolecular simulations. From the scale of the atom to a whole organism as illustrated in Figure 2.1, various types of simulations have been performed to investigate electronic, atomic and physical properties of molecules. Ab initio and semi-empirical quantum mechanical calculations permit the study of chemical reactions in electronic detail within single molecules and small proteins while molecular dynamics simulations allow for the study of biological phenomena from the individual protein level to large subcellular organelles, and at all levels in between (Perilla et al., 2015). Various calculations methods of CAMD were applied in this study to gain insight into the structural information in drug design. Two maims approaches of drug design, ligand based drug design (LBDD) and structure based drug design (SBDD) were applied. In addition, virtual screening technique was applied to identify potential hits for anti-TB agents and anti-cancer agents.

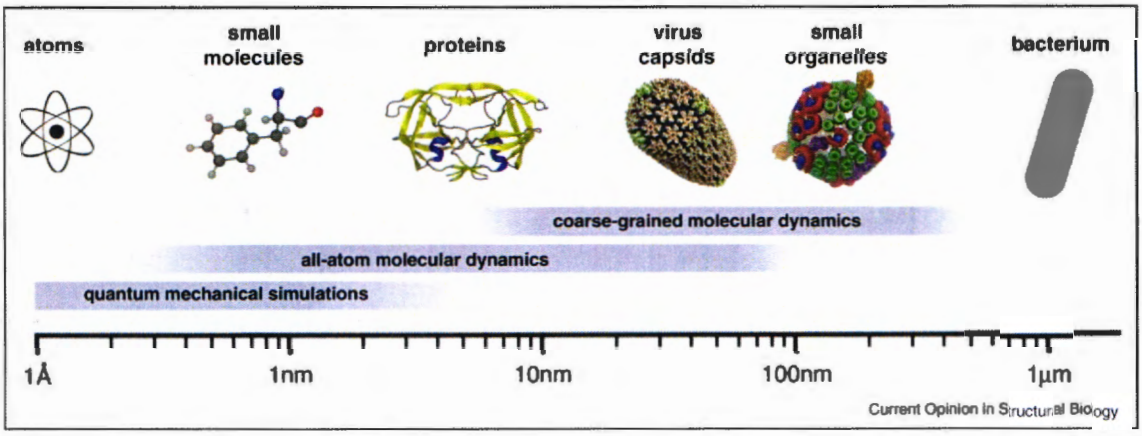


Figure 2.1 Characteristic time and length scales for various biological processes
Resource: Perilla et al. (2015)

2.1.1 Ligand based drug design (LBDD)

Ligand based drug design (LBDD) is the drug design methods that used the information of ligands or small molecules to design new potential compounds. This approach was used when the absence of an experimental 3D structure of the target. This work, quantitative structure–activity relationships (QSAR) was applied. QSAR is the ligand based drug design methods that developed to understand the relationship between chemical structure and biological affects with the aims to obtain a reliable statistical model for prediction of the activities of new chemical entities and comprehend and rationalize the mechanisms of action within a series of chemicals as shown in equation. QSAR is the mathematic equation that used to explain the relationship between the biological effects with the molecular properties of ligands as shown the simple mathematic equation in equation 1.

$$B = f(x) \tag{1}$$

B is the biological activities of ligands. f(x) is functions of x (molecular descriptors) that representation of their ligand structure. Based on dimensionality, QSAR methods are classified into following classes, based on the structural representation (Verma, Khedkar and Coutinho, 2010; Cherkasov et al, 2013).

1D-QSAR correlating activity with molecular formula and global molecular properties like pKa, log P etc.

2D-QSAR correlating activity with two-dimensional structural formula (structural patterns like connectivity indices, 2D-pharmacophores etc.) without taking into account the 3D-representation of these properties

3D-QSAR correlating activity with non-covalent interaction fields (conformation-dependent) surrounding the molecules

4D-QSAR additionally including ensemble of ligand configurations in 3D-QSAR

5D-QSAR explicitly representing different induced-fit models in 4D-QSAR

6D-QSAR further incorporating different solvation models in 5D-QSAR

The most commonly used mathematical technique in classical QSAR work is multiple regression analysis. In the classical QSAR studies, biological activities of ligands with atomic, group or molecular properties have been correlated.

2.1.1.1 Comparative Molecular Field Analysis (CoMFA)

Three-dimensional quantitative structure-activity relationships (3D-QSAR) were developed. 3D-QSAR approaches attempt to map a receptor surface by analyzing a QSAR equation for noncovalent interactions of the different positions of substitution of ligands (Kubinyi, 1993(a)). Comparative molecular field analysis (CoMFA) is one of the most popular 3D-QSAR methods used to investigate the structural requirements of ligands to improve the biological activity. The idea underlying the CoMFA methodology is that differences in biological activity are often related to differences in the magnitudes of molecular fields surrounding the investigated receptor ligands (Bordas, Komives and Lopata, 2003). Receptor binding is directly proportional to the biological activity (Verma, Khedkar and Coutinho, 2010). CoMFAs describe 3D structure activity relationships in a quantitative manner. Two observations were decried as (1) the interactions at the molecular level which produce an observed biological effect are usually non-covalent; and (2) molecular mechanics force fields, most of which treat noncovalent (non-bonded) interactions only as steric and electrostatic forces, can account precisely for a great variety of observed molecular properties (Cramer, Patterson and Bunce, 1988). The CoMFA approach uses in its standard implementation only steric forces (Lennard-Jonesas potential shown in equation 2) and electrostatic forces (Coulomb potentials as shown in equation 3) (Cramer, Patterson and Bunce, 1988) as shown in Figure 2.2. In close proximity to the surface of the atoms both potentials have very steep slopes. They

approach infinite values if the atom positions of two molecules overlap. To avoid this, arbitrary cut-offs are defined and all larger positive (or negative) values are set to these cut-off values.

$$E_{vdW} = \sum_{i=1}^{n} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) \tag{2}$$

E_{vdW} =van der Waals, interaction energy, r_{ij} = distance between atom i of the molecule and the grid point j where the probe atom is located; A_{ij} and B_{ij} are constants that depend on the van der Waals, radii of the corresponding atoms.

$$E_C = \sum_{i=1}^{n} \frac{q_i q_j}{D r_{ij}} \tag{3}$$

E_C = coulomb interaction energy, q_i = partial charge of atom i of the molecule, q_j = charge of the probe atom, D = dielectric constant, r_{ij} = distance between atom i of the molecule and the grid point j, where the probe atom is located

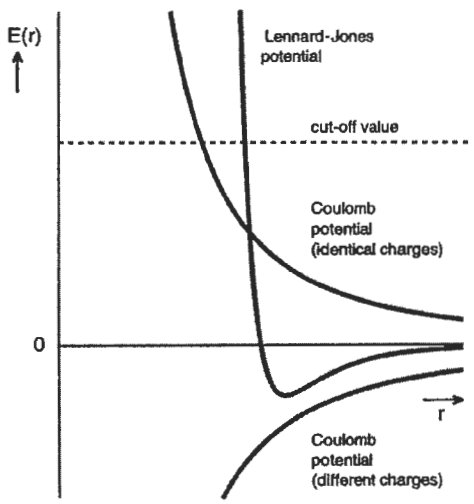


Figure 2.2 Electrostatic and steric fields in CoMFA studies are calculated from Coulomb and Lennard-Jones potentials, respectively

Resource: Kubinyi (1998)

The process of CoMFA as shown in Figure 2.3, the 3D-structure of ligands was sampled at the intersections of a 3D-lattice to calculate the steric and electrostatic interactions fields between the compounds of interest, and a "probe atom" placed at the various intersections of a 3D-lattice (Cramer, Patterson and Bunce, 1988). Then, QSAR models were set up using partial least squares (PLS) (an iterative regression method that produces its solutions based on linear transformation of a large number of molecular descriptors to a small number of new orthogonal terms of molecular descriptors) using cross-validation to maximize the likelihood that the results have predictive validity (Cramer, Patterson and Bunce, 1988; Clark et al, 1990). A cross-validated r^2 (r^2_{cv} or q^2) or predictive ability much be higher than 0.6 as shown in equation 4 was used to determine the performance of QSAR models.

$$r^2_{cv}, q^2 = 1 - \left(\frac{PRESS}{SSY}\right) = 1 - \frac{\sum (y_{\exp} - y_{pred})^2}{\sum (y_{\exp} - y_{mean,\exp})^2} \qquad (4)$$

Where y_{exp} is experimental biological activity, y_{pred} is predicted biological activity and $y_{mean,exp}$ is average experimental biological activity. Finally, graphic presentation of QSAR results, as contoured three-dimensional coefficient plots were derived to determine the structural requirements of ligands to improve their biological activity.

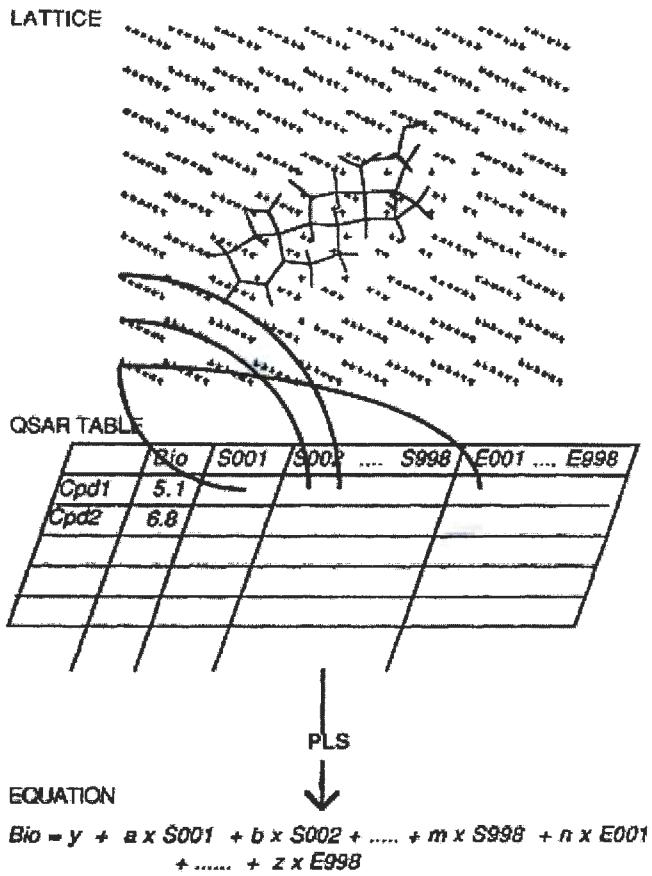


Figure 2.3 Standard CoMFA process

Resource: Cramer et al. (1988)

2.1.1.2 Comparative molecular similarity indices analysis (CoMSIA)

Comparative molecular similarity indices analysis (CoMSIA) was developed to solve the problems from CoMFA method (Klebe, Abraham, and Mietzner, 1994). In CoMFA method, the Lennard-Jones potential is very steep close to the van der Waals surface and the potential energy expressed at grid points in the proximity of the surface changes dramatically. It is likely that values from this region display significant descriptors in a QSAR. CoMSIA method was developed to compute property fields based on similarity indices of drug molecules that have been brought into a common alignment. The fields of different physicochemical properties use Gaussian-type distance dependence and no singularities occur at the atomic positions. The most important advantage of the CoMSIA fields is their 'smooth' nature (Figure 2.4). The slopes of the underlying Gaussian functions are not as steep as

the Coulomb and Lennard-Jones potentials; therefore, no cut-off values need to be defined. Moreover, additional molecular descriptors were added to gain insight into the crucial structural requirement of ligands. Similar to the usual CoMFA approach, a data table has been constructed from similarity indices calculated via a common probe atom which is placed at the intersections of a regularly spaced lattice. A grid spacing of 1 A has been used throughout this study. Similarity indices A F , k between the compounds of interest and a probe atom, systematically placed at the intersections of the lattice, have been calculated according to (e.g., at grid point *q* for molecule *j* of the data set) .

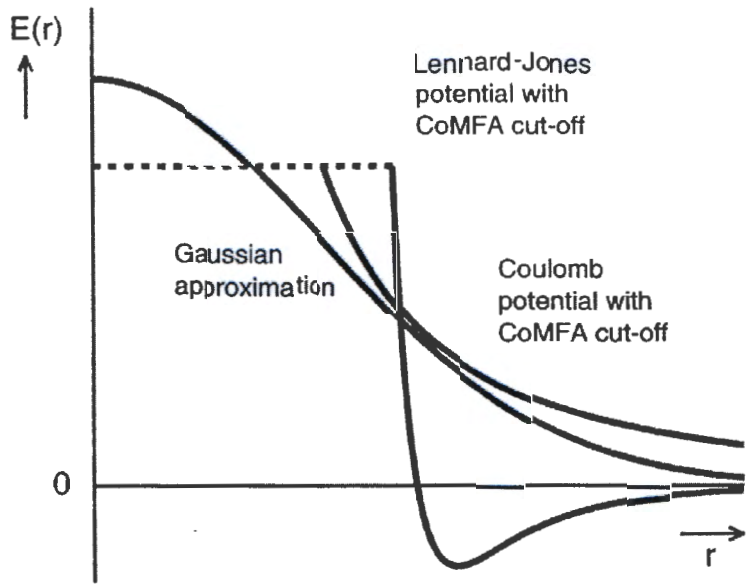


Figure 2.4 Bell-shaped Gaussian functions of CoMSIA fields

Resource: Kubinyi (1998)

$$A_{F,k}^{q}(j) = -\sum_{i=1}^{n} w_{probe,k} w_{i,k} e^{-\alpha r_{iq}^{2}} \quad (5)$$

Where i = summation index over all atoms of the molecule j under investigation; $w_{i,k}$ = actual value of the physicochemical property k of atom i ; $w_{probe,k}$ = probe atom with charge +1,radius 1 Å, and hydrophobicity +1; α= attenuation factor; and r_{iq} = mutual distance between probe atom at grid point *q* and atom *i* of the test molecule.

2.1.2 Structure based drug design (SBDD)

Structure based drug design (SBDD) is the drug design approaches that used an experimental 3D structure available. There are many available structure based drug design approaches used to rational design potential drug. This works, molecular docking calculations and molecular dynamics (MD) simulation combined with binding free energy calculations were applied.

2.1.2.1 Molecular docking calculations

Molecular docking calculations is a simple structure based drug design (the use of three-dimensional structural information gathered from biological targets) that among the most frequently used SBDD strategies (Kalyaanamoorthy and Chen, 2011). This method can go hand in-hand with major phases of drug discovery that ideally exhibit some degree of potency and specificity against the target. There are three important steps of docking, target identification, binding site recognition and scoring function, respectively.

1) Target site identifications

A typical molecular docking method begins with the identification and validation of the drug binding target structure (Kalyaanamoorthy and Chen, 2011). The choice and preparation of the structural model of a drug targeted binding site are important variables (Kitchen et al., 2004). Normally, the information of drug binding site was derived from experimental methods such as X-ray crystallography or NMR techniques. However, some of drug binding targets are no experimentally determined structures. Computational modeling approaches, such as ab initio modeling, threading and comparative modeling can be used to predict 3D structures of drug binding targets.

2) Binding site recognition

The drug binding site is a small region, where ligands can best fit or bind to activate the target and produce the desirable effect (Kalyaanamoorthy and Chen, 2011). In molecular docking approaches, conformation search is especial step to find best fit binding mode of ligand. Docking algorithms are methods to sampling the orientation of ligand in the binding site. There are three types of docking algorithms. There are different levels of approximation in docking. The basic approximations in docking algorithms are flexible ligand-search docking (very popular approaches) and

flexible protein docking, respectively. A list of widely used docking algorithms categorized according to the conformational search methodology is provided in Table 2.1 (Sousa, Fernandes and Ramos, 2006). Two types of docking algorithms, systematic and random/stochastic search algorithms are wildly used to develop docking programs. Listed docking programs with different docking algorithms were summarized in Table 2.2.

Table 2.1 Types of flexible-ligand and flexible receptor search algorithms

Flexible-ligand docking (flexible-rigid docking)	**Flexible-protein docking (flexible- flexible docking)**
1. Systematic	Molecular dynamics (MD)
1.1 Conformational	Monte Carlo (MC)
1.2 Fragmentation	Rotamer libraries
1.3 Database	Protein-ensemble grids
2. Random/stochastic	Soft-receptor modeling
2.1 Monte Carlo (MC)	
2.2 Genetic algorithm (GA)	
2.3 Tabu Search	
3. Simulation methods	
3.1 Molecular dynamics (MD)	
3.2 Energy minimization	

Resource: Sousa, Fernandes and Ramos (2006)

Table 2.2 Examples of docking programs with conformational search algorithms

Systematic	Random/Stochastic
eHiTS	AutoDock
FRED	Gold
Surflex-Dock	PRO_LEADS
DOCK	EADock
Glide	ICM
EUDOC	LigandFit
FlexX	Molegro Virtual Docker
Hammerhead	CDocker
Flog	GlamDock
SLIDE	PLANTS
ADAM	MolDock
	MOE_Dock

Resource: Ferreira et al. (2015)

3) Scoring function

To select the best fit binding mode of ligand in the receptor binding pocket, molecular docking program used scoring functions to estimate the binding-mode prediction (binding energetics), relative affinity ranking, and/or estimation of absolute binding free energy of the predicted ligand-receptor complexes (Ferreira et al., 2015; Foloppe and Hubbard, 2006; Sousa, Fernandes and Ramos, 2006; Meng et al., 2011). The purpose of the scoring function is to delineate the true ligand binding modes from incorrect ligand binding modes, or binders from inactive compounds in a reasonable computation time (Meng et al., 2011; Sousa, Fernandes and Ramos, 2006). Estimating binding free energies accurately is a time-consuming process (Brooijmans and Kuntz, 2003). In docking program, fast and highly efficient methods to discriminate the ligand-protein binding affinity are required simplified scoring functions. Therefore, scoring function for docking studies has led to a number of different functions that can be divided into three main classes, namely force-field

based scoring functions, empirical methods, and knowledge-based potentials. The listed of docking with programs with different docking scoring functions were summarized in Table 2.3.

Table 2.3 Scoring functions implemented in widely used molecular docking programs

Force-field based	Empirical	Knowledge-based
DOCK	AutoDock	SMoG
AutoDock	GlideScore	DrugScore
GoldScore	ChemScore	PMF_Score
ICM	X_Score	MotifScore
LigandFit	F_Score	RF_Score
Molegro Virtual Docker	Fresno	PESD_SVM
SYBYL_G-Score	SCORE	PoseScore
SYBYL_D-Score	LUDI	
MedusaScore	SFCscore	
	HYDE	
	LigScore	
	PLP	

Resource: Ferreira et al. (2015)

4) Docking programs

The most commonly used docking programs as shown in Figure 2.5 are Autodock (25.87%), GOLD (16.69 %) and Glide (14.91 %), respectively (Chen, 2015). Therefore, three of most used docking programs were selected to predict the binding modes of ligands into the binding site in this study. In addition, MOE docking program was used.

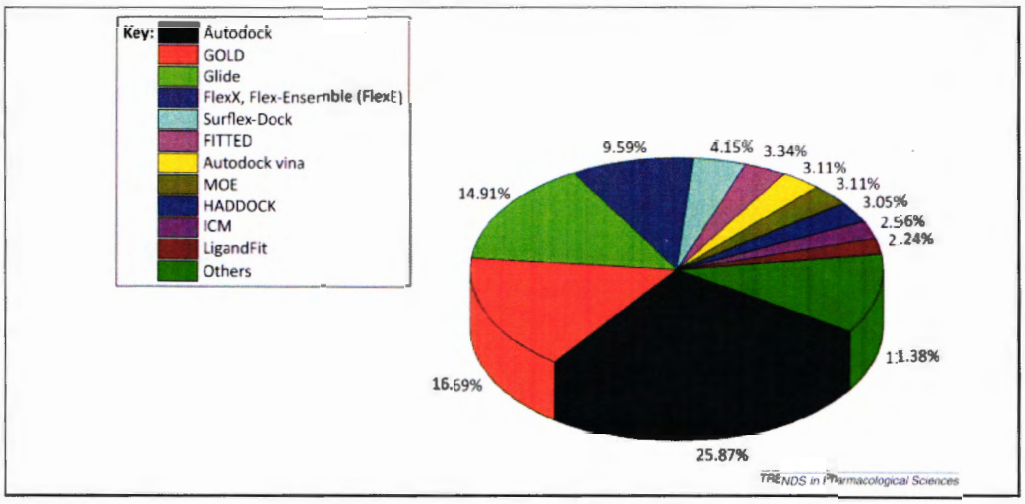


Figure 2.5 All docking publications from 1990 to 2013

Resource: Chen (2015)

4.1) Autodock

Autodock program uses a Lamarckian genetic algorithm (LGA) as the docking algorithm to generate ligand orientations in the active site of receptors. This program was developed to provide an automated procedure for predicting the interaction of ligands with biomacromolecular targets (Morri et al., 2001). A semi-empirical free energy force field to evaluate conformations during docking of Autodock was used. The force field evaluates binding of ligand in the active site in two steps as shown in Figure 2.6. The ligand and receptor start in a receptor-ligand unbound state. In the first step, the intramolecular binding energetics is estimated for the transition from these unbound states to the ligand and protein in the bound state. The second step then evaluates the intermolecular energetics of combining the ligand and protein in their bound stage (Huey et al., 2007(b); Olson et al., 2012). The docking scoring function of Autodock program divided into six pair-wise evaluations (V) and an estimate of the conformational entropy lost upon binding (ΔS_{conf}) as shown in equation 6.

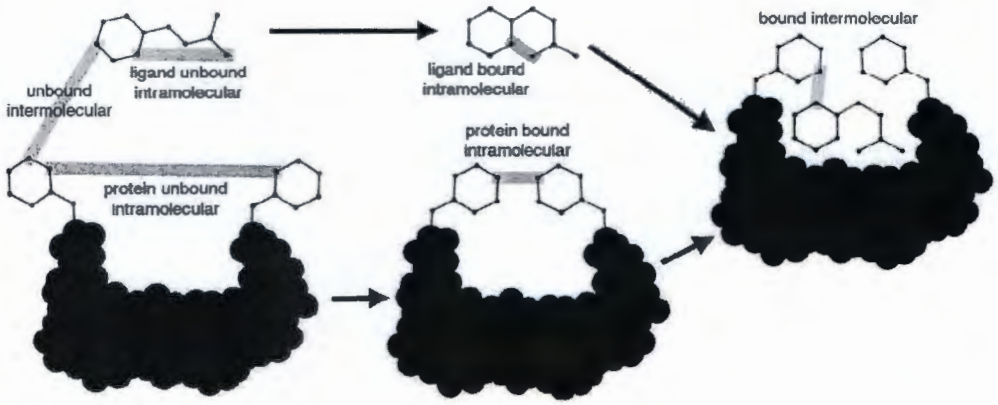


Figure 2.6 Force field evaluates binding in two steps. The ligand and protein start in an unbound conformation

Resource: Huey et al. (2007(b))

$$\Delta G = \left(V_{bound}^{L-L} - V_{unbound}^{L-L}\right) + \left(V_{bound}^{R-R} - V_{unbound}^{R-R}\right) + \left(V_{bound}^{R-L} - V_{unbound}^{R-L} + \Delta S_{conf}\right) \tag{6}$$

Where L refers to the "ligand" and R refers to the "ligand receptor" in a ligand-protein docking calculation. Each of the pair-wise energetic terms (ligand-ligand, receptor-receptor and receptor-ligand) includes evaluations for dispersion/repulsion, hydrogen bonding, electrostatics, and desolvation terms as shown in equation 7.

$$V = W_{vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{il}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + W_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} + \frac{D_{ij}}{r_{ij}^{10}} \right) +$$

$$W_{elec} \sum_{i,j} \frac{q_i q_j}{e(r_{ij}) r_{ij}} + W_{solv} \left(S_i V_j + S_j V_i \right) e^{\left(-r_{ij}^2 / \sigma^2\right)} \tag{7}$$

The weighting constants W were optimized to calibrate the empirical free energy in a set of experimentally determined binding constants (Huey et al., 2004). The intermolecular potentials were calculated by summations over all pairs of ligand atoms, i, and protein atoms, j, as a function of their distances (r). The dispersion/repulsion energy (van der Waal's interaction energy), first term is a typical 6/12 potential for dispersion/repulsion interactions. The parameters of

dispersion/repulsion energy term (A and B) are based on the Amber force field. AMBER force field energy functions were designed to determine potential energy of two atoms in the gas phase, which is only one component of the free energy change in a receptor–ligand binding process (Liu and Wang, 2015; Weiner et al., 1984). The second term is a directional hydrogen bond interaction term based on a 10/12 potential. The hydrogen bond parameters (C and D) are assigned to give a maximal well depth of 5 kcal/mol at 1.9Å for hydrogen bonds with O-H and N-H, and a well depth of 1 kcal/mol at 2.5Å for hydrogen bonds with (S-H) (Goodford, 1985). The function E(t) is a function of the angle (t) of the probe atom away from the ideal position for hydrogen bonding (Huey et al., 2004). The third term is a screened Coulomb potential for electrostatics with a distance-dependent dielectric screening (ε).Where qi and qj are the atomic charges of ligand atom i and receptor atom j. The radial permittivity function, $\varepsilon(r_{ij})$ is usually set to $4r_{ij}$, reflecting the screening effect of water on electrostatic interactions (Mehler and Solmajer, 1991; Huang and Zou, 2010). The final term is a desolvation potential based on the volume of atoms (V) that surround a given atom and shelter it from solvent, weighted by the solvent-accessible surfaces of ligand (S_i) and receptor (S_j); and an exponential term with distance-weighting factor σ=3.5Å (Huey et al., 2007(b); Forli and Olson, 2012).

The entropy of ligand binding (ΔS_{conf}) in receptor is included to account for the loss of degrees of freedom upon binding, which is proportional to the number of sp^3 bonds in the ligand (N_{tor}) (Huey et al., 2007(b); Forli and Olson, 2012).

$$\Delta S_{conf} = W_{conf} N_{tor} \tag{8}$$

The number of rotatable bonds includes all torsional degrees of freedom, including rotation of polar hydrogen atoms on hydroxyl groups and the like.

4.2) GOLD

Genetic Optimisation for Ligand Docking, GOLD docking program is an automated ligand docking used a genetic search algorithm based search method for generating ligand poses and allows for full ligand flexibility (Jones et al.,

1995; 1997; Sousa et al., 2013). GoldScore, ChemScore, Astex Statistical Potential (ASP), and Piecewise Linear Potential (PLP) were developed ad scoring functions to estimate the binding affinity of ligand in the active site of receptor.

The GoldScore fitness function is the original scoring function provided with GOLD. The Goldscore fitness function, a molecular-mechanics-like function has four components as shown in equations 9.

$$GoldScore\,fitness = S_{hb\_ext} + S_{vdw\_ext} + S_{hb\_int} + S_{vdw\_int} \tag{9}$$

where S_{hb_ext} is the receptor–ligand hydrogen-bond score and S_{vdw_ext} is the receptor–ligand van der Waals score. S_{hb_int} is the contribution to the Fitness due to intramolecular hydrogen bonds in the ligand; S_{vdw_int} is the contribution due to intramolecular strain in the ligand. Normally, the best result is obtained by letting the internal hydrogen bonding (S_{hb_int}) tend to zero. Therefore, the Goldscore fitness was calculated from equation 10.

$$GoldScore\,fitness = \sum_{ij}\left(\frac{A_{ij}}{r_{ij}^{8}} - \frac{B_{ij}}{r_{ij}^{4}}\right) + \sum_{ij}\left[(E_{da} + E_{ww}) - (E_{dw} + E_{aw})\right] + \left\{\sum_{ligand}\left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{6}}\right) + \sum_{ligand}\frac{1}{2}V\left[1 + \frac{n}{|n|}\cos(|n|\omega)\right]\right\} \tag{10}$$

The first term, complexed van der Waals score was determined from the placement of the ligand into the active site of the receptor, a 4-8 potential with linear cut-off (Jones et al., 1997). Where complexed van der Waals score was the energy of interaction between two atoms (ligand atom I and receptor atom j) and r_{ij} was the distance between them. The parameters A and B were the pairwise interaction, which it is much softer than the standard 6-12 potential. Complexed van der Waals score was zero for the interaction between a donor hydrogen atom and an acceptor, while the distance between the donor and acceptor is scaled by a factor of 1.43.

Second term, hydrogen bond energy was obtained for the complex. This energy was the sum of all individual bond energies found from all

combinations of ligand donor hydrogen atom and receptor acceptor and all combinations of ligand acceptor and receptor donor hydrogen atom (Jones et al., 1995; 1997). The hydrogen-bond energy between a donor and an acceptor is an important component of the fitness function since each hydrogen-bonding pair contributes to the overall energy of binding. Initially the donor (d) and the acceptor (a) are in solution but on coming together (da) water (w) is stripped off.

The last term is the internal energy term included the ligand van der Waals score and torsional energies. The steric energy was determined using a 6-12 potential. The energy of association between two molecules can be represented using a Lennard-Jones 6-12 potential, where the second term accounts for the attractive dispersion energy between two molecules. C and D parameters were calculated as the pairwise interactions.

The ChemScore fitness function incorporates a term, ΔG, that represents the total free energy change that occurs on ligand binding and was trained by regression against binding affinity data for 82 complexes (Eldridge et al., 1997; Baxter et al., 1998). The ChemScore fitness function also incorporates a protein-ligand atom clash term and an internal energy term. ChemScore takes account of hydrophobic-hydrophobic contact area, hydrogen bonding, ligand flexibility and metal interaction. The empirical scoring function can be written in the form equation 11

$$\Delta G_{bind} = \Delta G_0 + \Delta_{hbond} \sum_{iI} g_1(\Delta r) g_2(\Delta \alpha) + \Delta G_{metal} \sum_{aM} f(r_{aM}) +$$

$$\Delta G_{lipo} \sum_{1L} f(r_{1L}) + \Delta G_{rot} H_{rot} \qquad (11)$$

The ΔG coefficients are unknown and will be obtained by multiple linear regression. The hydrogen bond term, $\sum_{iI} g_1 g_2$, is calculated for all complementary possibilities of hydrogen bonds between ligand atoms, i, and receptor atoms, I. The functions g_1 and g_2 are of the same form as used by Böhm as shown in equation 12 and 13:

$$g_1(\Delta r) = \begin{cases} 1 & if\ \Delta r \le 0.25\,\text{Å} \\ 1-(\Delta r - 0.25)/0.4 & if\ 0.25\,\text{Å} \le \Delta r \le 0.65\,\text{Å} \\ 0 & if\ \Delta r > 0.65\,\text{Å} \end{cases} \tag{12}$$

$$g_2(\Delta\alpha) = \begin{cases} 1 & if\ \Delta\alpha \le 30^\circ \\ 1-(\Delta r - 30)/50 & if\ 30^\circ \le \Delta\alpha \le 80^\circ \\ 0 & if\ \Delta\alpha > 80^\circ \end{cases} \tag{13}$$

Δr is the deviation of the H...O/N hydrogen bond length from 1.85 Å and $\Delta\alpha$ is the deviation of the hydrogen bond angle N/O-H...O/N from its ideal value of 180°. The hydrogen-bond term is computed as a sum over all possible donor-acceptor pairs (Equation 14)

$$\Delta G_{hbond} = \sum_{all\ donor-acceptor\ pairs} B'(\Delta r, \Delta r_{ideal}\Delta r_{\max}, \sigma).B'(\Delta\alpha, \Delta\alpha_{ideal}, \Delta\alpha_{\max}, \sigma_\alpha).$$

$$B'*(\Delta\beta, \Delta\beta_{ideal}, \Delta\beta_{\max}, \sigma_\beta) \tag{14}$$

Where r is the ideal hydrogen...acceptor (H...A) distance in Å (Default value is 1.85 Å). Δr is the absolute deviation of the actual H...A separation from r. Δr_{ideal} is the tolerance window around the H...A distance, r, within which the H-bond is regarded as ideal. This value was set at 0.25. Δr_{max} is The maximum possible deviation from the ideal distance; above this, the interaction is not regarded as an H-bond. The default value was set as 0.65. σ_r is the Gaussian smearing sigma associated with this term (0.1). α is the ideal D-H...A angle (in degrees) (Default is 180 degree). The absolute deviation ($\Delta\alpha$) of the actual D-H...A angle from α was calculated for each H-bond. $\Delta\alpha_{ideal}$ is the he tolerance window around the D-H...A angle, α, within which the H-bond is regarded as ideal (30 degree). $\Delta\alpha_{max}$ is the maximum possible deviation from the ideal D-H...A angle; above this, the interaction

is not regarded as an H-bond (80 degree). σ_α is the e Gaussian smearing sigma associated with this term (10.0). B′* is the sum of all possible values for a given hydrogen bond. β is the ideal H...A-X angle (in degrees) (default is 180 degree). Δβ is the absolute deviation of the actual H...A-X angle from β. $\Delta\beta_{ideal}$ is the tolerance window around the H...A-X angle, β, within which the H-bond is regarded as ideal at 70 degree. $\Delta\beta_{max}$ is e maximum possible deviation from the ideal H...A-X angle; above this, the interaction is not regarded as an H-bond (80 degree). σ_β is the Gaussian smearing sigma associated with this term (10). The metal-binding term in ChemScore is computed as a sum over all possible metalion... acceptor pairs defined in equation 15

$$P_{metal} = \sum_{all\,ligand\,acceptors} \sum_{all\,protein\,metal} B(r_{aM}, R_{ideal}, R_{max}, \sigma_{metal}) \tag{15}$$

r_{aM} is the actual acceptor-metal distance (in Å) that calculated for each acceptor-metal pàir. R_{ideal} is the ideal acceptor-metal distance. This default was set at 2.6 Å. R_{max} is the maximum acceptor-metal distance to be considered a binding interaction (3.0 Å). σ_{metal} is Gaussian smearing sigma associated with this term (0.1). The lipophilic term (equation 16) is defined in a similar way:

$$P_{lipo} = \sum_{all\,ligand\,lipophilic\,atoms} \sum_{all\,protein\,lipophilic\,atoms} B(r_{ll}, R_{ideal}, R_{max}, \sigma_{lipo}) \tag{16}$$

r_{ll} is the actual distance (in Å) betbeen the pair of lipophilic atoms that calculated for each atom-atom pair. R_{ideal} is the ideal acceptor-metal distance. This default was set at 4.1 Å. R_{max} is the maximum acceptor-metal distance to be considered a binding interaction (7.1 Å). σ_{lipo} is Gaussian smearing sigma associated with this term (0.1). The following formula is used to estimate the entropic loss that occurs when single, acyclic bonds in the ligand become non-rotatable upon binding in equation 17:

$$P_{rot} = 1 + \left(1 - \frac{1}{N_{rot}}\right) \sum_r \frac{(P_{nl}(r) + P'_{nl}(r))}{2} \tag{17}$$

N_{rot} is the number of frozen rotatable bonds in the ligand (The expression is deemed to have a value of zero if there are no rotatable bonds in the ligand. $P_{nl}(r)$ and $P'_{nl}(r)$ are the percentages of non-hydrogen atoms on either side of the rotatable bond that are not lipophilic. $P_{nl}(r)$ and $P'_{nl}(r)$ are 30% and 10%, respectively.

The Astex Statistical Potential (ASP) fitness function is an atom-atom distance potential derived from a database of protein-ligand complexes and can be compared to other knowledge-based scoring potentials. ASP score (Equation 18) incorporates some ChemScore terms that can be rewritten as (Mooij and Verdonk, 2005)

$$ASP(i,j,r) = -\ln \frac{c_{obs}(i,j,r)}{\langle c_{obs}(i,j,r') \rangle_{r'=6.0}^{r'=8.0}} \tag{18}$$

Defining c_{obs} (i,j,r), the volume-corrected density of observations at distance r for the atom types i and j is equation 19.

$$c_{obs}(i,j,r) = \frac{n_{obs}(i,j,r)}{\left(f_p(i,r) \cdot f_l(j,r) \cdot 4\pi r^2 \Delta r\right)} \tag{19}$$

The fraction of available volume for a protein atom type i is calculated as the fraction of non-protein grid points at distance r, averaged over all protein atoms of type i that are within 8.0 Å of a ligand atom, in all complexes in the database in equation 20.

$$f_p(i,r) = \frac{gridpts_i^{total}(r) - gridpts_i^{protein}(r)}{gridpts_i^{total}(r)} \tag{20}$$

The ligand volume correction is calculated similarly, averaging over all ligand atoms of type j in all complexes in equation 21.

$$f_l(j,r) = \frac{gridpts_j^{total}(r) - gridpts_j^{protein}(r)}{gridpts_j^{total}(r)} \tag{21}$$

Piecewise Linear Potential (CHEMPLP) uses the ChemScore hydrogen bonding term and multiple linear potentials to model van der Waals and repulsive terms. CHEMPLP is empirical fitness functions optimized for pose prediction in equation 22 and 23.

$$fitness_{PLP} = -\begin{pmatrix} W_{PLP} \cdot f_{PLP} + W_{lig-clash} \cdot f_{lig-clash} + W_{lig-tors} \cdot f_{lig-tors} + \\ f_{chem-cov} + W_{prot} \cdot f_{chem-prot} + W_{cons} \cdot f_{cons} \end{pmatrix} \tag{22}$$

$$fitness_{CHEMPLP} = fitness_{PLP} - (f_{chem-hb} + f_{chem-cho} + f_{chem-met}) \tag{23}$$

In both cases, the f_{PLP} is used to model the steric complementarity between protein and ligand, while for CHEMPLP additionally the distance- and angle-dependent hydrogen and metal bonding terms from ChemScore are considered ($f_{chem-hb}$, $f_{chem-cho}$ and $f_{chem-met}$). The internal score of the ligand consists of the heavy-atom clash potential ($f_{lig-clash}$) as well as the torsional potential used within ChemScore ($_{flig-tors}$). Both fitness functions are capable of covalent docking ($f_{chem-cov}$), considering flexible side-chains ($f_{chem-prot}$) and explicit water molecules as well as handling constraints (f_{cons}).

4.3) Glide

Grid-based ligand docking with energetics, Glide was developed to perform as close to an exhaustive search of the positional, orientational, and conformational space. This program used a series of hierarchical filters to search for possible locations of the ligand in the binding site of receptor. The shape and properties of the receptor in docking procedures are represented on a grid. Initial ligand conformations are selected from an exhaustive enumeration of the minima in the ligand. Docking poses that pass these initial screens enter the final stage of the algorithm, which involves evaluation and minimization of a grid approximation to the OPLS-AA non bonded ligand-receptor interaction energy. There are three different choices of docking precision in Glide docking .Glide offers the full range of speed vs

accuracy options of docking scoring functions, from the high-throughput virtual screening (HTVS), to the standard precision (SP) mode, to the extra precision (XP). These scoring functions in Glide are specifically designed to bring speed, efficiency, and accuracy to lead discovery efforts at different numbers of ligand databases as shown in Figure 2.7. Glide scoring is the empirically based ChemScore function, which can be written as equation 24

$$\Delta G_{bind} = C_0 + C_{lipo}\sum f(r_{1r}) + C_{hbond}\sum g(\Delta r)h(\Delta\alpha) + C_{metal}\sum f(r_{1m}) + C_{rotb}H_{rotb} \quad (24)$$

The summation in the second term extends over all ligand-atom/receptor-atom pairs defined by ChemScore as lipophilic term. The third term extends over all ligand-receptor hydrogen-bonding interactions. f, g, and h are functions that give a full score (1.00) for distances or angles that lie within nominal limits and a partial score (1.00-0.00) for distances or angles that lie outside those limits but inside larger threshold values. For example, g(Δr) is 1.00 if the H---X hydrogen bond distance is within 0.25 Å of a nominal value of 1.85 Å but tails off to zero in a linear fashion if the distance lies between 2.10 and 2.50 Å. Similarly, h(ΔR) is 1.00 if the Z-H---X angle is within 30° of 180° and decreases to zero between 150° and 120°.GlideScore modifies and extends the ChemScore function as follows (Friesner et al., 2004(a); Halgren et al., 2004):

$$\Delta G_{bind} = C_{lipo-lipo}\sum f(r_{1r}) + C_{hbond-neut-neut}\sum g(\Delta r)h(\Delta\alpha) + C_{hbond-neut-charged}\sum g(\Delta r)h(\Delta\alpha) + C_{hbond-charged-charged}\sum g(\Delta r)h(\Delta\alpha) + C_{max-metal-ion}\sum f(r_{1m}) + C_{rotb}H_{rotb} + C_{polar-phob}V_{polar-phob}C_{coul}E_{coul} + C_{vdW}E_{vdW} + solutionterms \quad (25)$$

The lipophilic-lipophilic term is defined as in ChemScore. The hydrogen-bonding term also uses the ChemScore form but is separated into differently weighted components that depend on whether the donor and acceptor are

both neutral, one is neutral and the other is charged, or both are charged. In the optimized scoring function, the first of these contributions is found to be the most stabilizing and the last, the charged-charged term, is the least important. The metal-ligand interaction term (the fifth term in equation 20) uses the same functional form as is employed in ChemScore but varies in three principal ways.

The XP Glide scoring function is presented in equation 26. E_{bind} is defined in in equation 27 and $E_{penalty}$ is defined in equation 28 (Friesner et al., 2006(b)).

$$XP\ GlideScore = E_{coul} + E_{vdW} + E_{bind} + E_{penalty} \tag{26}$$

$$E_{bind} = E_{hyd-enclosur} + E_{hb\_nn\_motif} + E_{hb\_cc\_motif} + E_{PI} + E_{hb\_pair} + E_{phobic\_pair} \tag{27}$$

$$E_{penalty} = E_{desolv} + E_{ligand\_strain} \tag{28}$$

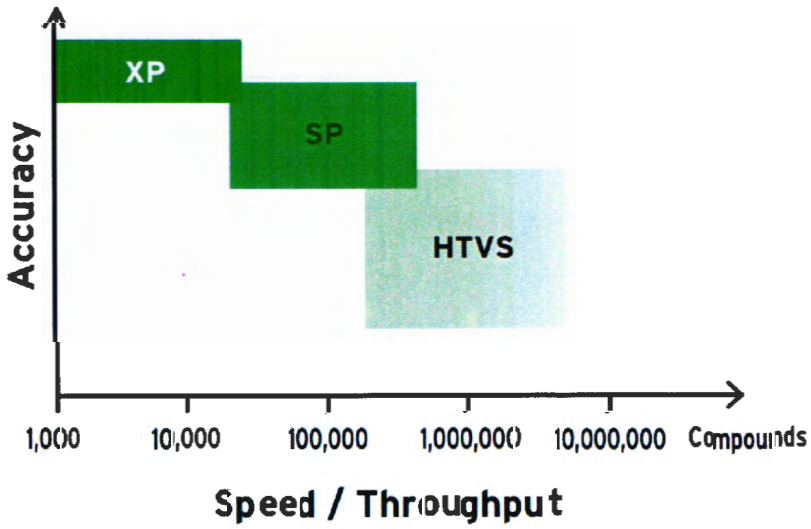


Figure 2.7 Accuracy vs speed of calculations of different Glide scoring functions

Resource: Glide (2010)

4.4) MOE

Molecular Operating Environment (MOE) is a comprehensive software system for Life and Material Science developed by Chemical Computing Group Inc. (CCG). This program strongly supports drug design through molecular simulation, protein structure analysis, data processing of small compounds, molecular docking study of receptor and small molecules (Hongmao, 2016). The MOE dovking architecture consists of four major components: (1) ligand-conformation generation (2) optional pharmacophore filtering (3) ligand placement and scoring in the pocket, and (4) flexible receptor and ligand refinement with re-scoring (Corbeil, Williams and Labute, 2012). Three ligand placement algorithms, Alpha PMI, Alpha Triangle and Triangle Matcher are entirely shape-based, geometrical whole molecule methods.

Affinity dG scoring function, this scoring function estimates the enthalpic contribution to the free energy of binding using a linear function in equation 29:

$$\Delta G = C_H \sum_{hbonds:\, i-j} f_H(r_{ij}) + C_M \sum_{metal-lig:\, i-j} f_M(r_{ij}) + C_I \sum_{ionic:\, i-j} q_i q_j f_I(r_{ij}) + C_B \sum_{contacts:\, i-j} f_B(r_{ij}) \qquad (29)$$

H-bonds (f_H) are between donor (i) and acceptor heavy atoms (j). Metal ligations (f_M) are between transition metal (i) and heteroatoms (–O, –N and –S) (j). Ionic contacts (f_I) are between functional groups (not just ions). This term also includes hydrophobic interactions. Contacts (C_H, C_M, C_I and C_B) are between heavy atoms of receptor and ligand. f_I and f_B functions have 7.5 Å cutoff.

The London dG scoring function implemented in MOE docking to estimate the binding free energy was defined as follow

$$\Delta G = c + E_{flex} + \sum_{h-bonds} c_{hb} f_{hb} + \sum_{m-lig} c_m f_m + \sum_{atoms\ i} \Delta D_i \qquad (30)$$

Here c, c_{hb} and c_m are constants defined as average entropy loss/gain due to rotational/translational motion, H-bond maximum energy and metal ligation maximum energy, respectively. E_{flex} is a topological estimate of ligand entropy. Both f_{hb} and f_m are measures of geometric imperfections of protein–ligand

and metal–ligand interactions. ΔDi is the desolvation energy term which is approximated using a volume integral London dispersion similar to that found in GB/VI. (Corbeil, Williams and Labute, 2012). For London dG Desolvation Model (ΔDi) is calculated according to the equation 30

$$\Delta D_i = c_i R_i^3 \left\{ \iiint_{u \in A \cup B} |u|^{-6}\, du - \iiint_{u \in B} |u|^{-6}\, du \right\} \tag{31}$$

where A and B are the protein and/or ligand volumes with atom i belonging to volume B; Ri is the solvation radius of atom i (taken as the OPLS-AA van der Waals sigma parameter plus 0.5 Å); and ci is the desolvation coefficient of atom i. The coefficients {c, c_{hb}, c_m, c_i} were fitted from ~400 x-ray crystal structures of protein-ligand complexes with available experimental pK_i data. Atoms are categorized into ~12 atom types for the assignment of the ci coefficients. The triple integrals are approximated using Generalized Born integral equations.

2.1.2.2 Molecular dynamics (MD) simulations

Molecular dynamics (MD) simulations are one of the most versatile, widely applied computational techniques and useful tools for the study of biological macromolecules and structure based drug design (Alonso, Bliznyuk and Gready, 2006; Hansson, Oostenbrink and van Gunsteren, 2002; Karplus and McCammon, 2002; Norberg and Nilsson, 2003; Durrant and McCammon, 2011; Dror et al., 2012; Zhao and Caflisch, 2015; De Vivo et al., 2015; Nair and Miners, 2014). They are very valuable for understanding the physical basis of the structure and function of proteins and drug with receptor at different timescales, from fast internal motions to slow conformational changes or protein folding processes (Snow et al., 2005; Karplus and McCammon, 2002; Dror et al., 2012; Perilla et al., 2015). These simulations appear poised to exert a significant impact on how new drugs are found, perhaps even transforming the very process of drug discovery (Borhani and Shaw, 2012, De Vivo et al., 2015).

In MD simulations the forces between atoms and the potential energy of the system are defined by molecular mechanics biomolecular force fields. Molecular mechanic force field is a mathematical expression describing the

dependence of the energy of a system on the coordinates of its particles. This is obtained by solving the second-order differential equations represented by Newton's second law.

$$f_i(t) = m_i a_i(t) = -\frac{\partial V(x(t))}{\partial x_i(t)} \tag{32}$$

where $f_i(t)$ is the net force acting on the *i*th atom of the system at a given point in time t, $a_i(t)$ is the corresponding acceleration, and m_i is the mass. In equation 26, the instantaneous configuration of the system is represented by the vector x(t), which describes the position of the *N* interacting atoms in the Cartesian space ($x = \{x_1, y_1, z_1, x_2, y_2, z_2, ..., xN, yN, zN\}$). The forces acting on each of the system atoms are then estimated from an equation like that shown in Figure 2.8 (Cornell et al., 2003; Durrant and McCammon, 2011; González, 2011; Nair and Miners, 2014; Paquet and Viktor, 2015). These force fields are parameterized to fit quantum-mechanical calculations and experimental spectroscopic data such as neutron, X-ray and electron diffraction, NMR, infrared, Raman and neutron spectroscopy, etc (Cornell et al., 1995; González, 2011; De Vivo et al., 2015; Durrant and McCammon, 2011; Nair and Miners, 2014). In brief, forces between atoms and the potential energy arising from interactions between bonded and non-bonded atoms contribute (Durrant and McCammon, 2011). The forces parameterization involve definition of chemical bonding, atomic angles and dihedral angles (that is, rotations about a bond) are modeled using a sinusoidal function that approximates the energy differences between eclipsed and staggered conformations. The determination of partial atomic charges for calculation of the electrostatic-interaction energies modeled using Coulomb's law, identification of appropriate van der Waals atomic radii modeled using the Lennard-Jones 6-12 potential.

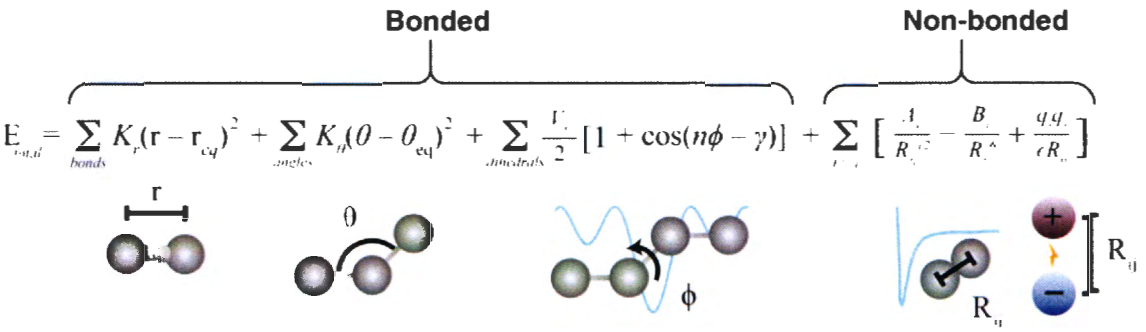


Figure 2.8 An equation used to approximate the atomic forces

Resource: Durrant and McCammon (2011)

In Figure 2.8, the first three terms represent intramolecular interactions of the atoms. The first and second terms, bond stretching and bending contributions share the same functional form, as they are both described by harmonic potentials that control the length of covalent bonds with reference values r_{eq} (the bond length at equilibrium) and θ_{eq} (the bond angle at equilibrium) and force constants K_r and K_θ, respectively. Reasonable values for r_{eq} can be obtained from X-ray diffraction experiments, while the spring constant may be estimated from infrared or Raman spectra. K_r is the bond force constant, whereas K_θ is bond angle force constant. The third term, dihedral angles is usually represented by a cosine function. V_n is the improper dihedral angle. Where ϕ is the torsional angle, γ is the phase, n defines the number of minima or maxima between 0 and 2π.

Non-bonded interactions, fourth term represents van der Waals and electrostatic interactions between atoms. These contributions act on every pair of atoms in the system that is not already covered by the bonded counterpart. Van der Waals interactions between two atoms arise from the balance between repulsive and attractive forces. The 12-6 Lennard-Jones (LJ) potential is very often used to represent these interactions. A and B are are constants that depend on the van der Waals. ε is the effective dielectric constant. q_i and q_j are the partial charges of a pair of atoms. The energy is expressed as an inverse power function of the distance between the considered atoms, R_{ij}.

2.1.2.3 Binding free energy calculations

Molecular mechanics (MM) has broad applications in studying biological systems for its simplicity and efficiency (Wang, Hou and Xu, 2006). From a

medical perspective, one of the ultimate goals in CAMD method is the accurate prediction of ligand-binding affinities to a macromolecular target (Karaman and Sippl, 2015; Wichapong, et al.. 2010(a); 2014(b)), which can facilitate and speed the routine identification of new candidates in early stage drug discovery projects (Gilson and Zhou, 2007; Hayes and Leonidas, 2010). Additionally traditional scoring methods, more sophisticated and demanding calculations of binding free energy (BFE) such as linear interaction energy (LIE) (Aqvist, Medina and Samuelsson, 1994), molecular mechanic/Poisson–Boltzmann (Generalized Born) surface area (MM-PB(GB)SA) (Srinivasan et al., 1998), free-energy perturbation (FEP), thermodynamic integration (TI) (Kirkwood, 1935; Straatsma and McCammon, 1991) and water swap (Woods et al., 2011(a); 2014(b)) have been developed. Although some studies show successful examples of lead optimization using FEP and TI methods (Archontis et al., 2005; Pearlman, 2005; Homeyer and Gohlke, 2012), they are still rarely applied in the drug discovery process, mainly due to the high computational costs. Indeed, the success of exhilarating the use of these techniques (especially in industrial settings) depends mainly on the minimization of human resources for data analysis. The binding free energy calculations were calculated using the thermodynamics cycle in Figure 2.9. The binding free energies (ΔG_{bind}) were obtained as shown in Equations (33) and (34).

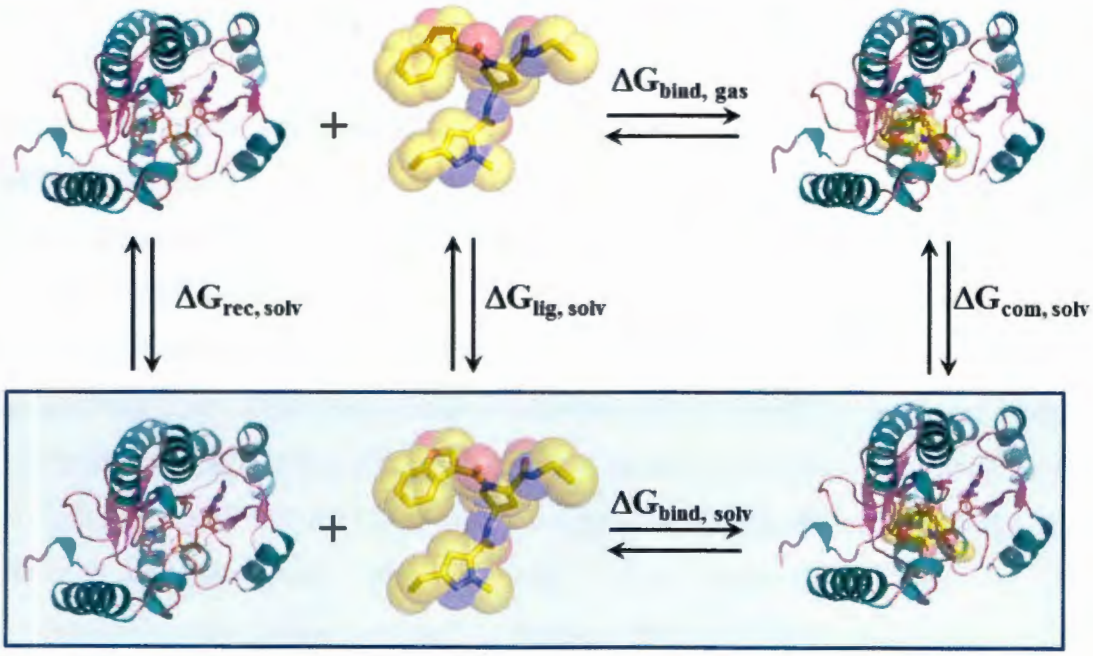


Figure 2.9 Thermodynamic cycles for binding free energy calculations for a protein–ligand complex

$$\Delta G_{bind} = G_{com} - (G_{rec} + G_{lig}) \quad (33)$$

$$\Delta G_{bind} = \Delta H - T\Delta S \quad (34)$$

$$\Delta H = \Delta E_{MM} + \Delta G_{solv} \quad (35)$$

Where G_{com}, G_{rec} and G_{lig} are the free energies of the complex, receptor and ligand, respectively. In general, the binding free energy composes of an enthalpic (ΔH) and an entropic contribution ($-T\Delta S$). The enthalpic contribution (ΔH) contains the gas phase molecular mechanics energy (ΔE_{MM}) calculated with a sander module and the solvation free energy (ΔG_{solv}) calculated with the PBSA program of the AMBER suite as shown in Equation (36).

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{solv} - T\Delta S \quad (36)$$

ΔE_{MM} is divided into non-covalent van der Waals component (ΔE_{vdw}), electrostatic energies component (ΔE_{ele}) and internal (bond, angle and dihedral) energies (ΔE_{INT}) in Equation (37).

$$\Delta E_{MM} = \Delta E_{vdw} + \Delta E_{ele} + \Delta E_{INT} \quad (37)$$

ΔG_{solv} is demined from the summation of electrostatic solvation energy (polar contribution) and non-electrostatic solvation energy (non-polar contribution). The electrostatic solvation energy in this study was calculated using two different methods, Poisson-Boltzmann (PB) and Generalized Born (GB) methods. For GB methods, different GB models were developed and implemented in AMBER program. Therefore, different of GB solvation models, igb1, igb2, igb5, igb7 and igb8 were selected to calculate the binding free energy in this study. The first selected GB model, igb1 is GB model with parameters developed by Tsui and Case (Tsui and Case, 2001). For two GB solvation models, igb2 and igb5 were developed. (Onufriev, Bashford and Case, 2004). The PB and GB solvation models were compared. The dielectric constant for GB calculations of implicit solvent and solute was set to 80 and 1, respectively. The default value of solvent probe radius (1.4 Å) was selected.

$$\Delta G_{solv} = \Delta G_{solv,\ polar\ (PB/GB)} + \Delta G_{solv,\ non\text{-}polar\ (SA)} \quad (38)$$

For non- electrostatic solvation energy, hydrophobic contribution (non-polar contribution) to the solvation free energy was estimated by calculating the Solvent Accessible Surface Area (SASA) with Molsurf method as shown in equation 39. Where, γ is a surface tension parameter and is a parameterized value. This work, defaults values of γ and β were used.

$$\Delta G_{solv,\ non\text{-}polar\ (SA)} = \gamma \cdot SASA + \beta \quad (39)$$

The estimated entropy changes based on the number of rotatable bonds (N_{Rot}) of ligands have been reported. Server reported have been applied this estimations to get the entropy changes of compounds for docking and binding free energy calculations. The estimated entropy change was approximated as shown in equation 40 (Raha and Merz, 2004; Hayik, Dunbrack, and Merz, 2010).

$$T\Delta S N_{Rot} = \text{number of rotatable bonds} \times 1.0\ \text{kcal/mol} \quad (40)$$

2.1.3 Virtual screening

The identification of lead compounds showing pharmacological activity against a biological target and the progressive optimization of the pharmacological properties and potency of these compounds are the focal points of early-stage drug discovery. To this end, the pharmaceutical industry has adopted the experimental screening of large libraries of chemicals against a therapeutically-relevant target (high throughput screening or HTS) as a means to identify new lead compounds. Through HTS, active compounds, antibodies or genes, which modulate a particular biomolecular pathway, may be identified; these provide starting points for drug discovery and for understanding the role of a particular biochemical process in biology. Although HTS remains the method of choice for drug discovery in the pharma industry, the various drawbacks of this method, namely the high cost, the time-demanding character of the process as well as the uncertainty of the mechanism

of action of the active ingredient have led to the increasing employment of rational, drug design with the use of computational methods.

Computer-aided drug discovery thought virtual screening (VS) has recently had important successes: new biologically-active compounds have been predicted along with their receptor-bound structures and in several cases the achieved hit rates (ligands discovered per molecules tested) have been significantly greater than with HTS (Lavecchia and Di Giovanni, 2013; Benod et al., 2013; Cheng et al., 2012; Andricpoulo, Salum and Abraham, 2009). Moreover, while it is rare to deliver lead candidates in the nM regime through VS, several reports in the recent literature describe the identification of nM leads directly from VS; these strategies will be discussed herein (Heifetz et al., 2013; Schröder et al., 2013; Kolb et al., 2009). Therefore, computational methods play a prominent role in the drug design and discovery process within the context of pharmaceutical research.

Virtual screening (VS) is a powerful technique for identifying hit molecules as starting points for medicinal chemistry. VS is a detailed, knowledge-driven, compound database searching approach that attempts to find novel compounds and chemotypes with a required biological activity as alternatives to existing ligands or sometimes to make first inroads into finding ligands for unexplored putative drug targets for which crystal structures, solution structures or high confidence homology models are available. VS is a knowledge-driven approach. The quality and the amount of information regarding the system under inspection is a critical factor when designing a computer-assisted drug design experiment (Klebe, 2006). Knowledge about the substrates may be the starting point to retrieve similar compounds by 2D/3D similarity or pharmacophore searches. In cases where the target structure is available, docking methods that sample ligand poses with respect to the receptor binding site can be used. An available VS methodologies and to categorize them into various groups were summarized in Figure 2.10 and Table 2.4 (Cheng et al., 2012; Muegge and Oloff, 2006; Gohlke and Klebe, 2002; Ripphausen, Nisius and Bajorath, 2011; Sun, 2008; Kar and Roy, 2013; Bielska et al., 2011; Lionta et al., 2014).

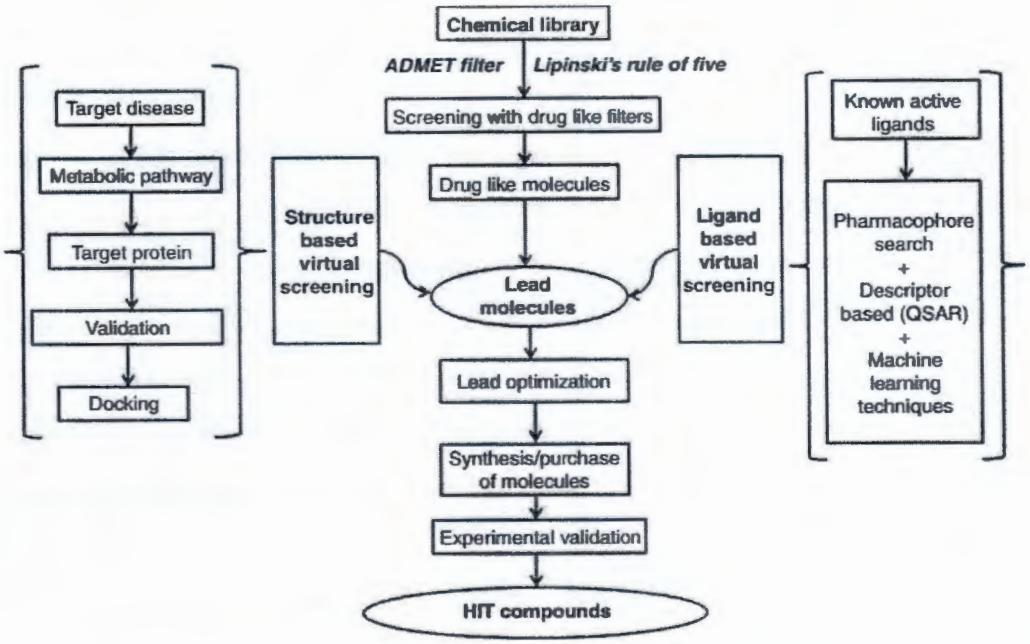


Figure 2.10 Schematic representation of commonly practiced VS process

Resource: Kar and Roy (2013)

Table 2.4 Classification of virtual screening methods based on the amount and type of information available about the system under inspection

	Known ligand(s)	**Unknown ligands**
Known structure of target or close homologue	structure-based virtual screening: protein-ligand docking	*de novo* structure-based virtual screening: protein-ligand docking
Unknown target structure	ligand-based virtual screening: few ligands: similarity searches several ligands: pharmacophore searches	virtual screening cannot be applied

Resource: Bielska et al. (2011)

2.1.4 Quantum chemical calculations

2.1.4.1 Brief overview of quantum chemical methods

Solving the Schrödinger equation for a molecule is a very complex problem. There are many degrees of freedom involved: for a molecule with N nuclei and n electrons we have 3(N + n) variables. The general strategy is based on the

principle of divide and conquers! This means divide the large problem into several smaller problems, which can be solved separately. This general theme is repeated many times throughout the quantum chemistry. As we will see, some errors are always introduced by dividing the system, but those can be subsequently corrected after the smaller, easier problems are dealt (Hehre, 2003, Kubelka, 2014).

2.4.1.2 Quantum mechanics and the Schrodinger equation

Computational quantum chemistry is concerned with predicting the properties of atomic and molecular systems. It is based upon the fundamental laws of quantum mechanics and uses a variety of mathematical transformation and approximation techniques to solve the fundamental equations (Szabo and Ostlund, 1989).

The quantum mechanical system is completely described by its wave function Ψ. The product of Ψ with its complex conjugate (Ψ*Ψ, often written as $|\Psi|^2$) is interpreted as the probability distribution of the particle. For this reason $|\Psi|^2$ has to be normalized to 1. The wave function is found by solving the Schrodinger equation:

$$\hat{H}\psi(r) = E\psi(r) \tag{41}$$

where E is the energy of the particle, and is the Hamiltonian operator:

$$\hat{H} = -\frac{\hbar^2}{2m}\nabla^2 + \hat{V}(r) \tag{42}$$

m is the mass of the particle, (Planck's constant divided by 2π, is the potential field in which the particle is moving and is the Laplacian operator:

For a molecular system, the wave function Ψ is a function of the positions of the electrons and the nuclei within the molecule, which we will designate as r and R, respectively.

$$\Psi = \Psi(r, R) \tag{43}$$

These symbols are ashorthand for the set of component vectors describing the position of each particle. We'll use subscripted versions of r_i and R_I to denote the vector corresponding to a particular electron i or nucleus I. Note that electrons are treated individually, while each nucleus is treated as an aggregate; the component nucleons are not treated individually.

$$\hat{H} = -\sum_{i}^{electrons} \frac{\hbar^2}{2m_e}\nabla_e^2 - \sum_{I}^{nuclei} \frac{\hbar^2}{2M_I}\nabla_N^2 - \sum_{i}^{electrons}\sum_{I}^{nuclei} \frac{Z_I e'^2}{r_{iI}} + \sum_{I<J}^{nuclei} \frac{Z_I Z_J e'^2}{r_{IJ}} + \sum_{i<j}^{electrons} \frac{e'^2}{r_{ij}} \quad (44)$$

where: $r_{iI} = |r_i - R_I|$, $r_{IJ} = |R_I - R_J|$, $r_{ij} = |r_i - r_j|$, $e' = e/4\pi\varepsilon_0$, e is elementary charge and ε_0 vacuum permittivity.

In the above equation, the individual terms represent, respectively: kinetic energy of electrons, kinetic energy of the nuclei, electron-nuclear attraction, nuclear repulsion, electron repulsion.

The fundamental equations of quantum chemistry are usually expressed in atomic units, introduced at the beginning. In atomic units the Schrodinger equation simplifies

$$\hat{H} = -\sum_{i}^{electrons} \frac{1}{2}\nabla_e^2 - \sum_{I}^{nuclei} \frac{1}{2M_I}\nabla_N^2 - \sum_{i}^{electrons}\sum_{I}^{nuclei} \frac{Z_I}{r_{iI}} + \sum_{I<J}^{nuclei} \frac{Z_I Z_J}{r_{IJ}} + \sum_{i<j}^{electrons} \frac{1}{r_{ij}} \quad (45)$$

2.4.1.3 Linear combination of atomic orbitals

Linear combination of atomic orbitals (LCAO) is a simple method of quantum chemistry that yields a qualitative picture of the molecular orbitals (MOs) in a molecule. The overall wavefunction of molecule (Ψ) is treated as a product of *molecular orbitals* (ψ):

$$\Psi = |\psi_1\psi_2\psi_3....\psi_n| \quad (46)$$

One-electron molecular orbitals ψ are built up as linear combinations of atomic orbitals ϕ according to

$$\phi = \sum_{i=1}^{N} a_i\psi_i \quad (47)$$

In the LCAO approximation, each MO is treated as being made up of the AOs of the atoms in the molecule. The set of N atomic-orbital basis functions ϕ_i is called the "basis set" and each "basis function" has associated with it some coefficient a_i for any given MO. Use the variational principle to find the optimal coefficients. For a given one-electron orbital, two orbitals are permitted to interact then the general expression for the molecular orbital expressed as a linear combination of atomic orbitals.

$$\psi = c_1\phi_1 + c_2\phi_2 \tag{48}$$

The expression for E now becomes :

$$E = \frac{\int\left(\sum_i a_i^*\varphi_i^*\right)H\left(\sum_j a_j\varphi_j\right)dr}{\int\left(\sum_i a_i^*\varphi_i^*\right)\left(\sum_j a_j\varphi_j\right)dr} = \frac{\sum_{ij} a_i^* a_j \int \varphi_i^* H\varphi_j dr}{\sum_{ij} a_i^* a_j \int \varphi_i^*\varphi_j dr} = \frac{\sum_{ij} a_i^* a_j H_{ij}}{\sum_{ij} a_i^* a_j S_{ij}} \tag{49}$$

H_{ij} and S_{ij} are "resonance" and "overlap" integrals. Hamiltonian matrix element is define by

$$\begin{aligned} H_{11} &= \int \phi_1 \hat{H}\phi_1 dv \\ H_{12} &= \int \phi_1 \hat{H}\phi_2 dv = H_{21} \end{aligned} \tag{50}$$

Overlap matrix element is define by

$$\begin{aligned} S_{11} &= \int \phi_1\phi_1 dv = 1 \\ S_{12} &= \int \phi_1\phi_2 dv = S_{21} \end{aligned} \tag{51}$$

Note the symmetry $H_{12} = H_{21}$ and $S_{12} = S_{21}$ in these matrix elements or integrals.

Secular equations can be defined as:

$$\begin{aligned} (H_{11} - E)c_1 + (H_{12} - ES_{12})c_2 &= 0 \\ (H_{21} - ES_{21})c_1 + (H_{22} - E)c_2 &= 0 \end{aligned} \tag{52}$$

These equation have a "trivial" and useless solution $c_1 = c_2 = 0$. The condition that there should exist a nontrivial solution of these equations is that the secular determinant should be zero:

$$\begin{vmatrix} (H_{11} - E) & (H_{12} - ES_{12}) \\ (H_{21} - ES_{21}) & (H_{22} - E) \end{vmatrix} = 0 \tag{53}$$

Everything in this equation is a known number except E. The equation is therefore an equation for E. In the present case where the molecular orbital was a linear combination of just two atomic orbitals it is a quadratic equation and has two solutions for E. These two values of E are the molecular orbital energies. We always get the same number of molecular orbitals as atomic orbitals we start with.

If we multiply out the secular determinant we obtain:

$$\begin{aligned} &(H_{11} - E)(H_{22} - E) - (H_{12} - ES_{12})(H_{21} - ES_{21}) = 0 \\ &E^2 - (H_{11} + H_{22} - 2H_{12}S_{12})E + (H_{11}H_{22} - H_{12}^{\ 2}) = 0 \end{aligned} \tag{54}$$

which you should now be able to recognise as a quadratic equation for E. It has two solutions E_1 and E_2. These are the two molecular orbital energies.

These equations now have a nontrivial solution and we find the ratio c_2/c_1 by solving them. The absolute value of the coefficients c_1 and c_2 can only be found by using the normalization condition:

$$\begin{aligned} &\int \psi^* \psi dv = 1 \\ &\int (c_1\phi_1 + c_2\phi_2)(c_1\phi_1 + c_2\phi_2) dv = 1 \\ &c_1^{\ 2} \int \phi_1\phi_1 dv + c_1c_2 \int \phi_1\phi_2 dv + c_2c_1 \int \phi_2\phi_1 dv + c_2^{\ 2} \int \phi_2\phi_2 dv = 1 \\ &c_1^{\ 2} = 2c_1c_2S_{12} + c_2^{\ 2} = 1 \end{aligned} \tag{55}$$

2.4.1.4 Basis set

In quantum chemistry, the "basis set" often refers to any set of (usually nonorthogonal) one-particle functions used to build molecular orbitals. A basis set is a mathematical description of the orbitals within a system used to perform the theoretical calculation. Larger basis sets approximate more accurately the orbitals

by imposing fewer restrictions on the locations of the electrons in space (Eliav, 2016; Skylaris, 2016).

1) Minimal basis sets

Minimal basis sets contain the minimum number of basis functions that are needed for each atom. For example,

H: 1s

C: 1s, 2s, $2p_x$, $2p_y$, $2p_z$

Minimal basis sets use fixed size atomic type orbitals. The STO-3G basis set is a minimal basis set (through it is not the smallest possible basis set). It uses three 80 gaussian primitives per basis function ("3G"). "STO" stands for "Slater type orbitals", and the STO-3G basis set approximates Slater orbitals with 80aussian functions. Their general definition is

$$S_{nlm}^{\xi}(r,\vartheta,\varphi) = Nr^{n-1}e^{-\xi r}Y_l^m(\vartheta,\varphi) \tag{56}$$

with N being a normalization factor and Y ml being the spherical harmonics.

Gaussian type orbitals (GTOs) or Gaussian type functions (GTF), were proposed by S.F. Boys in 1950. Linear combinations of GTO's are used to approximate STOs (which are themselves approximations). STOs can be approximated as linear combinations of Gaussian orbitals. Gaussian type orbitals (GTOs) are defined as

$$\left| g(r,\theta,\phi) = \left[\frac{2^{(2n+1.5)}}{(2n-1)!\sqrt{\pi}} \right]^{0.5} \xi^{(2n+1)/4} r^{(n-1)} e^{(-\alpha r^2)} Y_l^m(\theta,\phi) \right| \tag{57}$$

Different types of basis sets use different numbers and types of GTOs. For the different types of basis sets were shown as below.

2) Split valence basis sets

One way to increase the size of a basis set is to take more basis functions per atom. Split valence basis sets, such as 3-21G and 6-31G basis sets, have two

(or more) sizes of basis function for each valence orbital. For example, in the above 3-21G and 6-31G basis sets we have:

H: 1s, $1s^1$

C: 1s, 2s, $2s^1$, $2p_x$, $2p_y$, $2p_z$, $2p_x^1$, $2p_y^1$, $2p_z^1$

Here the primed and unprimed orbitals differ in size. The double zeta valence basis sets form molecular orbitals from the linear combinations of two sets of functions for each atomic valence orbital. Similarly, triple split valence basis sets such as 6-311G, use three sets of contracted functions for each valence orbital type.

3) Polarized basis sets

Split valence basis sets could be improved by adding orbitals with different shapes. Polarized basis sets add orbitals with angular momentums going beyond of requirement for the proper description of the ground state of each atom at the HF level. For example, polarized basis sets add to carbon atoms and some of them add to hydrogen atoms. Examples for polarized basis sets are the 6- 31G(d) and the 6-311G(d, p) basis sets.

4) Diffused functions

Basis sets with additional diffuse functions are large by size versions of s- and p-type split valence basis sets. Diffuse orbitals occupy a larger region of space. Basis sets with diffuse functions are important for systems where electrons may be far from the nucleus. One example for diffuse basis function is the 6-311+G(d, p) basis set. The table below summarizes the number of basis functions of each and the number of primitive gaussian functions they are constructed of.

Table 2.5 summarizes the number of basis functions

Atomic number	1-2	3-10
n_1-$n_2n_3n_4$G Basis function	1s constructed of n_2 gaussians 1s′ constructed of n_3 gaussians 1s″ constructed of n_4 gaussians	1s constructed of n_1 gaussians 2s, 2p each constructed of n_2 gaussians 2s′, 2p′ each constructed of n_3 gaussians 2s″, 2p″ each constructed of n_4 gaussians
n_1-$n_2n_3n_4$G(d)	n_1-$n_2n_3n_4G$ basis function plus nothing	n_1-$n_2n_3n_4G$ basis function plus 3d each constructed of 1 gaussian
n_1-$n_2n_3n_4$G(d,p)	n_1-$n_2n_3n_4G$ basis function plus 2p each constructed of 1 gaussians	n_1-$n_2n_3n_4G(d)$ basis function plus nothing
n_1-$n_2n_3n_4$+G	n_1-$n_2n_3n_4G$ basis function plus nothing	n_1-$n_2n_3n_4G$ basis function plus 3s, 3p each constructed of 1 gaussian
n_1-$n_2n_3n_4$++G	n_1-$n_2n_3n_4G$ basis function plus 2s each constructed of 1 gaussian	n_1-$n_2n_3n_4G$ basis function plus nothing

Resource: Eliav (2016)

2.2 Rational drug design procedures

2.2.1 Rational design of anti-TB agents

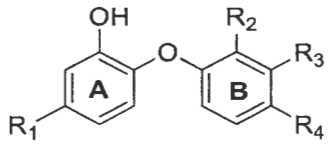
2.2.1.1 Molecular modeling and CAMD of InhA inhibitors

1) Diphenyl ether derivatives

1.1) Structure and biological activity

Fifty-four diphenyl ether derivatives (Sullivan et al., 2006; am Ende et al, 2008; Freundlich et al., 2009; Luckner et al., 2010; Pan, 2014) listed in Table 1 were selected for this work. Fifty-two compounds (cpd. **1-52**) were used to build the CoMSIA model. The experimentally obtained IC_{50} values of each compound for InhA inhibition were converted to the corresponding log ($1/IC_{50}$) values, and used as dependent variables for the QSAR model. The chemical structures of these compounds were constructed using standard tools available in GaussView 3.07 program and were then fully optimized using the ab initio quantum chemical method (HF/3-21G) implemented in the Gaussian 09 program. The compounds were divided into a training set of 43 compounds, and a test set of 9 compounds for model development and validation, respectively. The test set was randomly selected based on a structural diversity and wide range of activity in the data sets.

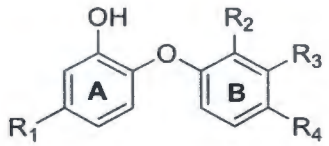
Table 2.6 Chemical structures and activities for InhA inhibition of 52 diphenyl ether derivatives



Cpd.	R_1	R_2	R_3	R_4	IC_{50}(nM)	Log(1/IC_{50})
1	Cl	Cl	H	Cl	1100	5.96
2	CH_3	Cl	H	Cl	800	6.10
3	CH_2Cy	Cl	H	Cl	110	6.96
4*	CH_2CH_3	Cl	H	Cl	120	6.92
5	$(CH_2)_2CH_3$	Cl	H	Cl	91	7.04
6	$(CH_2)_3CH_3$	Cl	H	Cl	55	7.26
7	$(CH_2)_2CH(CH_3)_2$	Cl	H	Cl	63	7.20
8	$CH_2CH(CH_3)CH_2CH_3$	Cl	H	Cl	130	6.89
9	CH_2(2-pyridyl)	Cl	H	Cl	29	7.54
10*	CH_2(3-pyridyl)	Cl	H	Cl	42	7.38
11	CH_2(4-pyridyl)	Cl	H	CN	75	7.12
12	*o*-CH_3-Ph	Cl	H	Cl	1300	5.89
13	*m*-CH_3-Ph	Cl	H	Cl	870	6.06
14	CH_2Ph	Cl	H	Cl	51	7.29
15	CH_2CH_2Ph	Cl	H	Cl	21	7.68
16*	$(CH_2)_3Ph$	Cl	H	Cl	50	7.30
17	$(CH_2)_5CH_3$	H	H	H	11	7.96
18	CH_2CH_3	H	H	H	2000	5.70
19	$(CH_2)_3CH_3$	H	H	H	80	7.10
20	$(CH_2)_4CH_3$	H	H	H	17	7.77
21	$(CH_2)_7CH_3$	H	H	H	5	8.30
22	$(CH_2)_{13}CH_3$	H	H	H	150	6.82
23*	$(CH_2)_5CH_3$	NO_2	H	H	180	6.74
24	$(CH_2)_5CH_3$	H	NO_2	H	48	7.32
25	$(CH_2)_5CH_3$	H	H	NO_2	90	7.05
26*	$(CH_2)_5CH_3$	NH_2	H	H	62	7.21
27	$(CH_2)_5CH_3$	H	NH_2	H	1090	5.96
28	$(CH_2)_5CH_3$	H	H	NH_2	55	7.26
29	$(CH_2)_5CH_3$	Br	H	H	10	8.00
30*	$(CH_2)_5CH_3$	CF_3	H	H	29.7	7.53
31	$(CH_2)_5CH_3$	F	H	H	12.1	7.92
32	$(CH_2)_5CH_3$	I	H	H	44.6	7.35
33	$(CH_2)_5CH_3$	OH	H	H	48	7.32
34	$(CH_2)_5CH_3$	CN	H	H	235.6	6.63
35	$(CH_2)_5CH_3$	Cl	H	H	49.5	7.31
36*	$(CH_2)_5CH_3$	CH_3	H	H	50.7	7.29

*test set

Table 2.6 Chemical structures and activities for InhA inhibition of 52 diphenyl ether derivatives (continued)



Cpd.	R_1	R_2	R_3	R_4	IC_{50}(nM)	Log(1/IC_{50})
37	$(CH_2)_5CH_3$	$NHCOCH_3$	H	H	1550	5.81
38	$(CH_2)_5CH_3$	H	H	$NHCONH_3$	1300	5.89
39	$(CH_2)_5CH_3$	$NHCOCO_2H$	H	H	2360	5.63
40	$(CH_2)_5CH_3$	H	$NHCOCO_2H$	H	580	6.24
41	$(CH_2)_5CH_3$	H	H	$NHCOC2O_2H$	1930	5.71
42*	$(CH_2)_5CH_3$	H	NHCO-Isoxazole	H	1220	5.91
43	$(CH_2)_5CH_3$	CH_2-*N*-CH_3-piperazine	H	H	1315	5.88
44	$(CH_2)_5CH_3$	H	H	CH_2-*N*-CH_3-piperazine	306	6.51
45	CH_2CH_2Ph	H	H	H	144.3	6.84
46	CH_2CH_2Ph	CH_3	H	H	360.1	6.44
47	CH_2Ph	Cl	H	H	20.08	7.70
48	CH_2Ph	H	H	H	49.6	7.30
49	CH_2Ph	CH_3	H	H	56.4	7.25
50	$CH_2CH_2CH_2OH$	CH_3	H	H	4326	5.36
51	$OCH_2CH_2OCH_3$	H	H	H	253.1	6.60
52*	$O(CH_2)_4CH_3$	H	H	H	94.2	7.03
53	$CH(CH_3)CH_3CH_3$	Cl	H	Cl	96	7.00
54	$(CH_2)_5CH_3$	CH_3	O	NO_2	50	7.30

*test set

1.2) 3D-QSAR based on CoMSIA study

The binding mode of compound **21** representing the best active compound for the InhA inhibition was taken from the X-ray structure (PDB code 2B37) (Sullivan et al., 2006) and used as a template for molecular alignment. The pharmacophore alignment module with the GALAHAD fit implemented in SYBYL 8.0 program was employed to align all compounds to the molecular template. SYBYL 8.0 molecular modeling software was then used to construct CoMSIA models (Klebe, Abraham, and Mietzner, 1994). Five CoMSIA descriptors including steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor fields were calculated using an sp^3 carbon probe atom, with a formal charge of +1, which

was placed at the intersections in a grid spacing of 2 Å. CoMSIA descriptors were set as independent variables and log ($1/IC_{50}$) values were used as dependent variables in the partial least square (PLS) analysis to derive a linear relationship between molecular descriptors and activities. The cross-validation was performed using the leave-one-out method with a 2.0 kcal/mol column filter to minimize the influence of noisy columns. A final non cross-validated analysis with the optimal number of components was sequentially performed and was then employed to analyse the results. The non-cross-validated correlation coefficient (r^2) and the leave-one-out (LOO) cross-validated correlation coefficient (q^2) were used to evaluate the predictive ability of the CoMSIA model. To estimate the predictive abilities of the best CoMSIA model, external validation using several statistical data was employed. According to Golbraikh and Tropsha (Golbraikh and Tropsha, 2002), the best CoMSIA model is considerably acceptable if they satisfy all of the following criteria: $q^2 > 0.50$, $r^2 > 0.60$, and $0.85 \leq k \leq 1.15$.

1.3) Molecular docking calculations

The X-ray crystal structure of diphenyl ether in complex with InhA (PDB code 2X23) (Luckner et al., 2010) was used as a template for molecular docking calculations. Two of docking programs were used to create the binding mode of diphenyl ether derivatives in this study.

Docking calculations for all 54 diphenyl ether derivatives were carried out by the Autodock 4.02 program using Lamarckian Genetic Algorithm (LGA) (Morris et al., 1998). Docking parameters were used as default values, except for the number of docking runs, which was set to 50. The parameters of the docking calculations were validated by successfully reproducing the X-ray conformation of the ligand in the PDB structure 2X23, as well as its orientation in the binding pocket. The RMSD value between original and docked coordinates was lower than 1 Å and therefore acceptable. For all 54 candidate compounds, the ligand pose with the lowest final docked energy was selected as the best binding mode of these potential InhA inhibitors.

Molecular docking calculations using Molecular Operating Environment (MOE) docking program were used. The high-resolution X-ray crystal structure of diphenyl ether/InhA complex was obtained from Protein Data Bank (PDB

code 2X23) (Luckner et al., 2010). For molecular docking parameters, alpha PMI placement strategies, affinity ΔG scoring function and 50 run were used.

1.4) MD simulations

Two sets, set 1 and set2 of diphenyl ether were selected to perform MD simulations in this study. Set 1, six compounds (cpd. **6**, **14**, **21**, **24**, **53** and **54**) were selected to investigate the effect of R_1, R_2, R_3 and R_4 in InhA binding pocket. Complex InhA structures of these compounds as generated by the previous docking calculations using MOE were used as initial coordinates for MD simulations. The AMBER12 software suite was used for all MD simulations to classically describe all relevant interactions within the system: InhA protein was described by the ff03 force field (Duan et al., 2003) while NAD^+ and diphenyl ether inhibitors were described by the general AMBER force field (GAFF) (Wang et al., 2004(b); 2006(c)). All missing hydrogen atoms of InhA were added using the LEaP module. To obtain the partial atomic charges of diphenyl ether derivatives and NAD^+, the geometry optimization and electrostatic potential calculation of each compound was first calculated at the HF/6-31G* level using the Gaussian 09 program. Then, RESP partial charges (Cornell et al., 1993; Bayly et al., 1993; Wang, Cieplak, and Kollman, 2000; Gavernet et al., 2010; Li et al., 2012) were assigned using the Antechamber module implemented in AMBER12. Each complex structure was solvated by TIP3P (Mahoney and Jorgensen, 2000) waters in a truncated octahedral box extending up to 10 Å from each solute species. Five Na^+ cations were added to neutralize the charge in each system. Non-bonded cut-off was set to 10 Å. To relieve bad steric interactions that originated from addition of the water molecules and ions, the systems were first minimized with atomic positions of all solute species restraint (using a force constant of 500 kcal/mol $Å^2$). Then, the whole systems were fully minimized without restraining conditions. The solvated systems were gradually warmed up from 0 K to 300 K in the first 20 ps followed by maintaining the temperature at 300 K during the last 10 ps. An integration time-step of 2 fs was used in a constant volume boundary. After minimization and heating, the position-restrained dynamics simulations were performed for 70 ps at 300 K under an isobaric condition to relax the positions of the solvent molecules. A weak force constant of 10 kcal/mol $Å^2$ restraint on solute species was also applied for each simulation. Then, a 5 ns production MD simulation without

restraints was performed on each system at a constant temperature of 300 K under isobaric condition. The Particle Mesh Ewald (PME) (Darden York and Pedersen, 1993) was applied to treat the long-range electrostatic interactions with a periodic boundary condition during the MD simulations. The cut-off distance for the long-range van der Waals interaction was set to 8 Å. The SHAKE (Ryckaert, Ciccotti, and Berendsen, 1977) method was applied to constrain the bond lengths of hydrogen atoms attached to heteroatoms. Coordinates and energy outputs during the MD simulation were collected every 2 ps. Finally, the root-mean-square deviations (RMSDs) of the InhA protein, NAD^+ and diphenyl ether ligand, respectively, were analyzed along the MD trajectory relative to the initial structures to determine the stability of the system. The binding free energies were calculated to evaluate the binding affinities of diphenyl ether derivatives in InhA binding pocket. For set 2, MD simulations were performed on compounds **17**, **18**, **19** and **29**, which are representative compounds that cover a wide range from highly active (**17** and **29**) to less active compounds (**18**) among the candidate series in this study. Compound **19** was also included in the simulations to represent a moderate inhibitory activity. The simulations details were similar to set 1, except the initial coordinate of InhA-inhibitors complexes. The initial complexed in set 2 were obtained from molecular docking using Autodock 4.2 program.

1.5) Binding free energy calculation

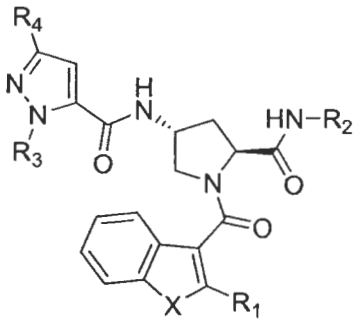
The free energy of binding between InhA and diphenyl ether inhibitors were calculated using the Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA) (Homeyer and Gohlke, 2012; Wang et al., 2001(a), 2006(c); Hou et al., 2011) and Normal-mode (Kaledin et al., 2011) methods. For MM-PBSA calculation, 125 snapshots were generated for each complex from the last 1 ns of MD trajectory with an interval of 8 ps. The entropic contribution to the binding free energy was estimated using normal-mode analysis with AMBER Nmode module. Due to a highly computational cost in the entropy calculation, the residues around the ligand (less than 12 Å) were only considered as the receptor for normal mode calculations (Hou et al., 2011; Xue et al, 2012). For this calculation, 50 snapshots were extracted from the last 1 ns of MD trajectory with an interval of 20 ps.

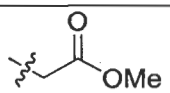
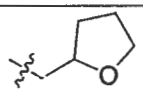
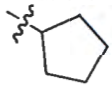
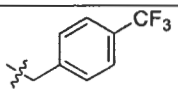
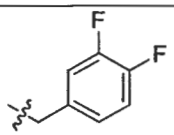
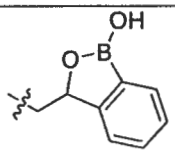
2) Benzofuran pyrrolidine pyrazole derivatives

2.1) Structure and biological activities

Thirty-four benzofuran pyrrolidine pyrazole derivatives used for CoMFA and CoMSIA studies were identified from the published literature (Encinas et al., 2014). Chemical structures and experimental biological activities in terms of MIC_{90} and IC_{50} values of these compounds are shown in Table 3.2. MIC_{90} and IC_{50} values were nominally converted into log ($1/MIC_{90}$) and log ($1/IC_{50}$) values for CoMFA and CoMSIA studies. Based on the diversity of structures and wide range of activities, the data set of compounds was divided into 30 training set compounds for final model development and 4 test set compounds for model validation. All chemical structures of benzofuran pyrrolidine pyrazole derivatives were constructed using the standard tools available in the GaussView 3.07 program and were then fully optimized using the HF/6-31G method implemented in the Gaussian 09 program. The harmonic vibrational frequencies of the optimized geometries have also been calculated. All elements in the calculated Hessian matrix are positive, which indicate that the structures are true minima on the potential energy surface.

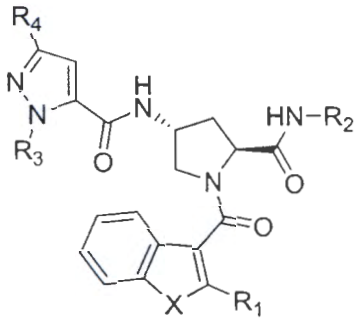
Table 2.7 Chemical structures and activities against InhA and *M. tuberculosis* of thirty four benzofuran pyrrolidine pyrazole derivatives

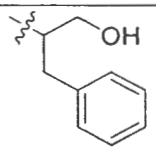
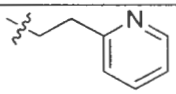
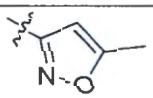
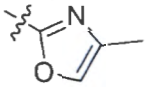
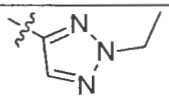
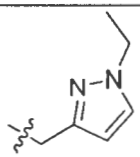
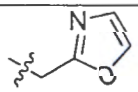


Cpd.	X	R_1	R_2	R_3	R_4	IC_{50} (μM)	MIC_{90} (μM)	$log(1/IC_{50})$	$log(1/MIC_{90})$
1	O	H	Et	Me	Et	0.034	8.00	7.47	5.10
2*	O	H		Me	Et	0.005	0.50	8.30	6.30
3	O	H	H	Me	Et	0.012	3.00	7.92	5.52
4	O	H	CH_2CF_3	Me	Et	0.046	4.00	7.34	5.40
5	O	H	$CH_2CH_2CH_3$	Me	Et	0.021	15.60	7.68	4.81
6	O	H	CH_2CH_2OMe	Me	Et	0.014	4.00	7.85	5.40
7*	O	H	CH_2CH_2COOEt	Me	Et	0.022	4.00	7.66	5.40
8	O	H		Me	Et	0.045	4.00	7.35	5.40
9	O	H		Me	Et	0.040	4.00	7.40	5.40
10	O	H		Me	Et	0.042	16.00	7.38	4.80
11*	O	H		Me	Et	0.009	2.00	8.05	5.70
12	O	H		Me	Et	0.035	3.00	7.46	5.52

*test set

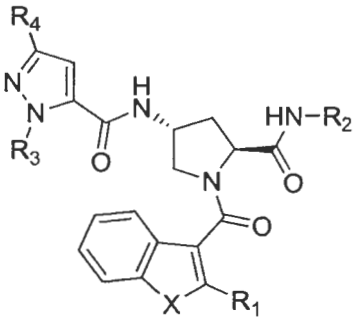
Table 2.7 Chemical structures and activities against InhA and *M. tuberculosis* of thirty four benzofuran pyrrolidine pyrazole derivatives (continued)

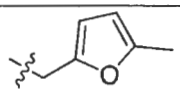
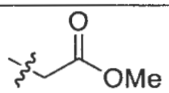
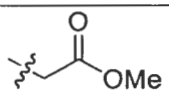
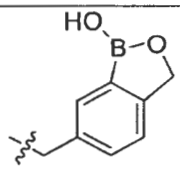
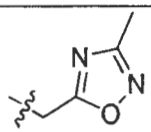
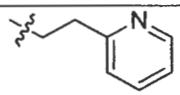
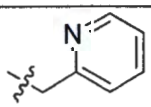
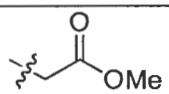


Cpd.	X	R_1	R_2	R_3	R_4	IC_{50} (μM)	MIC_{90} (μM)	$log(1/IC_{50})$	$log(1/MIC_{90})$
13	O	H		Me	Et	0.112	1.00	6.95	6.00
14	O	H		Me	Et	0.025	1.00	7.60	6.00
15	O	H		Me	Et	0.018	16.00	7.74	4.80
16	O	H		Me	Et	0.009	8.00	8.05	5.10
17	O	H		Me	Et	0.003	4.00	8.52	5.40
18	O	H		Me	Et	0.032	4.00	7.49	5.40
19	O	H		Me	Et	0.005	1.00	8.30	6.00

*tes: set:

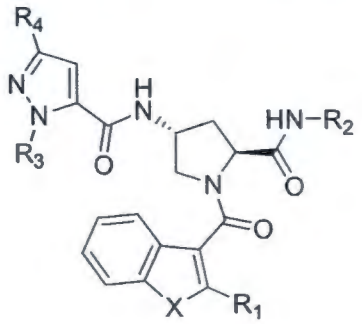
Table 2.7 Chemical structures and activities against InhA and *M. tuberculosis* of thirty four benzofuran pyrrolidine pyrazole derivatives (continued)

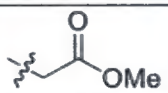
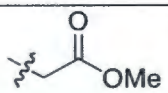
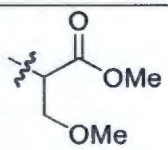
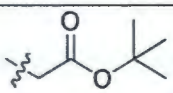
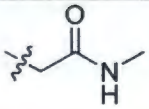


Cpd.	X	R_1	R_2	R_3	R_4	IC_{50} (μM)	MIC_{90} (μM)	$log(1/IC_{50})$	$log(1/MIC_{90})$
20	O	H		Me	Et	0.021	1.50	7.68	5.82
21	O	H		Me	Cyclopropyl	0.015	1.00	7.82	6.00
22	O	H		Et	Et	0.003	0.05	8.52	7.30
23	O	H	H	Et	Et	0.004	0.50	8.40	6.30
24	O	H		Et	Et	0.002	0.20	8.70	6.70
25	O	H		Et	Et	0.004	0.50	8.40	6.30
26	O	H		Et	Et	0.004	0.50	8.40	6.30
27	O	H	CH_2CH_2OH	Et	Et	0.003	1.00	8.52	6.00
28	O	H		Et	Et	0.002	0.70	8.70	6.15
29	O	Et		Me	Et	0.005	0.70	8.30	6.15

*test set

Table 2.7 Chemical structures and activities against InhA and *M. tuberculosis* of thirty four benzofuran pyrrolidine pyrazole derivatives (continued)



Cpd.	X	R_1	R_2	R_3	R_4	IC_{50} (μM)	MIC_{90} (μM)	$log(1/IC_{50})$	$log(1/MIC_{90})$
30	S	H		Me	Et	0.029	1.00	7.54	6.00
31*	O	Ph		Me	Et	0.003	1.50	8.52	5.82
32	O	H		Me	Et	0.018	2.00	7.74	5.70
33	O	H		Me	Et	0.007	2.00	8.15	5.70
34	O	H		Me	Et	0.008	2.00	8.10	5.70

*test set

2.2) 3D-QSAR studies based on CoMFA and CoMSIA methods

Two biological activities against InhA enzyme and *M. tuberculosis* whole cell, IC_{50} and MIC_{90} values of compounds were used to set up CoMFA (Crame et al., 1988) and CoMSIA (Klebe, Abraham, and Mietzner, 1994) models in order to evaluate the key structural features relating to the activity against both InhA and *M. tuberculosis*. The predicted binding modes of training set compounds obtained from molecular docking calculations were used for molecular alignment to set up CoMFA and CoMSIA models. SYBYL 8.0 molecular modelling

software was used to run CoMFA and CoMSIA models. Partial least square (PLS) analysis was employed to derive a linear relationship between CoMFA and CoMSIA descriptor fields and activities. The PLS analysis, using the leave-one-out (LOO) cross-validation method, was performed to determine the optimal number of components. Sequentially, a final analysis with the optimal number of components was performed to construct CoMFA and CoMSIA models that were not cross validated. The non-cross-validated correlation coefficient (r^2) and the leave-one-out cross-validated correlation coefficient (q^2) were used to evaluate the predictive ability of CoMFA and CoMSIA models. Selected CoMFA and CoMSIA models were employed to predict IC_{50} and MIC_{90} values of test set compounds that were not used to construct models. This was done to evaluate the external predictive ability of these models.

2.3) Molecular docking calculations

In this study, molecular docking calculations using the GOLD Program (Jones et al., 1995; 1997; Nissink et al., 2002; Verdonk et al., 2003(a)) were employed with the aims of generating the initial structure for MD simulations and performing molecular alignment to set up CoMFA and CoMSIA models. The available X-ray structure of InhA in a complex with compound **1** (PDB code 4COD) was used as an initial structure for molecular docking calculations. All atoms of the protein were kept rigid, whereas ligand was flexible during the molecular docking calculations. The number of Genetic Algorithm (GA) runs was set to 15 runs with the default search algorithm parameters. The docking calculations were validated using the root-mean-square deviation (RMSD) value between the docked and observed X-ray conformations of compound **1** in its pocket. A RMSD value lower than 1 Å was acceptable. Then, molecular docking calculations with validated parameters were used to dock all remaining compounds into the InhA binding pocket. The binding mode that showed the lowest binding energy was selected for each compound and was used to set up CoMFA and CoMSIA models. It was then used as the initial structure for MD simulations of compounds **2**, **22**, **23** and **28**.

2.4) Molecular dynamics simulations

Compound **28**, with the best IC_{50} value, was selected to investigate its binding mode in InhA. Moreover, the binding modes of compounds **2**,

22 and **23** were modelled by MD simulations in order to investigate the effect of R_2 and R_3 substituents on the IC_{50} value. The AMBER12 program was employed to perform molecular dynamics simulations. The complex structures of compounds **2**, **22**, **23** and **28** in InhA obtained from molecular docking calculations were used as the initial structure in MD simulations. The Amber ff03 force field was used for the physical description of InhA (Duan et al., 2003). The general Amber force field (GAFF) (Wang et al., 2004(a); 2006(b)) and restrained electrostatic potential (RESP) partial charges (Cornell et al., 1993; Bayly et al., 1993, Wang et al., 2000) of ligands and NAD^+ were generated by the antechamber module implemented in the AMBER12 package. To generate the system for MD simulations, the initial complex structure was solvated by TIP3P water (Mahone et al., 2000) in a truncated octahedral box extending up to 10 Å from the solute species. Five Na^+ ions were added to neutralize the system charge. Initially, the energy of system was minimized using a steepest decent method followed by the conjugate gradient method. Then, the system was gradually warmed from 0 K to 300 K in 30 ps by restraining all atoms of the complex with a restraint weight of 2 kcal/mol $Å^2$. This was followed by 70 ps of the position-restrained dynamics simulations with a restraining weight of 2 kcal/mol $Å^2$ at 300 K under an isobaric condition. Finally, 10 ns MD simulations without any restraints were performed using the same conditions. Long-range electrostatic interactions were applied using the Particle Mesh Ewald method (PME) (Darden et al., 1993) during the simulations. The cut-off distance for the long-range van der Waals interaction was set to 8 Å. The SHAKE method (Ryckaert, Ciccotti, and Berendsen, 1977) was applied to constrain the bond lengths of hydrogen atoms attached to heteroatoms. Coordinates and energy outputs during MD simulations were recorded at 2 ps intervals.

2.5) Binding free energy calculations

The Molecular Mechanics/Poisson–Boltzmann Surface Area (MM-PBSA) method (Homeyer and Gohlke, 2012; Wang et al., 2001(a); 2006(c) Hou et al., 2011) was employed for calculating the binding free energy of compounds **2**, **22** and **23** in InhA. In this calculation, 250 snapshots of the complex, receptor and ligand were extracted every 8 ps from the last nanosecond of the MD trajectory, which represents the equilibrium state. The entropy contribution was estimated using normal mode analysis with the Nmode module (Kaledin et al., 2004). The entropy

contribution was estimated using 250 snapshots for the binding free energy calculation.

2.2.1.2 Structure based virtual screening of InhA inhibitors

Initially, with a goal to identify new chemical scaffolds with InhA inhibitory activity, structure based virtual screening (SBVS) using docking calculations was performed. The details of virtual screening to identify new analogues of InhA inhibitors in this study are described below.

1) Ligand preparation

207,369 small compounds obtained from Spces database were selected as compound database for virtual screening in this study. LigPrep module of the Maestro molecular modeling package was used to produce the lowest energy conformations. Epik was used to generate the ionization/tautomeric states of compounds. The pH was set to 7.0. All the structure conformations of compounds were minimized using OPLS-2005 force field and at the most 32 conformations per ligands were generated.

2) Receptor preparation

InhA contained 269 amino acid residues. 60 % of amino acid residues in substrate binding pocket (the attractive target of InhA inhibitor) are hydrophobic residues. The binding site of InhA was reported as flexible binding pocket. There are "Y158-in" and "Y158-out" conformations of Tyr158 in catalytic site of InhA (Shirude et al., 2013). InhA complexed with diphenyl ether compound (PDB code 2X23) was selected as Y158-in receptor conformation (Luckner et al., 2010). For Y158-out receptor conformation, InhA complexed with benzofuran pyrrolidine pyrazole (PDB code 4COD) was selected (Encinas et al., 2014). The selected InhA enzymes were prepared using protein preparation module of Schrodinger's Maestro Molecular modeling suit. InhA enzyme with no missing residue and side chain were selected as receptor for structure based virtual screening. Selected proteins were prepared by Protein Preparation Wizard. Hydrogen atoms were added. The ionization/tautomeric states of amino acid residues were generated by Epik and were then optimized by OPLS-2005 force field.

3) Validations of docking parameters

Glide docking program (Friesner et al., 2004(a); Halgren et al., 2004) was used to prepare the docking grid for docking calculations. The binding site of InhA was set as the center of ligand. Two scoring functions, Glide SP (standard precision mode) and Glide XP (extra precision mode) (Friesner et al., 2006(b)) scoring functions were used to calculate the binding affinity of small compounds in InhA binding pocket for first screening and second screening. To confirm the accuracy of Glide docking program with different docking scoring, RMSD of ligand from X-ray conformations and docking conformations was calculated. The docking parameters produced RMSD value lower than 1 Å are acceptable.

4) Docking based virtual screening

In this study, we used SBVS using Glide docking with different docking scoring, to identify the suitable binding mode and affinity of ligands within the binding pockets of InhA. All compounds were docked into InhA binding pocket using Glide SP scoring function in first round screening. Top 2,000 ranked compounds with the highest Glide SP docking score were selected to re-dock into InhA binding site again using Glide XP scoring function. Glide XP docking scores of selected X-ray structures were used as cut off to select compounds from second screening. Then, compounds found in Y158-in and Y158-out conformations were selected as hit compounds.

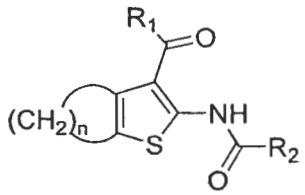
2.2.1.3 Molecular modeling and CAMD of benzothiophene derivatives as PknG inhibitors

1) Structure and biological activity

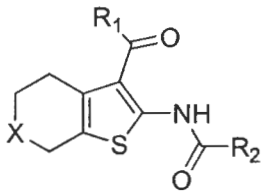
39 PknG inhibitors were selected to study in this work. There were classified as three main derivatives, thiophene(a), tetrahydrobenzothiophene(b) and benzothiophene(c) derivatives, respectively. Chemical structures and biological activities of these compounds against PknG enzyme were shown in Table 3.3 (Sipos et al., 2015; Bánhegyi, 2008). For 3D-QSAR study, all compounds were divided into a training set of 27 compounds and a test set of 4 compounds for model development and validation, respectively. The test set was randomly selected based on a structural diversity and wide range of activity in the data sets. The experimental IC_{50} values of each compound for PknG inhibition were converted to the corresponding log ($1/IC_{50}$)

values and used as dependent variables for the QSAR model. The three-dimensional structures were constructed using standard tools available in the GaussView 3.07 program and were then fully optimized using the ab initio quantum chemical method (HF/6-31G) implemented in the Gaussian 09 program.

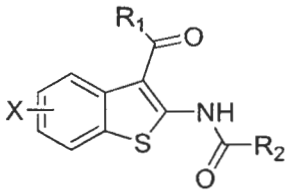
Table 2.8 Structure and biological activity of benzothiophene derivatives as PknG inhibitors



1-2a



3-20b

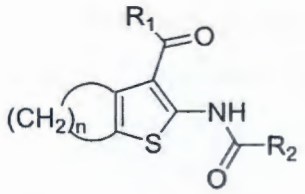


21-31c

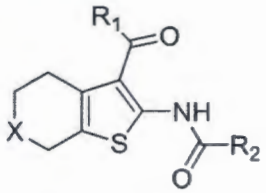
Cpd.	R_1	R_2	n	IC_{50} (μM)	$\log(1/IC_{50})$
1a	NH_2	Cyclopropyl	3	0.35	6.46
2a	NH_2	Cyclopropyl	5	68	4.17
Cpd.	**R_1**	**R_2**	**X**	**IC_{50} (μM)**	**$\log(1/IC_{50})$**
3b	NH_2	Cyclopropyl	CH_2	0.500	6.30
4b	NH_2	CH_3	CH_2	31	4.51
5b	NH_2	CH_2CH_3	CH_2	23	4.64
6b*	NH_2	$CH(CH_3)CH_2CH_3$	CH_2	17	4.77
7b	NH_2	$CH(CH_3)_2$	CH_2	4.79	5.32
8b	NH_2	Cyclobutyl	CH_2	1.66	5.78
9b*	NH_2	$C{=}CH_2CH_3$	CH_2	56	4.25
10b	NH_2	$CHCHCH_3$	CH_2	62	4.21
11b	NH_2	$CHC(CH_3)_2$	CH_2	0.41	6.39
12b	NH_2	2-thiophenyl	CH_2	26	4.59
13b	$NHCH_3$	Cyclopropyl	CH_2	22	4.66
14b*	NH_2	Cyclopropyl	O	0.63	6.20
15b	NH_2	Cyclopropyl	NCH_3	68	4.17
16b	NH_2	Cyclopropyl	S	0.74	6.13
17b	NH_2	Cyclopropyl	S=O	28	4.55
18b	NH_2	Cyclopropyl	$COCH_2CH_2O$	3.16	5.50
19b	NH_2	Cyclopropyl	C=O	0.49	6.31
20b	NH_2	NH-cyclohexyl	O	0.68	6.17
21c	NH_2	Cyclopropyl	H	0.085	7.07
22c	OCH_2CH_3	Cyclopropyl	H	58	4.24

* test set compounds

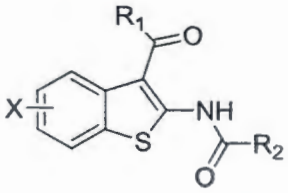
Table 2.8 Structure and biological activity of benzothiophene derivatives as PknG inhibitors (continued)



1-2a



3-20b



21-31c

Cpd.	R_1	R_2	X	IC_{50} (μM)	$\log(1/IC_{50})$
23c	NH_2	Cyclopropyl	5-Br	0.095	7.02
24c	NH_2	Cyclopropyl	5-NO_2	0.093	7.03
25c	NH_2	Cyclopropyl	5-NH_2	0.17	6.77
26c	NH_2	Cyclopropyl	6-OH	0.047	7.33
27c	NH_2	NH-6-benzo[d][1,3]dioxole	6-OH, 7-Cl	0.01	8.00
28c*	NH_2	$NHCH_2CH_2OH$	6-OH	0.05	7.30
29c	NH_2	Cyclopropyl	6-OH, 7-Br	0.02	7.70
30c	NH_2	Cyclopropyl	5-Cl, 6-OCH_3, 7-Cl	0.04	7.40
31c	NH_2	$NHCH(CH_3)_2$	6-OH	0.03	7.52
In1(b)	OCH_2CH_3	Cyclopropyl	CH_2	>100	<5.49
In2(b)	OH	Cyclopropyl	CH_2	>100	<5.49
In3(b)	NH_2	$C(CH_3)_3$	CH_2	>100	<5.49
In4(b)	NH_2	2-pyrroly	CH_2	>100	<5.49
In5(b)	NH_2	3-pyridinyl	CH_2	>100	<5.49
In6(b)	$NHNH_2$	Cyclopropyl	CH_2	>100	<5.49
In7(b)	$NHCH_2CH_2CH_3$	Cyclopropyl	CH_2	>100	<5.49
In8(b)	CN	Cyclopropyl	CH_2	>100	<5.49

* test set compounds

2) CoMSIA model

Molecular modeling software of SYBYL 8.0 with CoMSIA approach (Klebe, Abraham and Mietzner, 1994) was used to determine the relationship between the structures and the activities of tetrahydrobenzothiophene derivatives. The predicted binding modes of tetrahydrobenzothiophene derivatives derived from molecular docking calculations were used for molecular alignment to set up CoMSIA models. Five different physico-chemical properties of steric (S), electrostatic (E), hydrophobic (H), hydrogen bond donor (D) and hydrogen bond acceptor (A) fields are considered to develop a CoMSIA models. The attenuation factor was set as default

value of 0.3. The distance between inhibitor atoms and probe atoms of descriptors were calculated by gaussian–type functions at all grid points of 3D cubic lattice with grid spacing of 2 Å. Partial atomic charges of all compounds were calculated by the Gasteiger-Hückel charge calculation. To set up CoMSIA models, CoMSIA descriptors were used as independent variables and $\log(1/IC_{50})$ values were used as dependent variables. To derive a linear relationship between molecular descriptors and activities, the partial least squares (PLS) methodology analysis with the leave-one-out (LOO) cross-validation procedure was carried out to determine the optimal number of components. The cross-validation was performed using the leave-one-out method with a 2.0 kcal/mol column filter to minimize the influence of noisy columns. A final non cross-validated analysis with the optimal number of components was sequentially performed and was then employed to analyze the results. Non cross-validated correlation coefficient (r^2) and the leave-one-out cross-validated correlation coefficient (q^2) were used to evaluate the predictive ability of CoMSIA models.

3) Molecular docking calculations

Molecular docking calculations using Glide program (Friesner et al., 2004(a); Halgren et al., 2004) were used to predict the binding modes of PknG inhibitors in the PknG pocket. The X-ray structure of PknG enzyme complexed with tetrabenzophiophene, AX20017 compound taken from Protein Data Bank (PDB code: 2PZI) (Scherr et al., 2007) was employed for molecular docking calculations. All compounds were docked into the PknG pocket using Glide XP (Friesner et al., 2006(b)) docking scoring functions without applying any constraints. Molecular docking program and scoring functions were validated by root mean square deviations (RMSD) lower than 1 Å. All structures of inhibitors were optimized using the OPLS 2005 force field using the LigPrep module of Schrödinger. Tautomers and ionization states expected to occur in the pH range of 7.0 ± 2.0 were generated using the ionize module. Best docking pose for each compound was selected based on the lowest-energy docked.

4) MD simulations

MD simulations were performed in order to investigate the PknG-tetrahydrobenzothiophene and benzothiophene derivatives complexes. The initial coordinates for MD simulations were obtained from molecular docking

calculations using the Glide XP docking. AMBER12 using the Amber03 (Duan et al., 2003) force field and the general AMBER force field (GAFF) (Wang et al., 2004(b)) were employed for PknG and inhibitors, respectively. To obtain the restrained electrostatic potential (RESP) partial atomic charges of inhibitors, the geometry optimization and electrostatic potential calculation of each compound was first calculated at the HF/6-31G* level using the Gaussian09 program. Then, RESP partial charges were assigned using the antechamber module implemented in AMBER12. PknG- tetrahydrobenzothiophene and benzothiophene derivatives complexes were solvated by TIP3P waters in a truncated octahedral box extending up to 10 Å from the solute species. Eight Na^+ ions were added to neutralize the charge of each system. The added water molecules and ions in the solvated systems were minimized to relieve bad steric interactions using a steepest decent method. Non-bonded cutoff was set at 10 Å. The force of 500.0 kcal/mol was used to restrain the atom positions of all solute species. Thereafter, the whole system was minimized without restraint condition using conjugate gradient method. The systems were then gradually warmed up from 0 K to 300 K in the first 30 ps followed by maintaining the temperature at 300 K in the last 10 ps with 2 fs time steps in a constant volume boundary by restraining all atoms of the complex with a restraint weight of 2kcal/mol $Å^2$. This was followed by 70 ps of the position restrained dynamics simulations with a restraining weight of 2 kcal/mol $Å^2$ at 300 K under an isobaric condition. Finally, 15 ns MD simulations without any restraints were performed using the same conditions. Long-range electrostatic interactions were applied using the Particle Mesh Ewald method (PME) during the simulations. The cut-off distance for the long-range van der Waals interaction was set to 8 Å. The SHAKE method was applied to constrain the bond lengths of hydrogen atoms attached to heteroatoms. Coordinates and energy outputs during MD simulations were recorded at 2 ps intervals. Finally, the root-meansquare deviations (RMSDs) of the PknG enzyme and tetrahydrobenzothiophene and benzothiophene derivatives, respectively, were analyzed along the MD trajectory relative to the initial structures to determine the stability of the system. Energetics and structures of each system were analyzed based on the last 2 ns of the simulations, which represent the equilibrium state.

5) Binding free energy calculations

The binding free energy calculations of tetrahydrobenzothiophene and benzothiophene derivatives in the PknG pocket were calculated using the Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA) and Generalised Born Surface Area (MM-GBSA) models (Wang et al., 2001(a); 2006(c), Hou et al., 2011; Genheden and Ryde, 2015; Hou et al., 2011). MM-GBSA calculations were performed with the solation models. Amber03 force field (ff03) was used as protein parameters, whereas GAFF was used as parameter for inhibitors. AM1-BCC partial atomic charges of molecules were calculated with the Antechamber module of Amber12. The system was solvated using a truncated octahedral box of the water model TIP3P extending up to 10 Å from solute species. Eight Na^+ ions were used to neutralize. The refinement protocol consisted of two steps were carried out. In the first step 3000 iterations was performed to relieve the bad van der Waals (vdW) interactions. Atom coordinates for the PknG and ligand atoms water molecules were restrained to their initial coordinates with a force constant of 500 kcal mol^{-1} $Å^{-2}$. Non-bonded cutoff was set at 10 Å. In the second step 4000 iterations were applied to the system with the same first step conditions. The whole system was minimized freely to relieve bad contacts in the entire system of PknG-inhibitors. The MMPBSA.py script implemented in AMBER 12 was used for the calculation of energy components (Miller BR 3rd et al., 2012). In this calculation, single snapshot of the complex, receptor, and ligand were extracted from the energy minimizations step. The estimated entropy changes based on the number of rotatable bonds (NRot) of ligands have been reported. The estimated entropy change was approximated (Raha and Merz, 2004; Hayik, Dunbrack, and Merz, 2010).

2.2.1.4 Structure based virtual screening of PknG inhibitors

To identify novel specific kinase inhibitors for PknG, specific kinase database was selected as chemical database. GlaxoSmithKline (GSK) database contained 789 compounds were selected as chemical database for identification novel PknG inhibitors. The structure based virtual screening procedure in the present study is shown as below.

1) Ligand preparation

789 small compounds obtained from GSK database were prepared LigPrep method to produce the lowest energy conformations. Epik was used to generate the ionization/tautomeric states of compounds. pH was set at 7.0. All the structure conformations of compounds were minimized using OPLS-2005 force field and at the most 32 conformations per ligands were generated. Optimization was used to dock into PknG binding site.

2) Receptor preparation

PknG enzyme (751 amino acids) consisted three parts, *N*-terminal rubredoxin domain, a central kinase domain, and a C-terminal tetratricopeptide repeat (TPR) domain, respectively. The X-ray crystal structure of PknG complexed with trtrahydrobenzothiophene compound, AX20017 (PDB code 2PZI) was selected as receptor in this work. Selected proteins were prepared by Protein Preparation Wizard from Schrodinger's Maestro Molecular modeling suit. Hydrogen atoms were added. The ionization/tautomeric states of amino acid residues were generated by Epik and were then optimized by OPLS-2005 force field.

3) Validations of docking parameters

The docking grid for docking calculations was set as the center of ligand. This work, Glide XP (extra precision mode) scoring functions was used to calculate the binding affinity of small compounds in PknG binding. To confirm the accuracy of Glide docking program, RMSD of AX20017 from X-ray conformations and Glide XP docking conformations was calculated. The docking parameters produced RMSD value lower than 1 Å are acceptable. All compounds from GSK database were docked into PknG binding site using Glide XP with the same protocol with AX20017.

4) Docking and binding free energy based virtual screening

To identify ATP competitive inhibitors of PknG, we used SBVS using Glide XP docking with binding free energy calculation with MM-GBSA to identify high ligand binding affinity of ligands within the PknG binding pockets. All compounds were docked into ATP catalytic site in kinase domain of PknG using Glide XP scoring function in docking calculation step. 100 compounds with high binding affinity based on Glide XP scoring were selected as hit compounds. MM-GBSA with

IGB8 was used to estimate the binding free energy of 100 small compounds in PknG binding pocket.

2.2.2 Rational design of anti-cancer agents

2.2.2.1 Data sets and biological activities

Twenty-eight azanaphthoquinone annelated pyrrole derivatives taken from one laboratory were used in this work to ensure that all experimental data were determined under consistent assay conditions (Shanab et al.; 2007(a); 2010(b); 2011(c); Pongprom et al.; 2009(a); 2010(b); Pongprom and Pungpo, 2012). The general frame of these compounds is shown in Figure 3.1. Chemical structures and antiproliferative activities on the human cancer cell line of cervical carcinoma (KB/HeLa) of azanaphthoquinone annelated pyrrole derivatives are listed in Table 3.4. All chemical structures of these compounds were constructed using the standard tools available in the program GaussView 3.07 and were then fully optimized using the HF/3-21G method implemented in the Gaussian 03 program. For QSAR study, 28 azanaphthoquinone annelated pyrrole derivatives were divided into a training set of 24 compounds and a test set of 4 compounds for final model development and model validation, respectively. The representatives of the test set were manually selected and are covering the utmost range of activity and structural diversity of compounds in the data set.

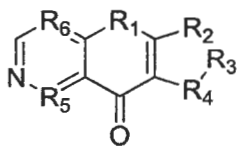
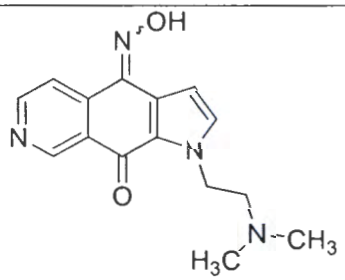
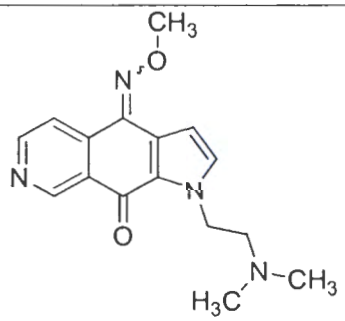
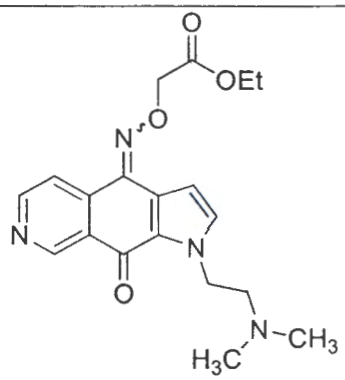
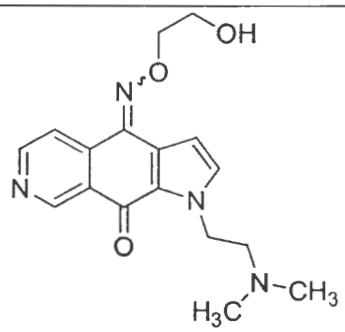


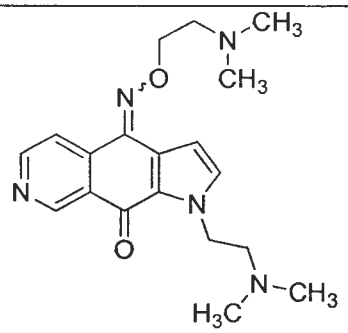
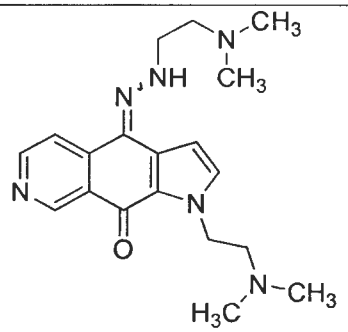
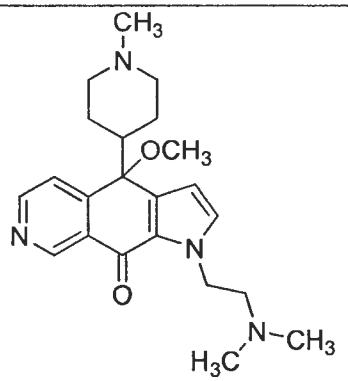
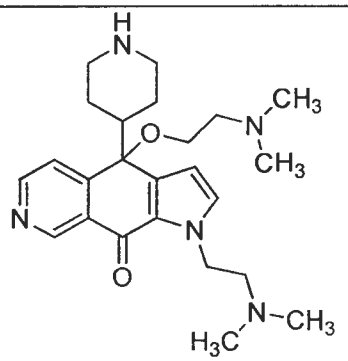
Figure 2.11 General frame of azanaphthoquinone annelated pyrrole derivatives

Table 2.9 Chemical structures and antiproliferative activities on cervical carcinoma of azanaphthoquinone annelated pyrrole derivatives

Cpd.	Structure	log(1/IC$_{50}$)
1		4.92
2		5.92
3		4.92
4		6.18

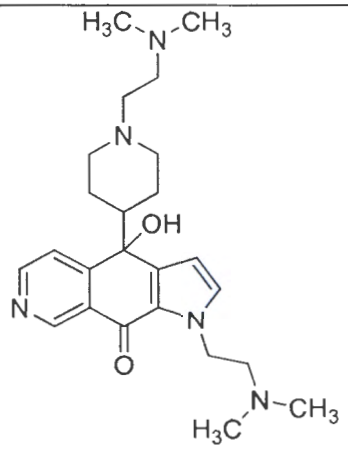
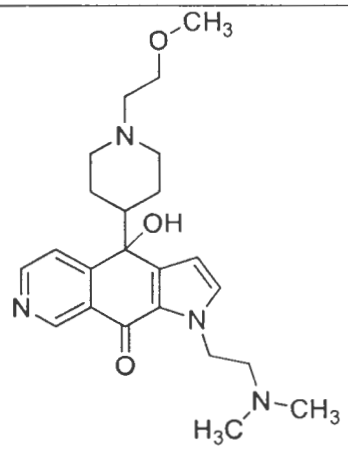
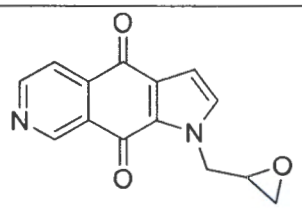
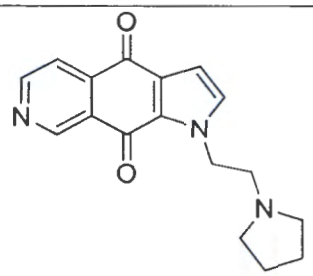
* The test set compound

Table 2.9 Chemical structures and antiproliferative activities on cervical carcinoma of azanaphthoquinone annelated pyrrole derivatives (continued)

Cpd.	Structure	log(1/IC$_{50}$)
5		6.17
6		6.55
7*		4.45
8		5.05

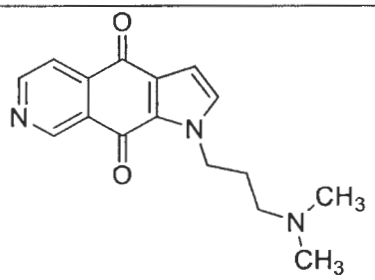
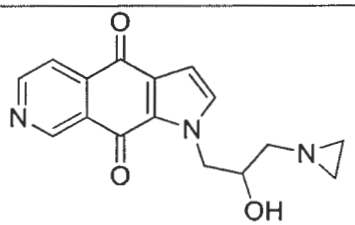
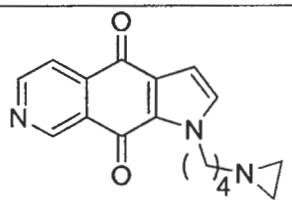
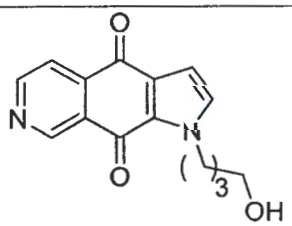
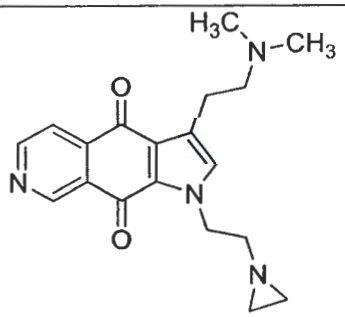
* The test set compound

Table 2.9 Chemical structures and antiproliferative activities on cervical carcinoma of azanaphthoquinone annelated pyrrole derivatives (continued)

Cpd.	Structure	log(1/IC_{50})
9		4.75
10		4.67
11		6.61
12		8.10

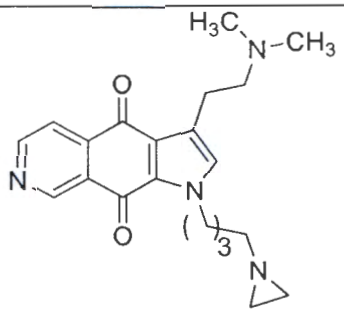
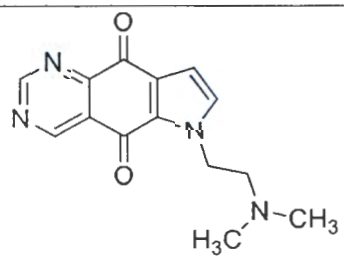
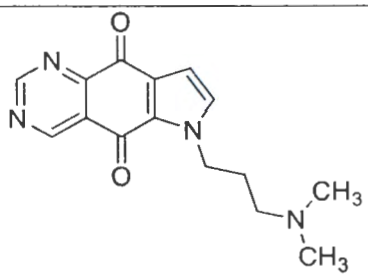
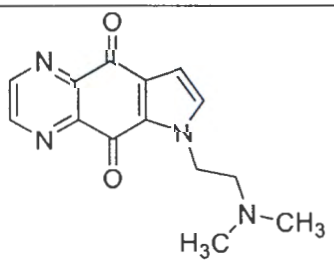
* The test set compound

Table 2.9 Chemical structures and antiproliferative activities on cervical carcinoma of azanaphthoquinone annelated pyrrole derivatives (continued)

Cpd.	Structure	log(1/IC$_{50}$)
13		6.85
14		5.60
15		5.73
16		5.13
17*		5.54

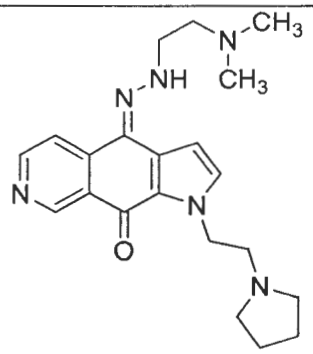
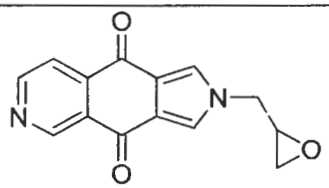
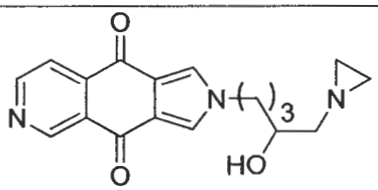
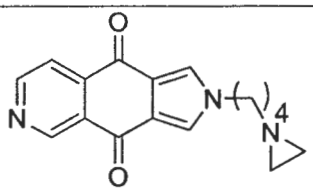
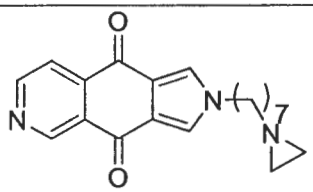
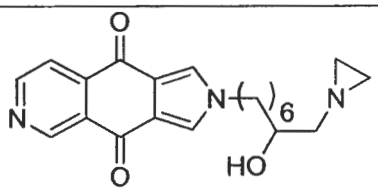
* The test set compound

Table 2.9 Chemical structures and antiproliferative activities on cervical carcinoma of azanaphthoquinone annelated pyrrole derivatives (continued)

Cpd.	Structure	log(1/IC$_{50}$)
18*		5.85
19		6.19
20*		6.01
21		5.30

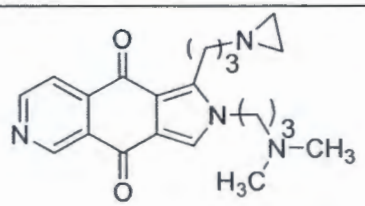
* The test set compound

Table 2.9 Chemical structures and antiproliferative activities on cervical carcinoma of azanaphthoquinone annelated pyrrole derivatives (continued)

Cpd.	Structure	log(1/IC$_{50}$)
22		4.58
23		5.49
24		5.66
25		5.55
26		5.68
27		5.55

* The test set compound

Table 2.9 Chemical structures and antiproliferative activities on cervical carcinoma of azanaphthoquinone annelated pyrrole derivatives (continued)

Cpd.	Structure	log(1/IC$_{50}$)
28		5.56

* The test set compound

2.2.2.2 3D QSAR technique

CoMSIA (Klebe, Abraham and Mietzner, 1994) was used to elucidate the relationship between the structures and the activities of azanaphthoquinone annelated pyrrole derivatives. Molecular modeling software of SYBYL 8.0 was used to calculate CoMSIA models. The molecular alignment for the setup of appropriate CoMSIA models was carried out by the Sybyl pharmacophore alignment module GALAHAD (Genetic Algorithm with Linear Assignment for Hypermolecular Alignment of Datasets). Five CoMSIA similarity index descriptors including steric, electrostatic, hydrophobic, hydrogen donor and hydrogen acceptor fields were derived with the grid spacing of 2 Å. There are no energy cutoffs for CoMSIA calculations. To generate a contour map with prominent molecular features in the CoMSIA study, an attenuation factor of 0.3 was used. To derive a linear relationship between molecular descriptors and activities, the partial least square (PLS) approach was employed, in which CoMSIA descriptors were set as independent variables and log (1/IC$_{50}$) values were used as dependent variables. The cross-validation was performed using the leave-one-out method with a 2.0 kcal/mol column filter to minimize the influence of noisy columns. A final non cross-validated analysis with the optimal number of components was sequentially performed and was then employed to analyze the results. The r^2 and q^2 values were used to evaluate the predictive ability of CoMSIA models.

2.2.2.3 MD Simulations

To obtain accuracy and reliability of the binding mode information, five compounds covering the range of the most active to the least active compounds in the series studied were selected for MD simulations. Compound **12** is represented as the most active compound, whereas compounds **7** and **22** are representative compounds possessing weak inhibitory activities in the dataset. Moreover, compounds **11** and **15** possessing moderate activities were also selected. The X-ray crystal structure with the pdb code of 2GB9 was used as the initial coordinates of $d(CGTACG)_2$. The initial coordinates of the selected compounds complexed with $d(CGTACG)_2$ were taken from molecular docking calculations using GOLD Program (Verdonk et al., 2003(a); 2005(b); Cole et al., 2005). AMBER12 using the Amber99 force field (Duan et al., 2003) for DNA duplex and the general Amber force field (GAFF) parameters (Wang et al., 2004(b)) for the selected compounds was employed for MD simulations. Each complex structure was solvated by TIP3P waters (Jorgensen et al., 1983) in an octahedral box extending up to 10 Å from each solute species, $d(CGTACG)_2$ and the selected compounds. 12 Na^+ cations were added to neutralize the charge of each system. The added water molecules and ions in the solvated systems were relaxed using the Sander program to relieve bad steric interactions. Non-bonded cutoff was set at 10 Å. The force of 500.0 kcal/mol was used to restrain the atom positions of all solute species. Thereafter, the whole system was minimized without restraint condition. The systems were then gradually warmed up from 0 K to 300 K in the first 20 ps followed by maintaining the temperature at 300 K in the last 10 ps with 2 fs time steps in a constant volume boundary. The solute species in the solvated systems were restrained to their initial coordinates with a weak force constant of 10 kcal/mol$Å^2$ during the temperature warming. Afterward, the position-restrained dynamics simulation using 2 fs time steps through 70 ps at 300 K under the isobaric condition was performed for each system to relax the positions of the solvent molecules. In this dynamics run, the positions of solute species were restrained with a weak force constant of 10 kcal/mol $Å^2$ during the position-restrained dynamics simulations. Finally, 20 ns MD simulations without the position restraints were performed under the same conditions. During the dynamics simulations, a nonbonded cutoff distance of 8 Å was applied to handle electrostatic interactions in periodic boxes

by the Particle Mesh Ewald method (Darden, York and Pedersen, 1993). The SHAKE method (Ryckaert, Ciccotti and Berendsen, 1977) was applied to constrain the bond lengths of hydrogen atoms attached to heteroatoms. Coordinates and energy outputs during molecular dynamics simulation were printed every 2 ps. MD trajectories were evaluated in terms of the root mean square deviation (RMSD), complex structure and binding free energy.

2.2.2.4 Binding energy calculation

The binding free energies of all selected compounds bound to $d(CGTACG)_2$ were calculated using the Molecular Mechanics/Poisson-Boltzmann Surface Area method (MMPBSA method) (Srinivasan et al., 1998) implemented in AMBER 12 package. In this work, 500 snapshots evenly from the last 10 ns on the MD trajectory with an interval of 20 ps were used in the MM-PBSA calculations. The entropy contribution ($T\Delta S$) to the binding free energy was estimated using normal-mode analysis with AMBER Nmode module. To save the computational time, 100 snapshots evenly from the last 10 ns on the MD trajectory with an interval of 100 ps were used to estimate the contribution of the entropy.

CHAPTER 3
RESULTS AND DISCUSSION

The results derived from molecular modeling and CAMD were divided into two sections. First part is rational design anti-tuberculosis agents. Second part is rational design anti-cancer agents. The details of results that observed in this study were discussed in this chapter.

3.1 Rational design novel anti-tuberculosis agents

3.1.1 Enoyl-ACP reductase inhibitors

3.1.1.1 Molecular modeling of diphenyl ether derivatives

1) 3D-QSAR study

1.1) CoMSIA model

The PLS results of CoMSIA models are summarized in Table 3.1. Ten CoMSIA models were constructed with various combinations of CoMSIA descriptors. Among all models, model 8 composing the steric, hydrogen bond donor, electrostatic and hydrophobic fields is the best CoMSIA model, giving the best statistical parameters with a q^2 value of 0.60 and a r^2 value of 0.95. There is a good correlation between actual and predicted activities of the training set based on the best CoMSIA model, as depicted in Figure 3.1. In order to assess the external predictive ability of this model, the InhA inhibitory activities of the test set were predicted. The predicted values of 9 test set compounds are within one logarithmic unit difference from the experimental values (Figure 3.1). Therefore, the best CoMSIA model is reliable with highly predictive ability and could be utilized to predict the InhA activities for newly designed diphenyl ether inhibitors.

Table 3.1 Summary of statistical results of CoMSIA models

Models	Statistical data						Fraction
	q^2	r^2	s	SSE	F	N	
1.S/E	0.29	0.93	0.70	0.21	85.48	6	38.6/61.4
2.S/H	0.08	0.69	0.75	0.44	43.53	2	38.8/61.2
3.S/D	0.54	0.89	0.56	0.27	50.30	6	53.7/46.3
4.S/A	0.13	0.88	0.76	0.28	54.86	5	53.5/46.5
5.S/D/E	0.58	0.93	0.54	0.22	77.34	6	27.9/41.7/30.5
6.S/D/H	0.56	0.93	0.55	0.21	85.19	6	29.9/37.6/32.5
7.S/D/A	0.51	0.93	0.58	0.22	76.55	6	39.9/33.1/27.0
8. S/D/E/H	**0.60**	**0.95**	**0.52**	**0.19**	**104.17**	**6**	**19.0/32.5/23.8/24.8**
9. S/D/E/A	0.50	0.93	0.58	0.22	78.66	6	23.9/34.3/25.9/15.9
10.S/D/E/H/A	0.55	0.95	0.55	0.19	103.76	6	17.2/26.6/21.4/22.1/12.8

N, optimum number of components; s, standard error of prediction; SEE, standard error of estimate; F, F-test value; S, steric field; E, electrostatic field; H, hydrophobic field; D, hydrogen donor field; A, hydrogen acceptor field

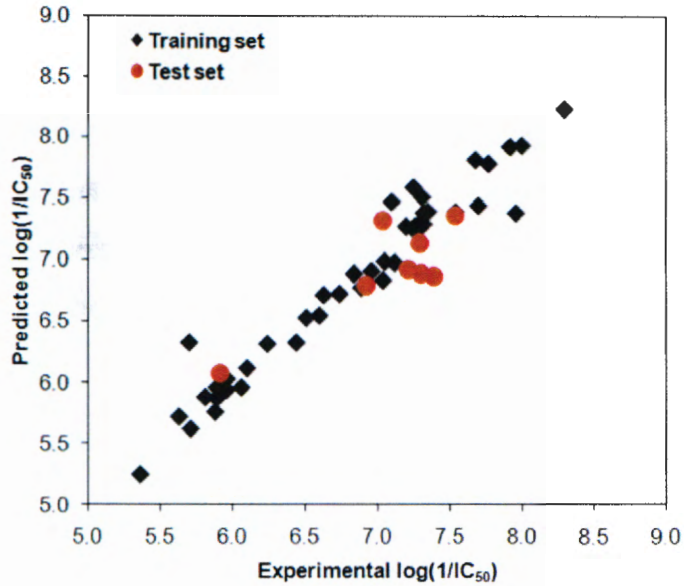


Figure 3.1 Plots between the experimental and predicted activities of the training and test sets derived from the CoMSIA model

The predictive abilities of the best CoMSIA model were determined from the test set including 9 compounds. For the best CoMSIA model, internal validation of leave-one-out cross validated q^2 and predicted r^2 ($r^2_{pred.}$ or r^2) were found to be 0.64 and 0.70, respectively. The calculated of square correlation coefficient values between the experimental and predicted values of the test set compounds with intercept set at zero (r^2_0) and without intercept (r^2) were 0.56 and 0.73, respectively. The slope of regression line through the origin (k) of the best CoMSIA model was 1.02, which is close to 1. Based on the statistical results, the best CoMSIA model could be considered reliable.

1.2) CoMSIA contour maps

To reveal the importance of molecular descriptor fields on InhA inhibitory activities of diphenyl ether derivatives, CoMSIA contour maps were established. Figure 3.2 and Figure 3.3 present the CoMSIA contour maps which reveal the influence of steric, electrostatic, hydrophobic and hydrogen donor fields to the activity of diphenyl ether derivatives. Green and yellow contours indicate areas where favorable and unfavorable steric bulks are predicted to enhance the activities of diphenyl ether derivatives. Blue and red contours indicate regions where electropositive and electronegative groups lead to increasing of the InhA inhibitory activity, respectively. Magenta and white contours represent areas, where the hydrophobic group and the hydrophilic group are predicted to favour the biological activities. The cyan and orange contours indicate regions that favor the hydrogen donor group and unfavor hydrogen donor group, respectively. The interpretation of CoMSIA contour maps reveals the structural requirement of each substituent position in the scaffold of diphenyl ether derivatives helpful for rational design of novel and potent InhA inhibitors.

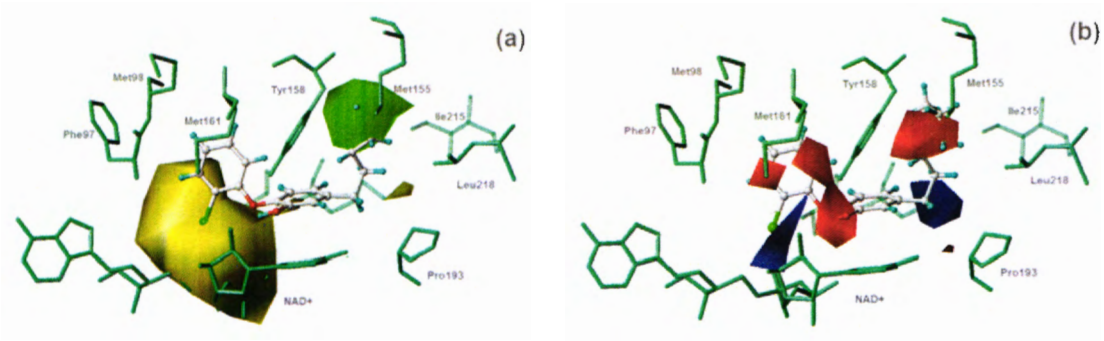


Figure 3.2 CoMSIA steric (a) and electrostatic (b) contours in combination with compound 29 (ball and stick in atom type colors) in InhA binding pocket (stick in greenblue)

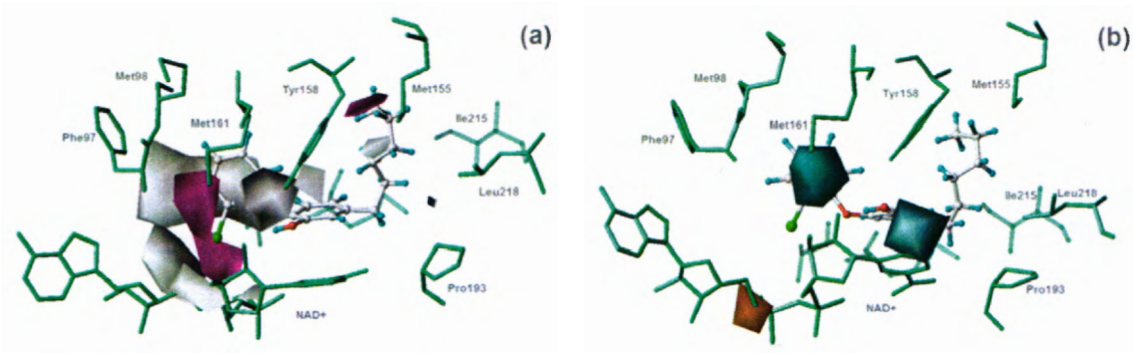


Figure 3.3 CoMSIA hydrophobic (a) and hydrogen bond donor (b) contours in combination with compound 29 (ball and stick in atom type colors) in InhA binding pocket (stick in greenblue)

1.3) Structural requirement for the R_1 positions on the phenyl A ring

The appearance of cyan contours near the OH group of the phenyl A ring emphasizes the important role of this moiety to the InhA inhibitory activity of diphenyl ether derivatives (Figure 3.3b). The C4 and C6 atoms of hexyl side chain of compound **29** are covered by green and red contours (Figure 3.2a and Figure 3.2b). Therefore, the R_1 substituent containing the bulky size and high electron density would be favorable for this region. In case of the R_1 substituent as the alkyl chain, the alkyl chain with the carbon atoms higher than two atoms should be preferable for the InhA inhibitory activity. As exemplified, compound **21** containing octyl group at the R_1 position possesses the most active compound in this series,

whereas compound **18** containing ethyl substituent exhibits much lower inhibitory activity than those of compounds **19, 20, 17** and **21** bearing butyl, pentyl, hexyl and octyl substituents at the R_1 position, respectively. Corresponding to the MD results, the longer alkyl chain at R_1 substituent could form hydrophobic interactions more than the shorter alkyl chain.

1.4) Structural requirement for the R_2, R_3, R_4 positions on the phenyl B ring

The unfavourable hydrophobic white contour and the unfavourable steric yellow contour present near the R_2, R_3 and R_4 substituents (Figure 3.2a and Figure 3.3a). These results indicate that the small hydrophilic substituents at the R_2, R_3 and R_4 positions are required for the InhA inhibitory activity of diphenyl ether derivatives. Therefore, compounds **37-44** containing the bulky hydrophilic substituents at the R_2, R_3 and R_4 positions show poor activities for InhA inhibition with IC_{50} more than 360 nM. These suggestions are in agreement with the binding modes of compounds **17, 18, 19** and **29** observed from the MD simulations that the R_2, R_3 and R_4 substituents are located near the pyrophosphate moiety of NAD^+, the hydrophilic backbones of Gly96 and Met98, respectively. Accordingly, the small substituent with hydrophilic property at the R_2, R_3 and R_4 substituents should be optimal for the InhA binding pocket. Moreover, the magenta and blue contours close to the R_2 substituent suggest additional structural requirement at this position, which should contain the hydrophobic property and less electron density. This suggestion is consistent with the MD results which indicate that the R_2 position can be substituted with hydrophobic or hydrophilic groups so that the phenyl B ring could be favorable in binding with the methyl side chain of Ala198, and the pyrophosphate moiety of NAD^+, respectively. Apart from hydrophobic property, the R_2 substituent with the less electron density should be optimal for the pyrophosphate moiety of NAD^+ presenting the negative charge.

2) MD simulations of diphenyl ether derivatives

To study the crucial information for binding of diphenyl ether derivatives in InhA binding pocket, MD simulations was performed. Selected diphenyl ether derivatives are classified into two groups. First group, the effect of R_1, R_2, R_3 and R_4 substituents were investigated. Based on experimental data, R_1 and R_2 substituent

diphenyl ether derivatives were found as potent InhA inhibitors. Therefore, the effect of R_1 and R_2 substituent diphenyl ether derivatives were studied using MD simulations. The obtained results derived from MD simulations of two different groups of selected diphenyl ether derivatives were integrated with the structural requirements derived from 3D-QSAR CoMSIA study. The results were discussed in this chapter.

2.1) MD simulations of diphenyl ether derivatives set 1

2.1.1) System equilibration of first group of diphenyl ether derivatives

To evaluate the reliable stability of the MD trajectories, the RMSDs for all atoms of the InhA, NAD^+ cofactor and diphenyl ether derivatives relative to the initial minimized structure over the 10 ns of simulation times were calculated and plotted in Figure 3.4. There are three solute species in each MD system including the InhA, NAD^+ and inhibitor. The plateau characteristic of the RMSD plot over the simulation time is the criteria to indicate the equilibrium state of each solute species. For the equilibrium state of each MD system, the RMSD plots of all solute species have to reach the plateau characteristic. The InhA, NAD^+ and inhibitor in each system reach the equilibrium state at the different time (Figure 3.4). For the system of compound **6**, NAD^+ and this compound reach the equilibrium state at the early time, whereas the InhA reaches the equilibrium state after 2 ns (Figure 3.1a). Therefore, after 2 ns the RMSD plots of all solute species reach the plateau characteristic indicating the equilibrium state of this MD system. In case of compound **53**, its MD system reaches the equilibrium after 1 ns (Figure 3.4b). For compounds **24** and **54**, the MD systems reach the equilibrium after 2 ns (Figures 3.4c and Figure 3.4d). The RMSD plots over 10 ns of these compound show large fluctuation in the range about 0.5-2.5 Å. The result could be accounted by the more flexibility of the long hexyl chain. Therefore, the information in terms of binding free energy, interaction energy and structure of each system after an equilibrium state were analyzed.

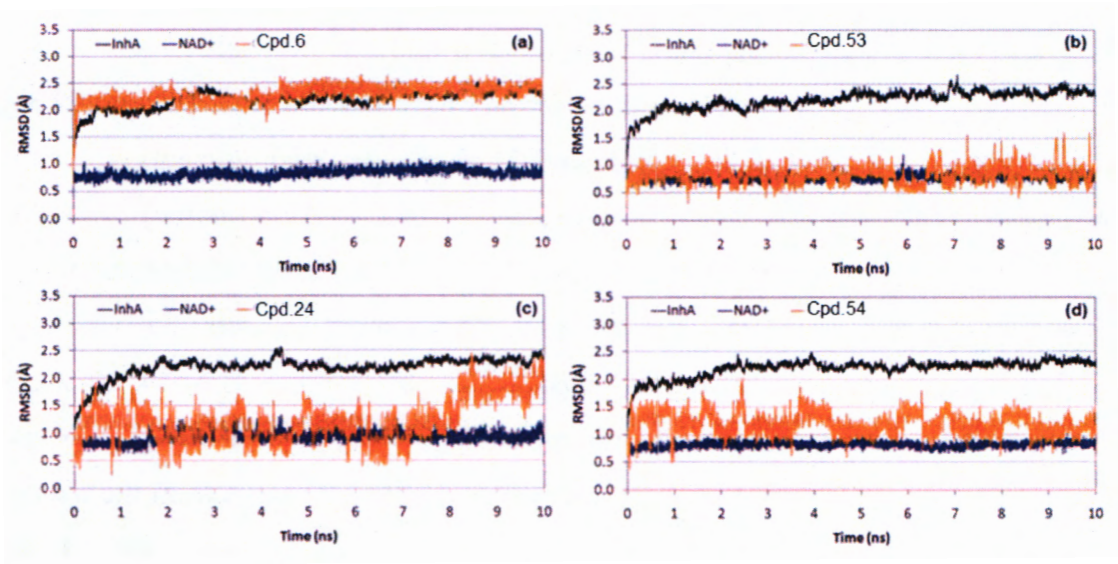


Figure 3.4 RMSDs of diphenyl ether derivatives, compounds 6 (a), 53 (b), 24 (c) and 54 (d) complexed with the InhA

2.1.2) Structural flexibility of InhA binding pocket of first group of diphenyl ether derivatives

The binding cavity volumes the InhA complexed with diphenyl ether derivatives were calculated to study the flexibility of the InhA binding pocket by computed atlas of surface topography of proteins (CASTp) method on the website (http://sts.bioengr.uic.edu/castp/index.php) (Liang, Edelsbrunner and Woodward, 1998). Amino acid residues within 10 Å from ligands-NAD^+ cofactor were used to calculate the binding cavity volumes of the InhA binding pockets. The binding cavity volumes of the InhA complexed with six diphenyl ether derivatives were summarized in Table 2. The binding cavity volumes ranging from 1,763 $Å^3$ to 2,465 $Å^3$ of six inhibitors were found. Moreover, the binding cavity volume of *trans*-2-hexadecenoyl-(n-acetyl-cysteamine)-thioester substrate (PDB code 1BVR) (Rozwarski et al., 1999) are 1,902 $Å^3$. In previous study (Pauli et al., 2013), the binding cavity volumes of direct InhA inhibitors obtained from the available X-ray structures of InhA inhibitors are ranging from 1,597 $Å^3$ to 3,047 $Å^3$. These results indicate that the binding pocket of InhA is flexible enough to bind with the substrate and diphenyl ether inhibitors. The binding cavity volume of compound **24** in the InhA binding pocket is higher than that of the highest compound **21** and *trans*-2-

hexadecenoyl-(*n*-acetyl-cysteamine)-thioester substrate. In addition, the binding cavity volumes of low potent compounds, compound **6** and **53** are higher than *trans*-2-hexadecenoyl-(n-acetyl-cysteamine)- thioester substrate and the most potent diphenyl ether inhibitor. The most potent compound **21** shows the binding cavity volume is comparable with that of *trans*-2-hexadecenoyl-(*n*-acetyl-cysteamine)- thioester substrate. Compounds **6**, **24**, **53** and **54** creating the InhA cavity volume larger than that of the substrate show poor activities to inhibit the InhA. To reveal the correlation between the InhA cavity volume, the molecular size of compounds **6**, **14**, **21**, **24**, **53** and **54** and their inhibitory activities, the molecular surface area of these compounds were calculated using Hyperchem 7.51, reported in Table 3.2. Apart from the *trans*-2-hexadecenoyl-(*n*-acetyl-cysteamine)-thioester substrate possessing the largest molecular surface area with 663 $Å^2$, the molecular surface area of compounds **21** with 612 $Å^2$ is the largest one. Compound **54** has the surface area of 611 $Å^2$ close to that compound **21** but its cavity volume is larger. In case of compounds **6**, **24** and **53**, these compounds show smaller surface areas than that of compound **21** but their pocket cavity volume are larger. These results imply that compound **21** having larger molecular size in its smaller cavity volume could induce the InhA pocket to well fit with its binding better than compounds **6**, **24**, **53** and **54**. Accordingly, compound **21** should produce more interactions with amino acid residues in the InhA pocket than compounds **6**, **24**, **53** and **54**, leading to show better inhibitory activity of this compound. In contrast, compound **14** has smaller surface area than that of compound **21** but its cavity volume is close to that of compound **21**. This result implies that the binding of compound **14** could not properly fit to the InhA pocket leading to loss its activity as compared with compound **21**.

Table 3.2 The binding cavity volume of diphenyl ether inhibitors

Compound	Ligand Surface Area (Å^2)	Binding cavity volume (Å^3)
6	546	2,318
53	530	2,042
24	574	2,465
54	611	2,039
14	575	1,763
21	612	1,876
Substrate [a]	663	1,902

[a] *trans*-2-hexadecenoyl-(*n*-acetyl-cysteamine)- thioester substrate

2.1.3) Binding free energy calculations of first group of diphenyl ether derivatives

To evaluate the binding affinity of diphenyl ether derivatives in the InhA binding pocket, the binding free energies (ΔG_{bind}) were calculated using the Molecular Mechanics/Poisson Boltzmann Surface Area (MM-PBSA) method. The binding free energies (ΔG_{bind}) were calculated by equation 34 (Chapter 2). The entropic (-TΔS), enthalpic (ΔH) energy and binding free energy of the diphenyl ether/InhA complexes are listed in Table 3.3. The binding free energies of compounds **6, 24, 53** and **54** bound to the InhA are −10.3, −7.9, −11.6 and −13.7 kcal/mol, respectively. Moreover, the binding free energy of the X-ray crystal structures of compound **14**/InhA and compound **21**/InhA complexes were calculated to compare with those of compounds **6, 24, 53** and **54**. The calculated and experimental binding free energies were compared as shown in Table 3.3. Moreover, a good linear correlation (r^2=0.78) between the experimental IC_{50} and calculated binding free energy is presented in Figure 3.5. It is notable that the calculated binding free energies of the selected compounds are in the correct order as compared with the IC_{50} values. The obtained results could be successfully used to validate the MD procedure in this study.

Table 3.3 Binding free energies in kcal/mol computed by the MM-PBSA method

Cpd.	Contribution							
	ΔG_{ele}	ΔG_{vdw}	ΔG_{MM}	ΔG_{sol}	ΔH	$-T\Delta S$	ΔG_{bind}	ΔG_{exp} [a]
6	-10.5 ± 2.5	-45.2 ± 2.3	-55.6 ± 5.6	27.2 ± 3.5	-28.4 ± 3.0	18.2 ± 0.8	-10.3 ± 1.4	-10.0
53	-12.4 ± 2.9	-47.1 ± 2.2	-59.4 ± 3.0	31.2 ± 2.5	-28.2 ± 2.8	20.3 ± 0.8	-8.0 ± 0.8	-9.6
24	-12.9 ± 3.2	-55.8 ± 2.7	-68.7 ± 3.5	31.2 ± 4.4	-37.5 ± 4.0	25.9 ± 1.9	-11.6 ± 1.5	-10.0
54	-9.0 ± 2.7	-50.3 ± 2.5	-59.3 ± 3.0	24.3 ± 2.2	-34.9 ± 2.8	21.3 ± 1.1	-13.7 ± 1.3	-10.1
14	-12.5 ± 2.0	-51.2 ± 2.2	-63.6 ± 3.2	32.8 ± 3.2	-30.8 ± 2.8	19.0 ± 1.0	-11.8 ± 1.3	-10.0
21	-16.6 ± 3.0	-55.3 ± 2.4	-71.9 ± 3.4	24.3 ± 2.5	-36.1 ± 5.5	22.1 ± 1.4	-13.9 ± 1.5	-11.4

[a] derived from $\Delta G = RT \ln[IC_{50}]$, R represents the gas constant (1.988 cal/mol K), T represents the temperature (300 K).

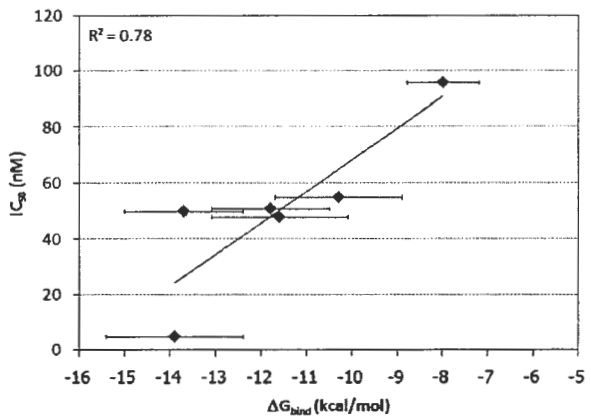


Figure 3.5 Correlation of experimental IC_{50} and calculated binding free energy using MM-PBSA method

According to the energy components of the binding free energies listed in Table 3.3, the van der Waals energy of diphenyl ether derivatives in the InhA binding pocket is the most contribution to the binding free energy due to the high hydrophobicity of the InhA binding pocket (Pauli et al., 2013). The van der Waals energy of diphenyl ether compounds **24, 54, 14** and **21** are lower than -50 kcal/mol. These results indicate that the increase of the hydrophobicity of inhibitors leads to increase the binding affinity of inhibitors in the InhA pocket. This result agrees with the experimental biological activity (Freundlich et al., 2009; Sullivan et al., 2006; am Ende et al., 2008; Luckner et al, 2010; Pan et al.,2014).

2.1.4) Per-residue binding energy decomposition of first group of diphenyl ether derivatives

The binding energies between diphenyl ether derivatives with each residue in InhA pocket were calculated by MM-GBSA method to understand the key interactions for binding of protein-ligand complexes. Figure 3.6 displays the binding energy decomposition of diphenyl ether derivatives. The obtained results indicate that nine residues including Phe97, Phe149, Tyr158, Met161, Lys165, Ala198, Met199, Val203 and NAD^+ cofactor show lower interaction energies with diphenyl ether inhibitors. Figure 3.7 shows the contribution of the van der Waals and electrostatic energies on the binding of diphenyl ether derivatives in the InhA binding pocket. Most of residues show the van der Waals energy lower than electrostatic energy. This result indicates that the van der Waals interactions display the important role on the binding of diphenyl ether derivatives in the InhA binding pocket. It is important to note that all repulsive energies observed for each compounds are generated from electrostatic energy. Particularly, the lower active compounds show more repulsive energies.

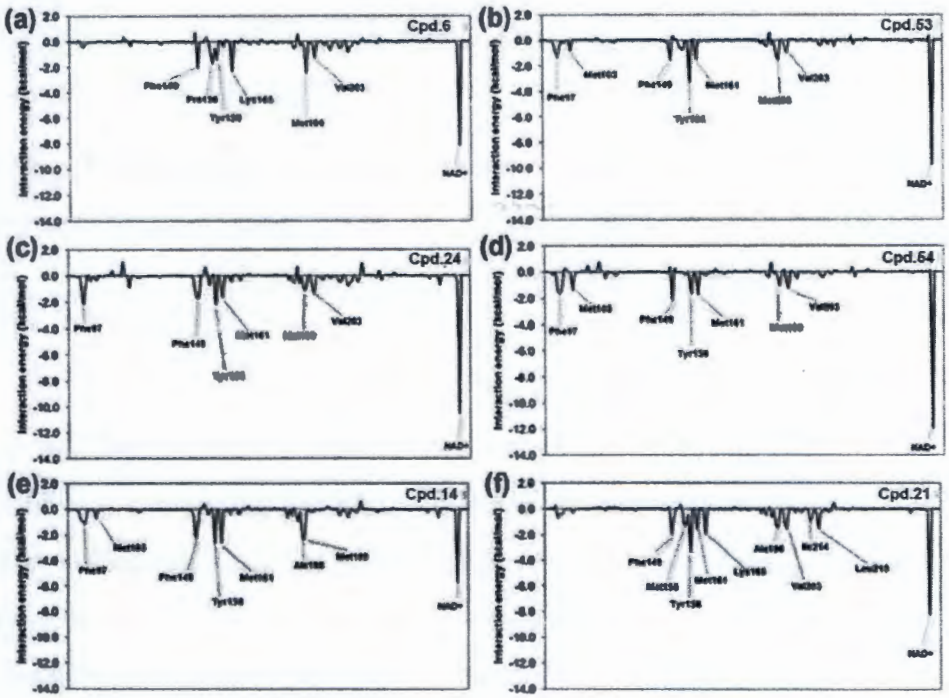


Figure 3.6 Per-residue binding energy decomposition of the selected diphenyl ether derivatives, compounds 6(a), 53(b), 24 (c) 54 (d), 14(e) and 21(f) using the MM-GBSA method

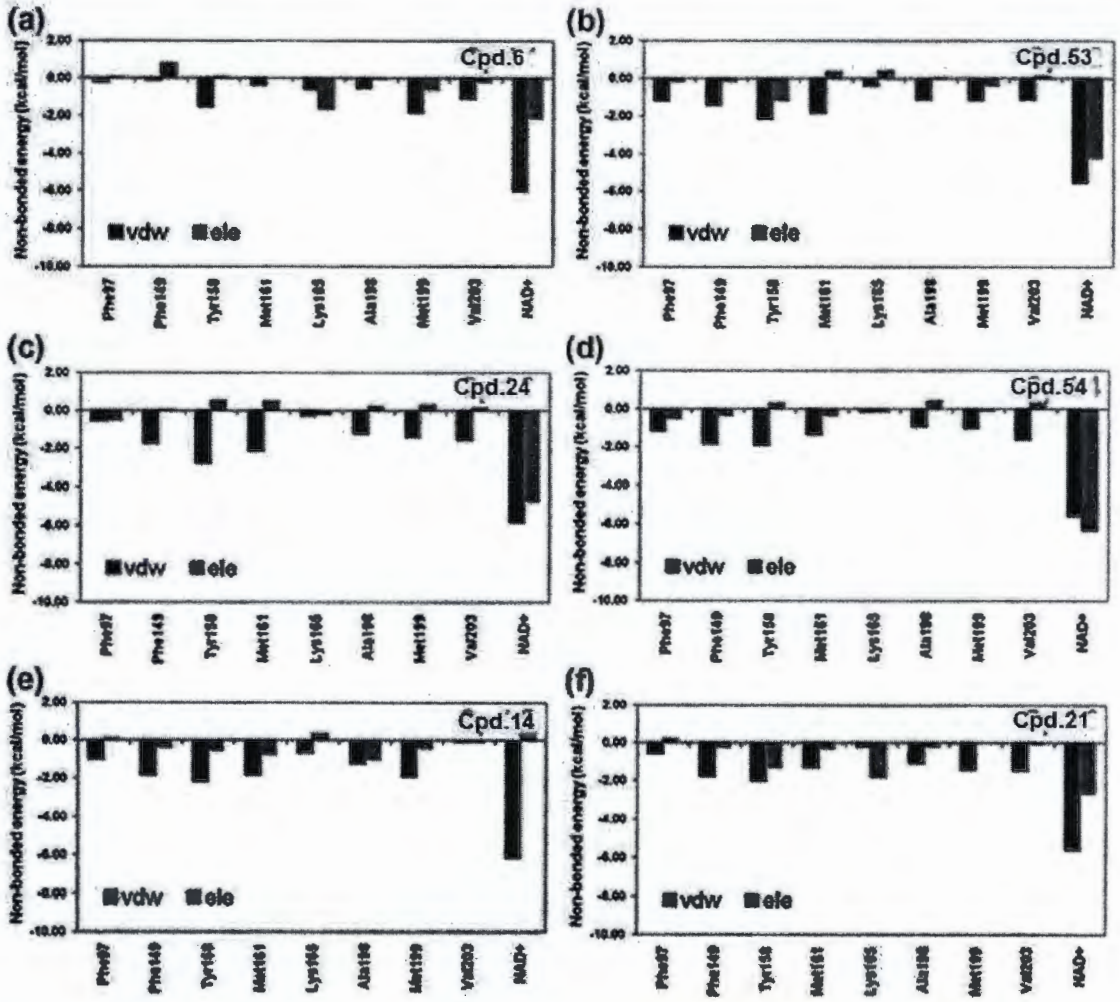


Figure 3.7 Plots of the decomposition energies in terms of van der Waals energy (vdw) and electrostatic energy (ele) for diphenyl ether derivatives of compounds 6(a), 53(b), 24(c) 54 (d), 5(e) and 6(f)

2.1.5) Binding interactions analysis

To elucidate the dynamic behavior of the diphenyl ether inhibitors in the InhA pocket, the binding modes and interaction of inhibitors in the InhA pocket obtained from MD simulations were analyzed. As shown in Figures 3.8a-3.8d, the binding modes of all compounds in the InhA pocket obtained from MD simulations are in the same manner. For compound **6**, the phenyl B ring reveals the pi-pi interaction with pyridine amide ring of NAD^+ cofactor. The hydrogen bond between the OH group of phenyl A ring and the NAD^+ cofactor is found. The substituent R_2 with Cl atom causes slightly repulsive interaction to the phenyl ring of Phe149. For

compound **53**, the binding interactions between diphenyl ether with NAD^+ can be explained as follows; (i) hydrogen bond interaction between –OH group of phenyl A ring with OH fragment of NAD^+ (ii) the pi-pi interaction of phenyl A ring and pyridine amide ring of NAD^+ cofactor (iii) weak hydrogen bond of Cl substituent and CH of NAD^+ cofactor. With regard to compound **24**, the crucial interactions of diphenyl ether compound 3 are hydrogen bond, pi-pi interactions and hydrophobic interactions with NAD^+ cofactor. The hydrogen bond interaction of OH group on phenyl A ring of diphenyl ether and OH of ribose fragment of NAD^+ cofactor is found. The pi-pi interactions between phenyl A ring of diphenyl ether and pyridine amide ring of NAD^+ cofactor can be formed and the hydrophobic interactions between phenyl B ring and all part of NAD^+ cofactor can be observed. Importantly, the hydrogen bond interactions between NO_2 substituent of diphenyl ether and CH_2 of Met98 are found. Moreover, the numerous hydrophobic interactions between the hexyl substituent attached to phenyl A ring of diphenyl ether with Tyr158 are observed. Compared to compound **24**, compound **54** also reveals similar interactions. However, only one hydrogen bond interaction between NO_2 substituent of diphenyl ether and Met98 is formed. The weaker interactions between diphenyl ether with Tyr158 are observed. Therefore, we can summarize the binding interaction of compounds **6, 24, 53** and **54** as follows. With regard to the phenyl B ring, the ring is surrounded by amino acid residues Met98, Phe97, Met161 and Ala198. The crucial interactions of this fragment are hydrogen bond interactions and van der Waals interactions. Regarding the phenyl A ring bearing the alkyl chain, the R_1 substituent of each compound is placed in the pocket of Phe149, Tyr158, Met199 and Val203. These residues show van der Waals interaction energies lower than electrostatic interaction energies for all compounds (Figure 3.7). This result could be accounted by forming of hydrophobic interactions between the alkyl chain at the R_1 substituent with surrounding residues. The OH group at the phenyl A ring of compounds **6, 24, 53** and **54** lies among the OH groups of Tyr158 and ribose fragment of NAD^+. Based on binding energy decomposition of each residue, NAD^+ shows the lowest interaction energy for all compounds (Figure 3.6) indicating the largest contribution of NAD^+ on binding of compounds **6, 24, 53** and **54** in the InhA pocket. All compounds could form two strong interactions with NAD^+. The first one is the hydrogen bond between the OH group of compound with the OH

group of ribose fragment. Another important interaction is the pi-pi interaction between the phenyl A ring of compounds with pyridine amide ring of NAD^+ (Figure 3.8). Considering the contribution of NAD^+ in terms of van der Waals and electrostatic interaction energies on the binding of compounds **6**, **24**, **53** and **54** in the InhA binding pocket (Figure 3.7), both van der Waals and electrostatic interactions show large attractive energy for all compounds. These results indicate that the pi-pi interaction as well as the hydrogen bond interaction with NAD^+ play important role on the binding of diphenyl ether derivatives in the InhA. Accordingly, apart from the phenyl B ring playing important role for forming strong hydrogen bond interactions to Met98, the key fragment for binding of diphenyl ether derivatives in the InhA is the phenyl A ring and the OH group that could generate the pi-pi interaction and the hydrogen bond interaction with NAD^+, respectively.

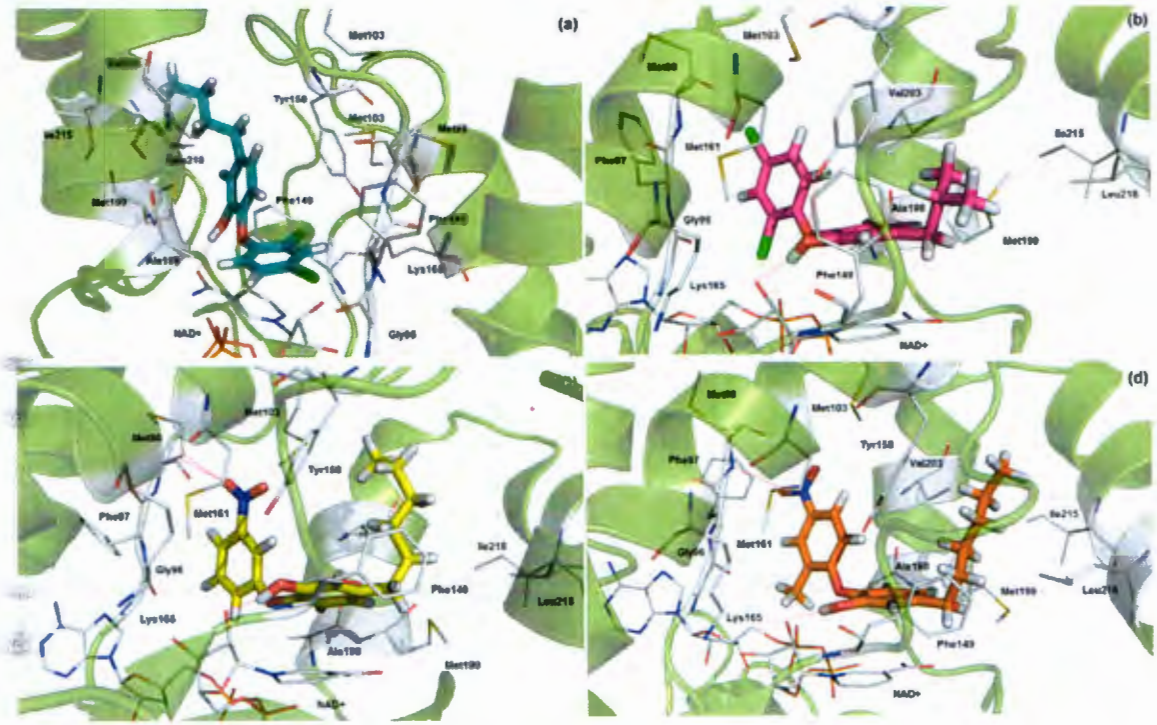


Figure 3.8 MD structure averaged over the last 2 ns of compounds 6(a), 53(b), 24 (c) and 54 (d) in the InhA binding pocket

2.1.6) Structural basis of diphenyl ether derivatives for rational inhibitor design

The reliable binding modes of diphenyl ether derivatives in the InhA binding pocket were obtained from MD simulations. The superposition of MD structures of compounds **6**, **24**, **53** and **54** and the X-ray structures of compounds **14** and **21** is shown in Figure 3.9. As mentioned above, the key fragment for binding of diphenyl ether derivatives in the InhA is the phenyl A ring and the OH group for generating the pi-pi interaction and the hydrogen bond interaction with NAD^+, respectively. The R_1 substituent on the phenyl A ring is oriented in the hydrophobic pocket of Phe149, Met155, Pro156, Ala157, Tyr158, Pro193, Met199, Val203, Leu207, Ile215 and Leu218 (Figure 3.9). Therefore, the R_1 substituent with highly lipophilic property i.e. hexyl, heptyl and octyl is optimal for forming hydrophobic interactions in this pocket. For the R_2, R_3 and R_4 substituents at the phenyl B ring, compound **21**, the highest active compound in this series, contains hydrogen atom at these positions. A hydrogen atom at the R_2 position of compound **21** could preferably form hydrophobic interaction with the methyl side chain of Ala198 and pyrophosphate group of NAD^+. The CH_3 group attached to the same position of compound **54** located close to these residues could also possibly form the hydrophobic interactions with Ala198 and pyrophosphate group of NAD^+. Whereas Cl atom attached to the R_2 position with respect to compounds **53** and **14** generate repulsive interaction with the pyrophosphate group of NAD^+. Accordingly, the hydrophobic substituents such as H and CH_3 should be optimal to interact with both methyl side chain of Ala198 and pyrophosphate group of NAD^+. In case of the R_3 position, H substituent at this position of compound **21** as well as those of compounds **6**, **53**, **54** and **14** point to the carbonyl backbone of Gly96. In contrast, the NO_2 substituent of compound **24** flips away from this direction (Figure 3.9). Therefore, H substituent as a hydrogen bond donor substituent would be the most suitable for the R_3 position to possibly form the hydrogen bond interaction with the carbonyl backbone of Gly96. For the R_4 substituent, H substituent at this position of compound **21** as well as Cl and NO_2 substituents of compounds **53**, **14** and **54**, respectively, are oriented to the NH backbone of Met98. Therefore, at the R_4 position, Cl and NO_2 substituents as well as hydrogen bond acceptor substituent that could form the hydrogen bond interaction

with the NH backbone of Met98 would be better for inhibitor-enzyme interaction than H substituent. Notably, the modification of the R_2, R_3 and R_4 substituents at the phenyl B ring should be optimal size because the steric effect of each substituent may cause the loss of the pi-pi interaction and the hydrogen bond interaction of the phenyl A ring bearing the OH group with NAD^+. To maintain these key interactions, size of the R_2, R_3 and R_4 substituents should be not too large.

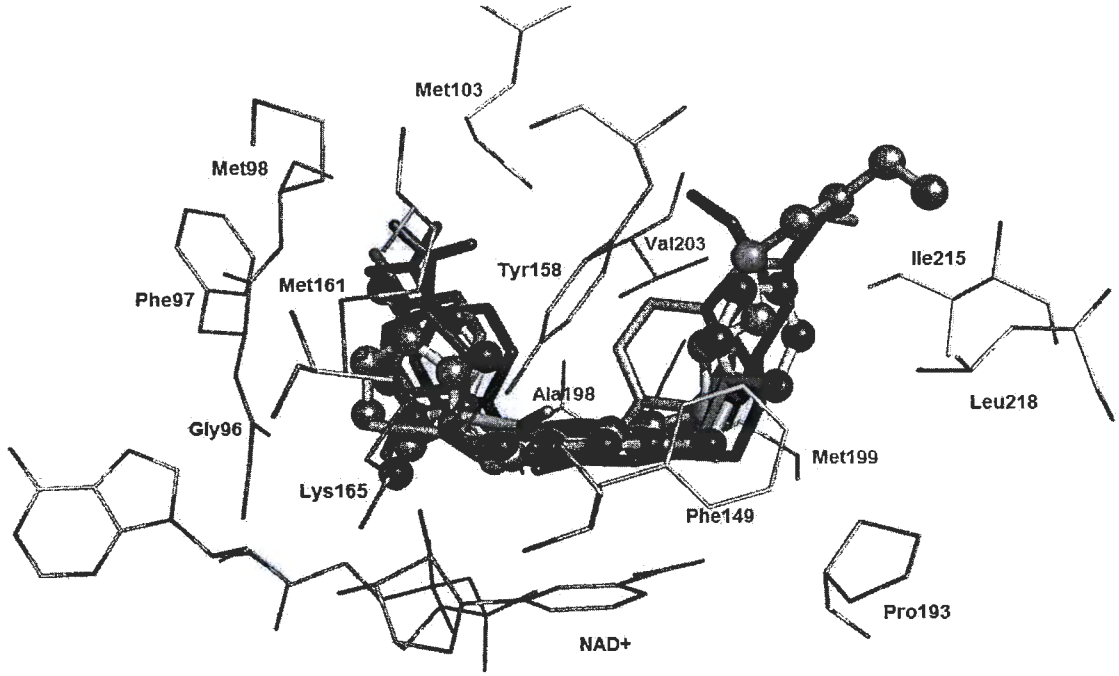


Figure 3.9 Superimposition of diphenyl ether derivatives in the InhA binding pocket. Compound 6 (stick in light grey color), Compound 53 (stick in grey color), Compound 24 (stick in dark grey color), Compound 54 (stick in black color), Compound 14 (ball and stick in dark grey color) and Compound 21 (ball and stick in light grey color)

2.2) MD simulations of diphenyl ether derivatives set 2

2.2.1) System equilibration

Four MD simulations of compounds **17**, **18**, **19**, and **29** bound with InhA were performed for 5 ns to evaluate the structural stability of the complexes and their binding strength. The RMSDs for all atoms of three different solute species (InhA, NAD^+ and inhibitor) relative to the initial structure over the 5 ns of simulation times were analyzed and plotted in Figure 3.10. The plateau

characteristic of the RMSD plot over the simulation time is the criteria to indicate the equilibrium state of each solute species. Figure 3.10 shows that NAD^+ and compounds **17**, **18**, **19** and **29** reach the equilibrium state at the early time. However, RMSDs of all compounds are more fluctuated, particularly compound **17**. InhA complexed with compounds **17**, **18**, **19** and **29** reach the equilibrium state after 1.0 ns (Figure 3.10a), 1.5 ns (Figure 3.10b), 2.5 ns (Figure 3.10c) and 1.0 ns (Figure 3.10d), respectively. Moreover, to reveal the energy stability of each system, the receptor-ligand interaction energies of compounds **17**, **18**, **19** and **29** over the 5 ns simulation time were calculated by MM-PBSA method. The receptor-ligand interaction energies of all compounds reach the equilibrium state at the beginning of the simulation time, except that of compound **29** which reaches the equilibrium state after the 0.5 ns simulation time (Figure 3.11). The average receptor-ligand interaction energies of compounds **17**, **18**, **19** and **29** are -58.85 ± 2.42, -49.27 ± 2.55, -53.85 ± 2.46 and -62.72 ± 3.55 kcal/mol, respectively. Based on the receptor-ligand interaction energy and RMSD plots, the all system are sufficiently stable and the production simulations are reliable. Therefore, the subsequent free energy calculation and free energy decomposition analysis based on snapshots extracted from the stable state are reasonable.

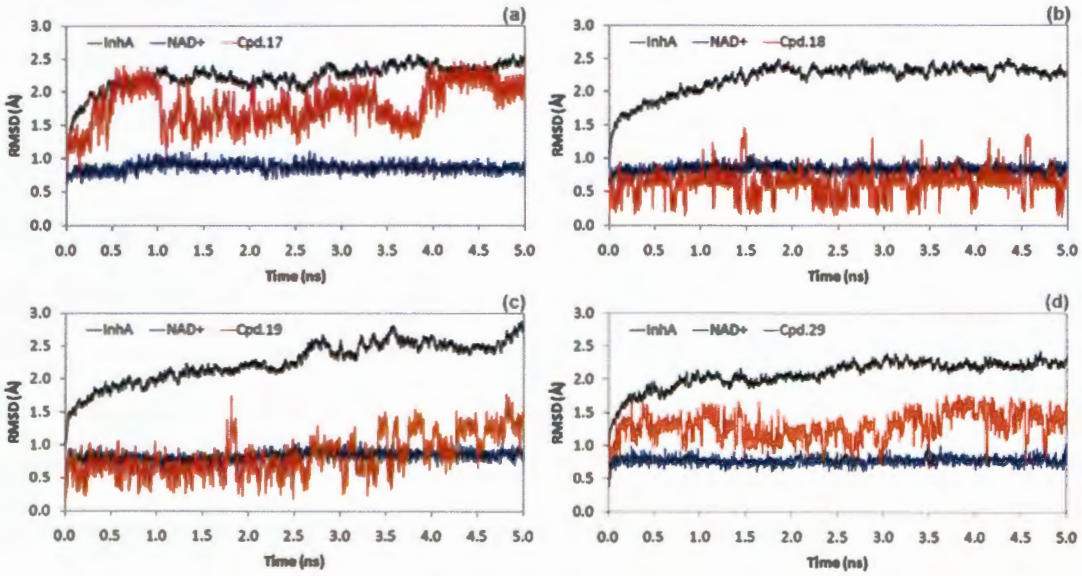


Figure 3.10 RMSD plots of compounds 17 (a), 18 (b), 19 (c) and 29 (d) in complexed with InhA and NAD^+

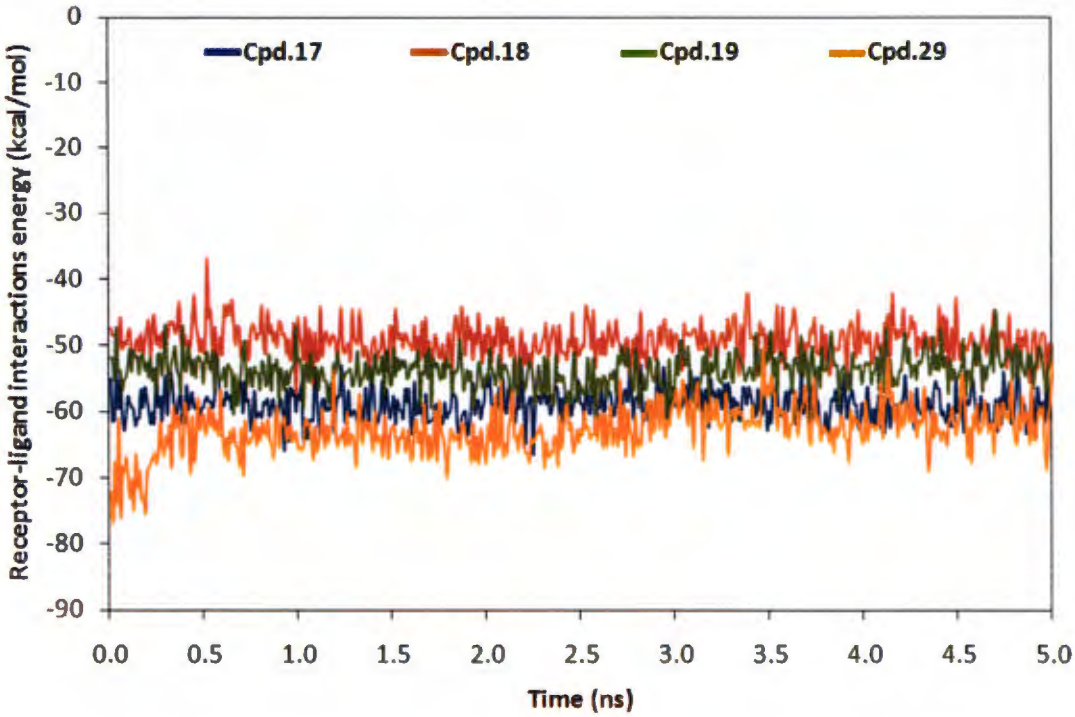


Figure 3.11 Recepter-ligand interactions energies for the systems of compounds 17 (a), 18 (b), 19 (c) and 29 (d) over the 5ns simulation

2.2.2) Binding free energy calculations

The MM-PBSA method was employed to calculate the binding free energies of compounds **17**, **18**, **19** and **29** in InhA and the results are shown in Table 3.4. The binding free energies (ΔG_{bind}) of compounds **17**, **18**, **19** and **29** bound to the InhA pocket are calculated to be -15.02, -9.03, -13.90 and -15.40 kcal/mol, respectively, which are in good agreement with those determined experimentally ($\Delta G_{exp.}$). Pearson correlation and Spearman rank correlation (Zar, 1998) were employed to determine the correlation between $\Delta G_{exp.}$ and ΔG_{bind}. The accepted values of correlation coefficient are in range of -1 to 1. Based on these methods, the correlation between ΔG_{exp} and ΔG_{bind} shows the correlation coefficient of Pearson correlation and Spearman rank correlation to be 0.98 and 1.00, respectively. Therefore, there is the correlation between $\Delta G_{exp.}$ and ΔG_{bind}.

Table 3.4 Binding free energies (kcal/mol) calculated by the MM-PBSA method

Component	Diphenyl ether-InhA complexes			
	17	**18**	**19**	**29**
ΔG_{MM}	-58.77 ± 2.59	-49.49 ± 2.29	-52.90 ± 2.57	-60.82 ± 2.91
$\Delta G_{solv.}$	21.59 ± 2.07	19.33 ± 1.23	20.13 ± 1.60	22.70 ± 2.69
ΔH	-37.18 ± 2.93	-30.16 ± 2.24	-32.77 ± 2.47	-38.06 ± 3.21
$-T\Delta S$	22.16 ± 0.85	21.13 ± 1.17	18.87 ± 1.06	22.66 ± 0.57
$\Delta G_{bind.}$	-15.02 ± 1.32	-9.03 ± .084	-13.90 ± 1.31	-15.40 ± 1.40
$\Delta G_{exp.}$ [a]	-10.93	-7.83	-9.75	-10.99

[a] derived from $\Delta G = RT \ln[IC_{50}]$, R represents the gas constant (1.988 cal/mol K), T represents the temperature (300 K).

2.2.3) The binding modes of diphenyl ether derivatives in InhA binding pocket

The binding modes of compounds **17**, **18**, **19** and **29** bound with InhA pocket observed from the simulations are superimposed and illustrated in Figure 3.12. In general, all compounds showed a similar binding mode and conformation: the OH group of the phenyl A ring lies in between the OH groups of Tyr158 and ribose fragment of NAD^+ to form the hydrogen bond interactions. The phenyl A ring forms the pi-pi interaction with pyridine amide ring of NAD^+. As the phenyl A ring bearing the R_1 substituent as the alkyl chain, it is placed in the hydrophobic pocket that is formed by Phe149, Met155, Pro156, Ala157, Tyr158, Pro193, Met199, Val203, Leu207, Ile215 and Leu218 (Figure 3.12). Compounds **17** and **29** that holding the hexyl substituents at the R_1 position could form stronger hydrophobic interactions with Phe149, Met155, Pro156, Ala157, Tyr158, Pro193, Met199, Val203, Leu207, Ile215 and Leu218 when comparing these interactions with compounds **18** and **19** that have shorter alkyl substituents (containing ethyl and butyl, respectively), loosing several hydrophobic interactions with Pro156, Ala157, Val203, Leu207 and Ile215. Therefore, the more hydrophobic interactions at the R_1 position of compounds **17** and **29** should account for better activities against InhA. The phenyl B

ring containing the R_2, R_3 and R_4 substituents is surrounded by the pyrophosphate moiety of NAD^+, the hydrophilic backbones of Gly96, Met98, Phe97 and the hydrophobic side chains of Met103, Met161, Ile202, Val203, Ala198. The H and Br substituents at the R_2 position for compounds **17** and **29**, respectively, are closed to the methyl sidechain of Ala198 and the pyrophosphate moiety of NAD^+ (Figure 3.13). The Br substituent of compound **29** contributes greatly a hydrophobic interaction to the methyl sidechain of Ala198 while the H substituent of compound **17** contributes a hydrophilic interaction to the ribose and pyrophosphate moieties of NAD^+. These results might be explained why compounds **29** and **17** show the InhA inhibitory activities in the same level with IC_{50} of 10 and 11 nM, respectively. Accordingly, the R_2 substituent would also be hydrophobic or hydrophilic groups. For R_3 position, the H substituents at this position for compounds **17**, **18**, **19** and **29** form a hydrogen bond interaction with the carbonyl backbone of Gly96 and, besides the H substituent, other hydrogen bond donor substituent would also be possible. This similar H-bond interaction was also found for the R_4 substituent where all four compounds point to the NH and carbonyl backbone of Met98.

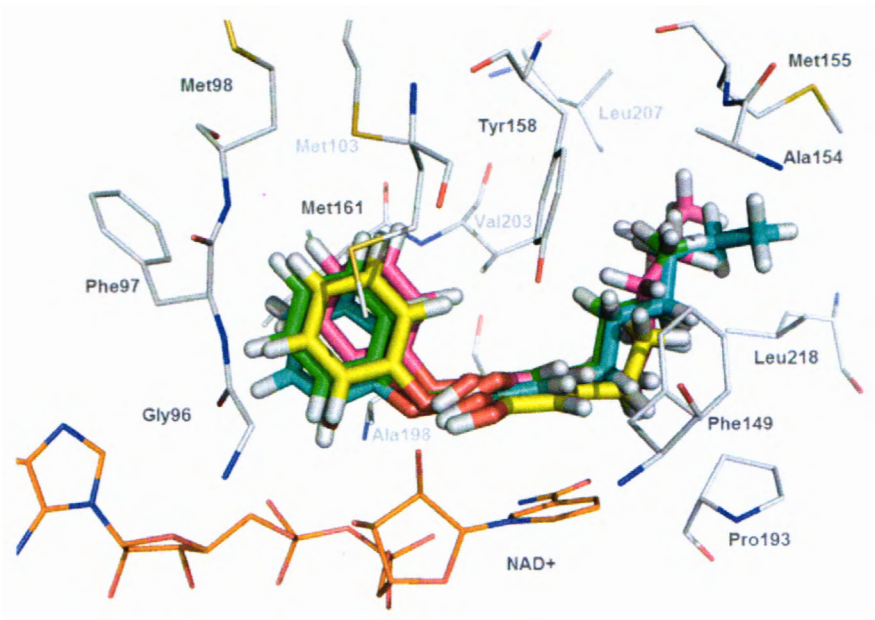


Figure 3.12 The superimposition of compounds 17 (stick in cyan color), 18 (stick in yellow color), 19 (stick in green color) and 29 (stick in pink color) in the InhA pocket obtained from MD simulation

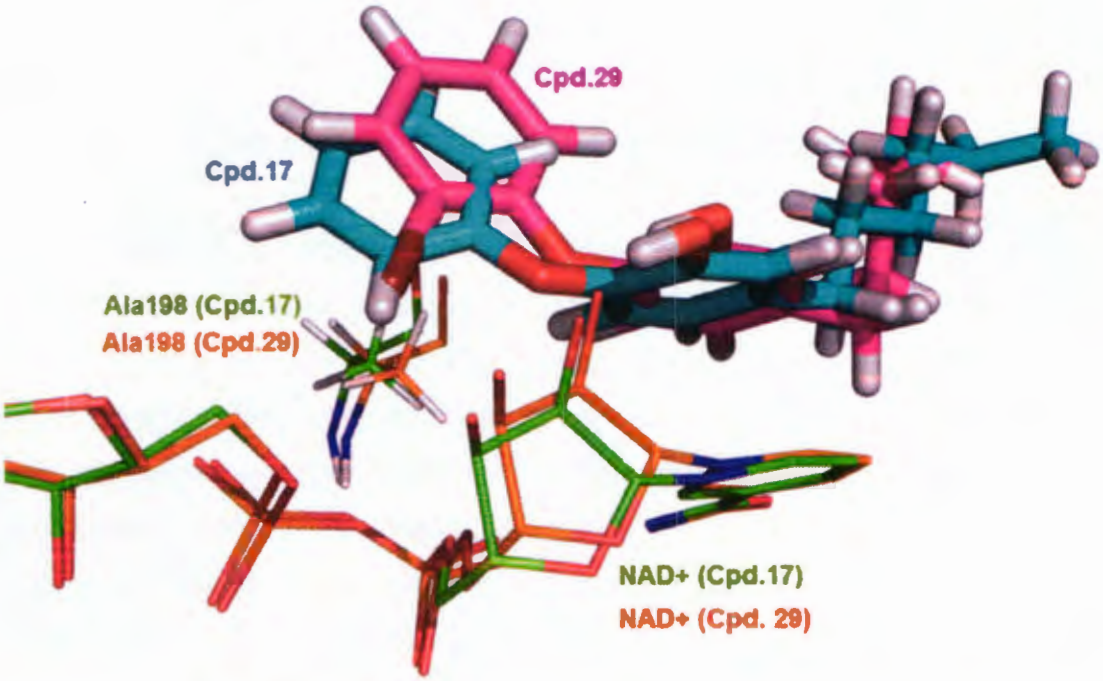


Figure 3.13 Interactions of the R_2 substituents of compounds 17 and 29 with Ala198 and the pyrophosphate moiety of NAD^+

3) Design new and more potent diphenyl ether derivatives as direct InhA inhibitors

The obtained results from CoMSIA and MD simulations are successful to recommend the structural requirement for designing new diphenyl ether derivatives. The substitutions of diphenyl ether derivatives were defined as R_1, R_2, R_3 and R_4 positions. The suggestion of CoMSIA contour maps and MD simulations, hydrophobic longer alkyl chain was required to form hydrophobic interaction with amino acids in InhA binding pocket at R_1 position. Therefore, two compounds were selected as a template scaffold for designing new compounds. Compound **21**, the highest active compound contained octyl alkyl chain at R_1 position was selected. The second highly active compound, compound **29**, **31** and **35** was selected as initial scaffold to design new diphenyl ether derivatives with hexyl substituent at R_1 position. The general structure of diphenyl ether derivatives is shown in Figure 3.14.

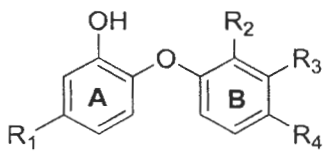


Figure 3.14 General structure of diphenyl ether derivatives

Compound **21** scaffold contained octyl chain at R_1 position was used to design new diphenyl ether derivatives (OD series). R_1, R_3 and R_4 of this compound were kept. The small substituent with hydrophilic or hydrophobic property was designed as R_2 substituent in this series. Twenty compounds (**OD1-OD20**) were designed. Halogen atoms, F, Cl, Br and I (**OD1-OD4**) were introduced into R_2 position. Based on predicted activity of halogen substituents (predicted activity in range of 8.21-8.28) seem to be produce predicted biological activity higher than parent compound **21**, except fluoro (**OD1**) substituent. Introduced small hydrophilic substituents, OH, NH_2 and SH (**OD5-OD7**) on R_2 position of compound **21**, the predicted biological activity of designed compounds (predicted activity in range of 7.73-8.19) was lower than parent compound **21**. The modification of R_2 position with small hydrophobic substituents, CH_3, OCH_3, $NHCH_3$ and SCH_3 (**OD8-OD11**) produced low predicted activity (predicted activity in range of 7.77-8.19) as compared to parent compound. Based on the results from design compounds **OD1-OD11**, halogen substituents at R_2 position were together modified. Compounds **OD12-OD20**, the mono-, di- and tri-halogenated methyl substituents at R_2 position were designed. The obtained results showed that tri-halogenated methyl substituents, CF_3 (**OD14**), CCl_3 (**OD17**) and CBr_3 (**OD20**) were slightly higher than parent compound with predicted activity of 8.24, 8.25 and 8.23 respectively.

Table 3.5 Chemical structure and predicted activity of new designed diphenyl ether derivatives utilizing from compound 21

Cpd.	R_1	R_2	CoMSIA[a]
21	$(CH_2)_7CH_3$	H	8.23
OD1	$(CH_2)_7CH_3$	F	8.21
OD2[b]	**$(CH_2)_7CH_3$**	**Cl**	**8.25**
OD3[b]	**$(CH_2)_7CH_3$**	**Br**	**8.27**
OD4[b]	**$(CH_2)_7CH_3$**	**I**	**8.29**
OD5	$(CH_2)_7CH_3$	OH	8.01
OD6	$(CH_2)_7CH_3$	NH_2	7.73
OD7	$(CH_2)_7CH_3$	SH	8.19
OD8	$(CH_2)_7CH_3$	CH_3	8.19
OD9	$(CH_2)_7CH_3$	OCH_3	8.10
OD10	$(CH_2)_7CH_3$	$NHCH_3$	7.77
OD11	$(CH_2)_7CH_3$	SCH_3	8.16
OD12	$(CH_2)_7CH_3$	CH_2F	8.14
OD13	$(CH_2)_7CH_3$	CHF_2	8.16
OD14[b]	**$(CH_2)_7CH_3$**	**CF_3**	**8.24**
OD15	$(CH_2)_7CH_3$	CH_2Cl	8.15
OD16	$(CH_2)_7CH_3$	$CHCl_2$	8.17
OD17[b]	**$(CH_2)_7CH_3$**	**CCl_3**	**8.25**
OD18	$(CH_2)_7CH_3$	CH_2Br	8.14
OD19	$(CH_2)_7CH_3$	$CHBr_2$	8.15
OD20[b]	**$(CH_2)_7CH_3$**	**CBr_3**	**8.23**

[a] predicted $\log(1/IC_{50})$ derived from best CoMSIA model

[b] Bold letter represents the better predicted biological activity

Based on parent compound **29**, R_1 and R_4 positions were kept as hexyl and hydrogen, whereas R_2 and R_3 were modified. The 155 new designed compounds were summarized in Table 3.6, Table 3.7 and Table 3.8. As considered experimental data, halogenate and small hydrophilic substitutions on R_2 position were reported. Therefore, small hydrophobic, mono-, di and tri-halogenated methyl were modified (**HD1-HD11**). The obtained results showed that predicted activity of new designed compounds were lower than parent compounds (predicted activity in range of 7.42-7.81). Therefore, the halogenate substitutions; Br, F and Cl from experimental data were used as starting structure to design new diphenyl ether derivatives (**HD12-HD155**). Small hydrophilic substituents suggested from CoMSIA and MD simulation were modified at R_3 position. New designed compounds with R_2 position contained Br with small hydrophilic substitutions at R_3 position showed higher predicted biological activity than parent compounds. New designed diphenyl ether with F substituent at R_2 position and small hydrophilic substituents at R_3 position (**HD60-HD107**) were considered. The obtained results showed that only **HD71** (predicted activity of 7.95) designed compound showed predicted activity higher than parent compounds **31** and **29**. For designed diphenyl ether with Cl at R_2 position and small hydrophilic substituents at R_3 position (**HD108-HD155**), high predicted activity than parent compound **35** of new designed compounds were obtained, except compound **HD138** (predicted activity of 7.95). Among of these compounds, 3 compounds (**HD114**, **HD115** and **HD123**) show higher predicted activity than compound **29**.

Table 3.6 Chemical structure and predicted activity of new designed diphenyl ether derivatives utilizing from compound 29

Cpd.	R_1	R_2	R_3	CoMSIA[a]
29	$(CH_2)_5CH_3$	Br	H	7.93
HD1	$(CH_2)_5CH_3$	OCH_3	H	7.44
HD2	$(CH_2)_5CH_3$	$NHCH_3$	H	7.42
HD3	$(CH_2)_5CH_3$	SCH_3	H	7.80
HD4	$(CH_2)_5CH_3$	CH_2F	H	7.72
HD5	$(CH_2)_5CH_3$	CHF_2	H	7.76
HD6	$(CH_2)_5CH_3$	CH_2Cl	H	7.76
HD7	$(CH_2)_5CH_3$	$CHCl_2$	H	7.81
HD8	$(CH_2)_5CH_3$	CH_2Br	H	7.76
HD9	$(CH_2)_5CH_3$	$CHBr_2$	H	7.77
HD10	$(CH_2)_5CH_3$	$CHCH_2$	H	7.74
HD11	$(CH_2)_5CH_3$	CCH	H	7.43
HD12	$(CH_2)_5CH_3$	Br	OH	7.83
HD13	$(CH_2)_5CH_3$	Br	SH	7.89
HD14	$(CH_2)_5CH_3$	Br	NH_2	7.84
HD15	$(CH_2)_5CH_3$	Br	COH	7.92
HD16[b]	**$(CH_2)_5CH_3$**	**Br**	**$COCH_3$**	**7.98**
HD17	$(CH_2)_5CH_3$	Br	$NHCH_3$	7.89
HD18[b]	**$(CH_2)_5CH_3$**	**Br**	**COOH**	**7.98**
HD19[b]	**$(CH_2)_5CH_3$**	**Br**	**$CONH_2$**	**8.01**
HD20[b]	**$(CH_2)_5CH_3$**	**Br**	**CSH**	**7.96**
HD21[b]	**$(CH_2)_5CH_3$**	**Br**	**$CSCH_3$**	**7.97**
HD22[b]	**$(CH_2)_5CH_3$**	**Br**	**CSOH**	**7.97**
HD23	$(CH_2)_5CH_3$	Br	$CSNH_2$	7.85
HD24	$(CH_2)_5CH_3$	Br	CH_2OH	7.71
HD25	$(CH_2)_5CH_3$	Br	CH_2SH	7.71
HD26	$(CH_2)_5CH_3$	Br	CH_2NH_2	7.82
HD27[b]	**$(CH_2)_5CH_3$**	**Br**	**$CONHCH_3$**	**8.01**
HD28[b]	**$(CH_2)_5CH_3$**	**Br**	**CONHOH**	**7.94**
HD29[b]	**$(CH_2)_5CH_3$**	**Br**	**$CONHNH_2$**	**7.94**

[a] predicted $\log(1/IC_{50})$ derived from best CoMSIA model

[b] Bold letter represents the better predicted biological activity

Table 3.6 Chemical structure and predicted activity of new designed diphenyl ether derivatives utilizing from compound 29 (continue)

Cpd.	R_1	R_2	R_3	CoMSIA[a]
HD30[b]	**$(CH_2)_5CH_3$**	**Br**	**CONHSH**	**7.97**
HD31	$(CH_2)_5CH_3$	Br	CH_2COH	7.88
HD32	$(CH_2)_5CH_3$	Br	CH_2COCH_3	7.90
HD33	$(CH_2)_5CH_3$	Br	CH_2COOH	7.89
HD34	$(CH_2)_5CH_3$	Br	CH_2CONH_2	7.89
HD35	$(CH_2)_5CH_3$	Br	CH_2CSH	7.89
HD36	$(CH_2)_5CH_3$	Br	CH_2CSCH_3	7.90
HD37	$(CH_2)_5CH_3$	Br	CH_2CSOH	7.89
HD38	$(CH_2)_5CH_3$	Br	CH_2CSNH_2	7.90
HD39	$(CH_2)_5CH_3$	Br	$NHCH_2OH$	7.90
HD40	$(CH_2)_5CH_3$	Br	$NHCH_2SH$	7.89
HD41[b]	**$(CH_2)_5CH_3$**	**Br**	**$NHCH_2NH_2$**	**7.94**
HD42	$(CH_2)_5CH_3$	Br	$NHNH_2$	7.22
HD43	$(CH_2)_5CH_3$	Br	$NHCH_2F$	7.89
HD44	$(CH_2)_5CH_3$	Br	$NHCH_2Cl$	7.88
HD45	$(CH_2)_5CH_3$	Br	$NHCH_2Br$	7.88
HD46	$(CH_2)_5CH_3$	Br	$NHCHF_2$	7.88
HD47	$(CH_2)_5CH_3$	Br	$NHCHCl_2$	7.86
HD48	$(CH_2)_5CH_3$	Br	$NHCHBr_2$	7.86
HD49	$(CH_2)_5CH_3$	Br	$NHCF_3$	7.87
HD50	$(CH_2)_5CH_3$	Br	$NHCCl_3$	7.83
HD51	$(CH_2)_5CH_3$	Br	$NHCBr_3$	7.83
HD52	$(CH_2)_5CH_3$	Br	NHCOH	7.93
HD53	$(CH_2)_5CH_3$	Br	$NHCOCH_3$	7.92
HD54	$(CH_2)_5CH_4$	Br	NHCOOH	7.93
HD55[b]	**$(CH_2)_5CH_5$**	**Br**	**$NHCONH_2$**	**7.95**
HD56	$(CH_2)_5CH_3$	Br	NHCSH	7.89
HD57	$(CH_2)_5CH_3$	Br	$NHCSCH_3$	7.88
HD58	$(CH_2)_5CH_3$	Br	NHCSOH	7.90
HD59	$(CH_2)_5CH_3$	Br	$NHCSNH_2$	7.92

[a] predicted $\log(1/IC_{50})$ derived from best CoMSIA model

[b] Bold letter represents the better predicted biological activity

Table 3.7 Chemical structure and predicted activity of new designed diphenyl ether derivatives utilizing from compound 31

Cpd.	R_1	R_2	R_3	CoMSIA[a]
31	$(CH_2)_5CH_3$	F	H	7.92
HD60	$(CH_2)_5CH_3$	F	OH	7.68
HD61	$(CH_2)_5CH_3$	F	SH	7.72
HD62	$(CH_2)_5CH_3$	F	NH_2	7.67
HD63	$(CH_2)_5CH_3$	F	COH	7.74
HD64	$(CH_2)_5CH_3$	F	$COCH_3$	7.80
HD65	$(CH_2)_5CH_3$	F	$NHCH_3$	7.69
HD66	$(CH_2)_5CH_3$	F	COOH	7.82
HD67	$(CH_2)_5CH_3$	F	$CONH_2$	7.84
HD68	$(CH_2)_5CH_3$	F	CSH	7.80
HD69	$(CH_2)_5CH_3$	F	$CSCH_3$	7.85
HD70	$(CH_2)_5CH_3$	F	CSOH	7.80
HD71[b]	**$(CH_2)_5CH_3$**	**F**	**$CSNH_2$**	**7.95**
HD72	$(CH_2)_5CH_3$	F	CH_2OH	7.57
HD73	$(CH_2)_5CH_3$	F	CH_2SH	7.58
HD74	$(CH_2)_5CH_3$	F	CH_2NH_2	7.65
HD75	$(CH_2)_5CH_3$	F	$CONHCH_3$	7.84
HD76	$(CH_2)_5CH_3$	F	CONHOH	7.78
HD77	$(CH_2)_5CH_3$	F	$CONHNH_2$	7.77
HD78	$(CH_2)_5CH_3$	F	CONHSH	7.80
HD79	$(CH_2)_5CH_3$	F	CH_2COH	7.69
HD80	$(CH_2)_5CH_3$	F	CH_2COCH_3	7.74
HD81	$(CH_2)_5CH_3$	F	CH_2COOH	7.71
HD82	$(CH_2)_5CH_3$	F	CH_2CONH_2	7.71
HD83	$(CH_2)_5CH_3$	F	CH_2CSH	7.74
HD84	$(CH_2)_5CH_3$	F	CH_2CSCH_3	7.77
HD85	$(CH_2)_5CH_3$	F	CH_2CSOH	7.74
HD86	$(CH_2)_5CH_3$	F	CH_2CSNH_2	7.74
HD87	$(CH_2)_5CH_3$	F	$NHCH_2OH$	7.71
HD88	$(CH_2)_5CH_3$	F	$NHCH_2SH$	7.70

[a] predicted $\log(1/IC_{50})$ derived from best CoMSIA model

[b] Bold letter represents the better predicted biological activity

Table 3.7 Chemical structure and predicted activity of new designed diphenyl ether derivatives utilizing from compound 31 (continued)

Cpd.	R_1	R_2	R_3	CoMSIA[a]
HD89	$(CH_2)_5CH_3$	F	$NHCH_2NH_2$	7.75
HD90	$(CH_2)_5CH_3$	F	$NHNH_2$	7.19
HD91	$(CH_2)_5CH_3$	F	$NHCH_2F$	7.70
HD92	$(CH_2)_5CH_3$	F	$NHCH_2Cl$	7.69
HD93	$(CH_2)_5CH_3$	F	$NHCH_2Br$	7.69
HD94	$(CH_2)_5CH_3$	F	$NHCHF_2$	7.69
HD95	$(CH_2)_5CH_3$	F	$NHCHCl_2$	7.68
HD96	$(CH_2)_5CH_3$	F	$NHCHBr_2$	7.68
HD97	$(CH_2)_5CH_3$	F	$NHCF_3$	7.68
HD98	$(CH_2)_5CH_3$	F	$NHCCl_3$	7.65
HD99	$(CH_2)_5CH_3$	F	$NHCBr_3$	7.65
HD100	$(CH_2)_5CH_3$	F	NHCOH	7.72
HD101	$(CH_2)_5CH_4$	F	NHCOOH	7.72
HD102	$(CH_2)_5CH_5$	F	$NHCONH_2$	7.74
HD103	$(CH_2)_5CH_3$	F	$NHCONH_2$	7.75
HD104	$(CH_2)_5CH_3$	F	NHCSH	7.69
HD105	$(CH_2)_5CH_3$	F	$NHCSCH_3$	7.69
HD106	$(CH_2)_5CH_3$	F	NHCSOH	7.71
HD107	$(CH_2)_5CH_3$	F	$NHCSNH_2$	7.73

[a] predicted $\log(1/IC_{50})$ derived from best CoMSIA model

[b] Bold letter represents the better predicted biological activity

Table 3.8 Chemical structure and predicted activity of new designed diphenyl ether derivatives utilizing from compound 35

Cpd.	R_1	R_2	R_3	CoMSIA[a]
35	$(CH_2)_5CH_3$	Cl	H	7.51
HD108	$(CH_2)_5CH_3$	Cl	OH	7.78
HD109	$(CH_2)_5CH_3$	Cl	SH	7.84
HD110	$(CH_2)_5CH_3$	Cl	NH_2	7.78
HD111	$(CH_2)_5CH_3$	Cl	COH	7.85
HD112	$(CH_2)_5CH_3$	Cl	$COCH_3$	7.92
HD113	$(CH_2)_5CH_3$	Cl	$NHCH_3$	7.83
HD114[b]	**$(CH_2)_5CH_3$**	**Cl**	**COOH**	**7.93**
HD115[b]	**$(CH_2)_5CH_3$**	**Cl**	**$CONH_2$**	**7.95**
HD116	$(CH_2)_5CH_3$	Cl	CSH	7.90
HD117	$(CH_2)_5CH_3$	Cl	$CSCH_3$	7.90
HD118	$(CH_2)_5CH_3$	Cl	CSOH	7.91
HD119	$(CH_2)_5CH_3$	Cl	$CSNH_2$	7.78
HD120	$(CH_2)_5CH_3$	Cl	CH_2OH	7.66
HD121	$(CH_2)_5CH_3$	Cl	CH_2SH	7.66
HD122	$(CH_2)_5CH_3$	Cl	CH_2NH_2	7.76
HD123[b]	**$(CH_2)_5CH_3$**	**Cl**	**$CONHCH_3$**	**7.95**
HD124	$(CH_2)_5CH_3$	Cl	CONHOH	7.89
HD125	$(CH_2)_5CH_3$	Cl	$CONHNH_2$	7.89
HD126	$(CH_2)_5CH_3$	Cl	CONHSH	7.91
HD127	$(CH_2)_5CH_3$	Cl	CH_2COH	7.82
HD128	$(CH_2)_5CH_3$	Cl	CH_2COCH_3	7.84
HD129	$(CH_2)_5CH_3$	Cl	CH_2COOH	7.83
HD130	$(CH_2)_5CH_3$	Cl	CH_2CONH_2	7.83
HD131	$(CH_2)_5CH_3$	Cl	CH_2CSH	7.84
HD132	$(CH_2)_5CH_3$	Cl	CH_2CSCH_3	7.85
HD133	$(CH_2)_5CH_3$	Cl	CH_2CSOH	7.84
HD134	$(CH_2)_5CH_3$	Cl	CH_2CSNH_2	7.84
HD135	$(CH_2)_5CH_3$	Cl	$NHCH_2OH$	7.84
HD136	$(CH_2)_5CH_3$	Cl	$NHCH_2SH$	7.83

[a] predicted $\log(1/IC_{50})$ derived from best CoMSIA model

[b] Bold letter represents the better predicted biological activity

Table 3.8 Chemical structure and predicted activity of new designed diphenyl ether derivatives utilizing from compound 35 (continued)

Cpd.	R_1	R_2	R_3	CoMSIA[a]
HD137	$(CH_2)_5CH_3$	Cl	$NHCH_2NH_2$	7.88
HD138	$(CH_2)_5CH_3$	Cl	$NHNH_2$	7.19
HD139	$(CH_2)_5CH_3$	Cl	$NHCH_2F$	7.83
HD140	$(CH_2)_5CH_3$	Cl	$NHCH_2Cl$	7.82
HD141	$(CH_2)_5CH_3$	Cl	$NHCH_2Br$	7.82
HD142	$(CH_2)_5CH_3$	Cl	$NHCHF_2$	7.82
HD143	$(CH_2)_5CH_3$	Cl	$NHCHCl_2$	7.80
HD144	$(CH_2)_5CH_3$	Cl	$NHCHBr_2$	7.80
HD145	$(CH_2)_5CH_3$	Cl	$NHCF_3$	7.80
HD146	$(CH_2)_5CH_3$	Cl	$NHCCl_3$	7.77
HD147	$(CH_2)_5CH_3$	Cl	$NHCBr_3$	7.77
HD148	$(CH_2)_5CH_3$	Cl	NHCOH	7.86
HD149	$(CH_2)_5CH_4$	Cl	NHCOOH	7.85
HD150	$(CH_2)_5CH_5$	Cl	$NHCONH_2$	7.87
HD151	$(CH_2)_5CH_3$	Cl	$NHCONH_2$	7.89
HD152	$(CH_2)_5CH_3$	Cl	NHCSH	7.82
HD153	$(CH_2)_5CH_3$	Cl	$NHCSCH_3$	7.82
HD154	$(CH_2)_5CH_3$	Cl	NHCSOH	7.86
HD155	$(CH_2)_5CH_3$	Cl	$NHCSNH_2$	7.86

[a] predicted $\log(1/IC_{50})$ derived from best CoMSIA model

[b] Bold letter represents the better predicted biological activity

4) Promising diphenyl ether based on rational design

155 designed diphenyl ether based on the structural requirement derived from 3D-QSAR and MD simulations. Among of these compounds, high predicted biological activity against InhA of 22 compounds was obtained as summarized in Table 3.9. To ensure that designed diphenyl ether compounds favorable to bind with InhA, molecular docking was used to predict the binding mode of new designed compounds in InhA binding site. Based on docking calculation, similar binding mode of ten compounds (**OD02, OD03, OD04, OD14, OD17, OD20, HD20,**

HD41, **HD115** and **HD123**) as compared to the X-ray binding mode of diphenyl ether compounds was obtained (Figure 3.15).

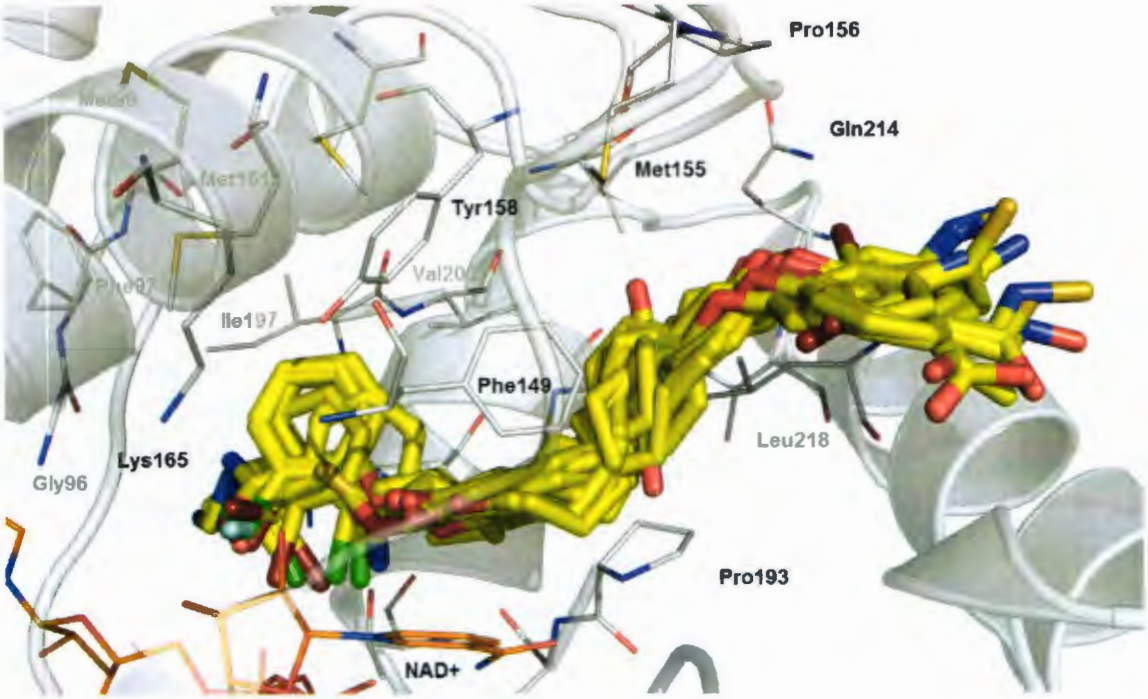
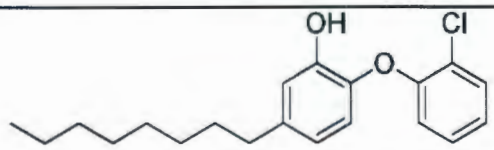
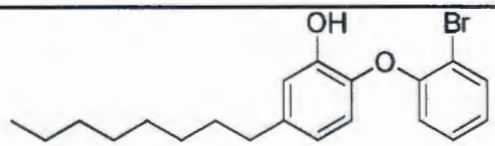
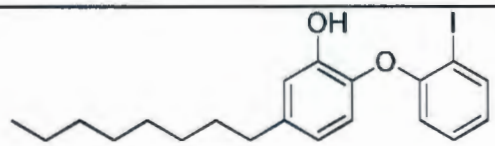


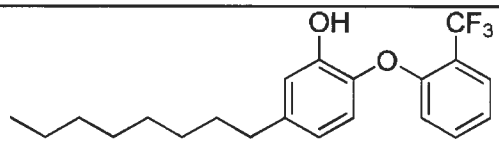
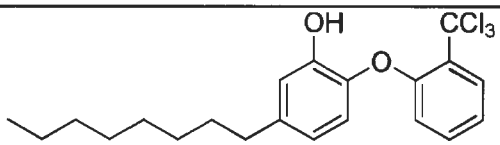
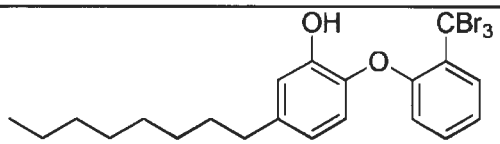
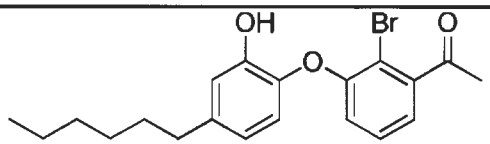
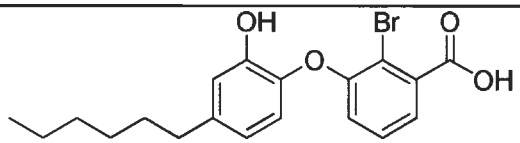
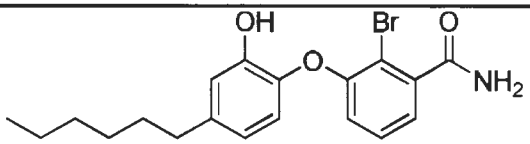
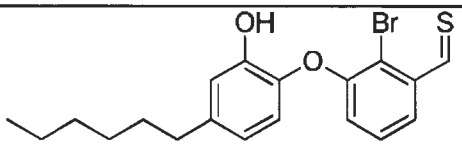
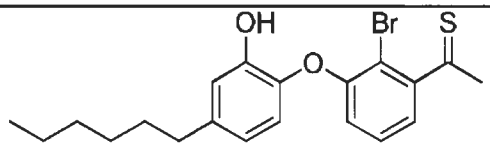
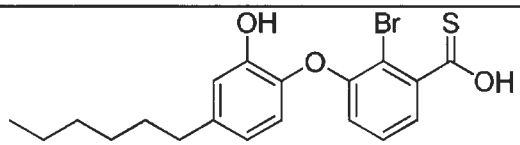
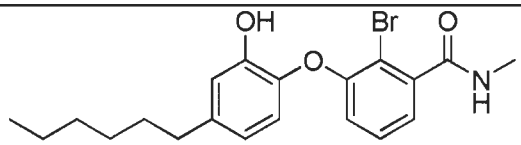
Figure 3.15 Superimposition of designed diphenyl ether in InhA binding pocket

Table 3.9 Highly predicted activity diphenyl ether compounds

Cpd.	Structure	CoMSIA[a]
OD2		8.25
OD3		8.27
OD4		8.29

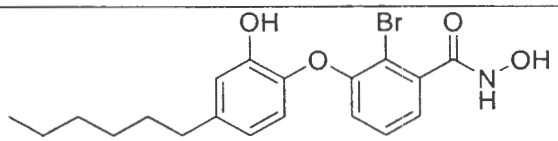
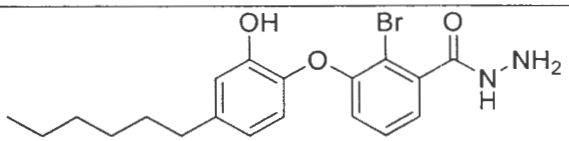
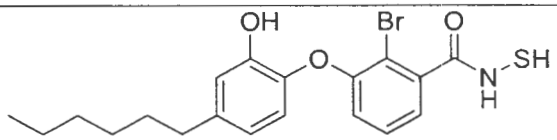
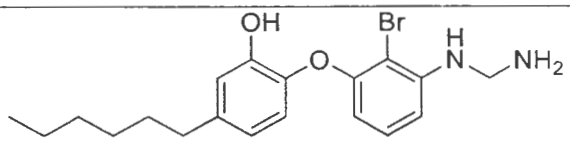
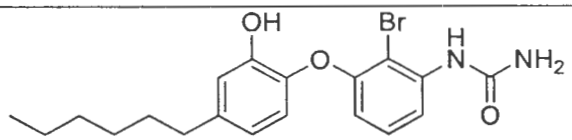
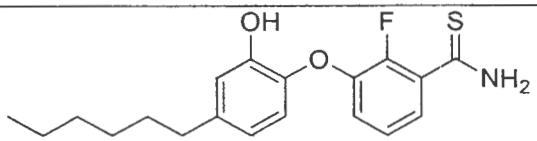
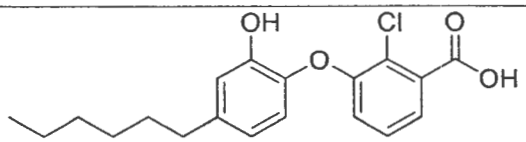
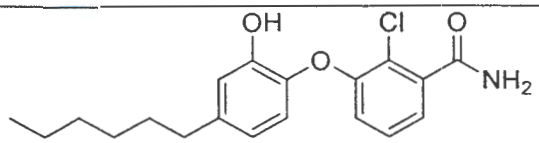
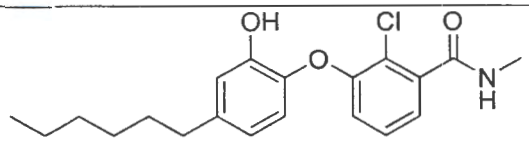
[a] predicted $\log(1/IC_{50})$ derived from best CoMSIA model

Table 3.9 Highly predicted activity diphenyl ether compounds (continued)

Cpd.	Structure	CoMSIA[a]
OD14		8.24
OD17		8.25
OD20		8.23
HD16		7.98
HD18		7.98
HD19		8.01
HD20		7.96
HD21		7.97
HD22		7.97
HD27		8.01

[a] predicted log(1/IC_{50}) derived from best CoMSIA model

Table 3.9 Highly predicted activity diphenyl ether compounds (continued)

Cpd.	Structure	CoMSIA[a]
HD28		7.94
HD29		7.94
HD30		7.97
HD41		7.94
HD55		7.95
HD71		7.95
HD114		7.93
HD115		7.95
HD123		7.95

[a] predicted log(1/IC_{50}) derived from best CoMSIA model

OD04 is the iodo (I) substituent at R_2 position. This compound showed the highest predicted biological activity in OD series (Octyl substituent at R_1 position). In addition, this compound displayed the highest binding energy (-13.25 kcl/mol) in InhA binding pocket (Figure 3.16) that obtained from molecular docking. U-shape of this compound recognized in the InhA binding site. Phenyl A ring that contained I substituent be vertical with Phenyl B ring. The hydroxyl group of phenyl A ring formed strong hydrogen bond network with hydroxyl group of Tyr158 (1.78 Å) and nicotinamide ribose (2.00 Å). Moreover, this phenyl ring bound with nicotinamide ring of NAD^+ cofactor via pi-pi interaction. Hydrophobic interactions of phenyl A ring with Phe149, Met199 and Val203. Octyl chain of R_1 substituent on phenyl A ring located at the substrate binding site of InhA and formed hydrophobic interaction with Phe149, Met155, Pro156, Ala157, Tyr158, Pro193, Ile194, Met199, Val203, Leu207, Cln214, Ile215 and Leu218. The position of iodine atom (Iodo substituent) of R_2 closed to the position of NAD^+ cofactor, Gly96 and Ala198. Hydrogen atom at para position of Phenyl B ring pointed to the position of nitrogen atom of Met98 backbone (distant between two atoms is 2.70 Å). Therefore, hydrogen atom of this position cloud is form weakly hydrogen bond interaction. Side chain of Met161 bound with phenyl B ring via methyl-pi interaction. Hydrophobic interaction of phenyl B ring with Met161, Ala198, Ile202 and Val203 were observed.

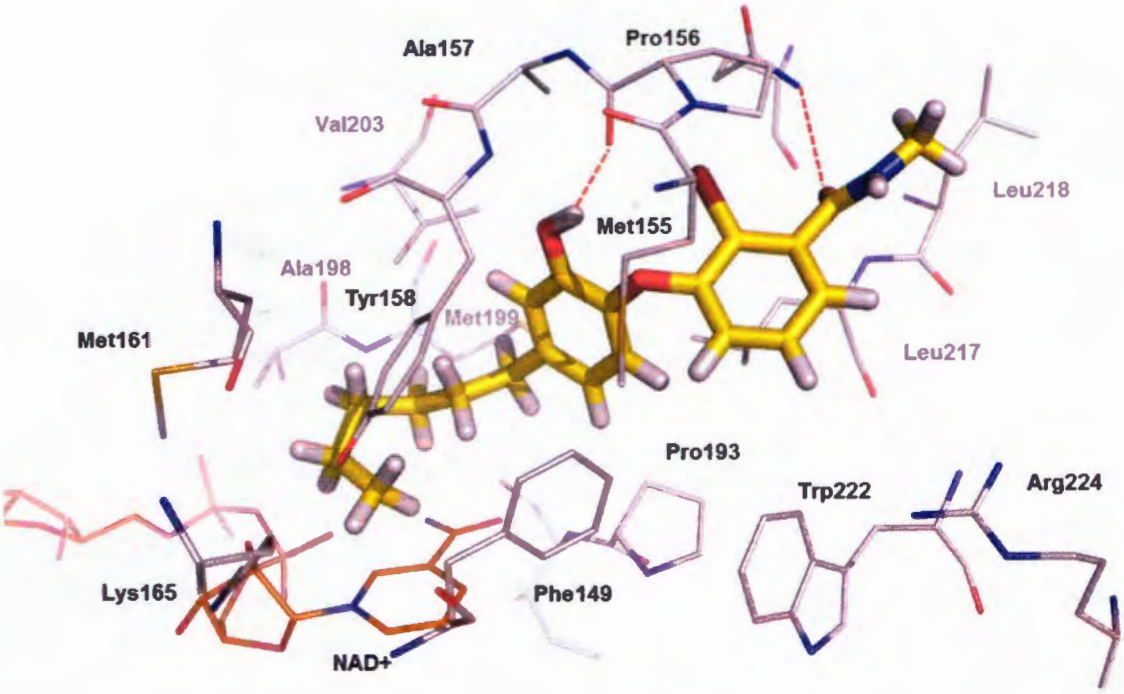


Figure 3.18 Binding mode of HD27 in InhA binding site derived from molecular docking

3.1.1.2 Molecular modeling of benzofuran pyrrolidine pyrazole derivatives

1) 3D-QSAR studies

1.1) CoMFA and CoMSIA models

In this study, CoMFA and CoMSIA models were constructed from IC_{50} and MIC_{90} where prefixed with IC_{50} and MIC_{90}, respectively. IC_{50} and MIC_{90} CoMSIA models were constructed based on various combinations of molecular descriptor fields, in order to develop a highly predictive CoMSIA model (Tables 3.10 and 3.11). An IC_{50} CoMSIA model constructed from the combination of steric (S), electrostatic (E), hydrophobic (H) and hydrogen acceptor (A) fields (Klebe et al., 1994) gave the highest q^2 (0.646), whereas an MIC_{90} CoMSIA model including steric, electrostatic, hydrophobic and hydrogen donor (D) fields (Klebe et al., 1994) showed the highest q^2 (0.639). Therefore, these models were selected for graphical interpretation of IC_{50} and MIC_{90} CoMSIA contour maps. In order to assess the predictive abilities of IC_{50} and MIC_{90} CoMSIA models, IC_{50} and MIC_{90} values of the test set were predicted. Both IC_{50} and MIC_{90} CoMSIA models showed good ability to predict IC_{50} and MIC_{90} values of the test set data as shown in Figure 3.19. In case of

IC_{50} and MIC_{90} CoMFA models, they had poor predictive ability with q^2 values of 0.464 and 0.432, respectively. Accordingly, these CoMFA models were not used further in this work.

Table 3.10 Statistical results of IC_{50} CoMFA and CoMSIA models

Models	Statistical parameters						Fraction
	q^2	r^2	s	SEE	N	F	
CoMFA							
S/E	0.464	0.996	0.392	0.035	6	909.618	60.3/39.7
CoMSIA							
S/E	0.084	0.977	0.512	0.081	6	162.845	32.1/67.9
S/H	0.465	0.950	0.383	0.118	5	90.431	29.1/70.9
S/D	0.624	0.923	0.321	0.145	5	57.579	54.3/45.7
S/A	0.146	0.970	0.495	0.093	6	123.724	39.7/60.3
S/E/H	0.260	0.981	0.460	0.074	6	194.704	16.6/44.5/38.9
S/E/D	0.592	0.980	0.342	0.076	6	185.576	21.0/53.7/25.3
S/E/A	0.281	0.975	0.454	0.085	6	149.701	22.5/42.8/34.7
S/E/H/D[a]	**0.646**	**0.990**	**0.318**	**0.055**	**6**	**363.962**	**13.1/35.8/28.5/22.6**
S/E/H/A	0.336	0.983	0.436	0.070	6	222.520	12.3/31.5/29.4/26.8
S/E/H/D/A	0.610	0.991	0.334	0.050	6	437.341	10.0/25.4/22.6/20.7/21.4

[a] Bold values indicate the best CoMSIA model

N optimum number of components; *s* standard error of prediction; *SEE* standard error of estimate; *F* F-test value; *S* steric field; *E* electrostatic field; *H* hydrophobic field; *D* hydrogen donor field; *A* hydrogen acceptor field

Table 3.11 Statistical results of MIC_{90} CoMFA and CoMSIA models.

Models	Statistical parameters						Fraction
	q^2	r^2	s	SEE	N	F	
CoMFA							
S/E	0.432	0.853	0.442	0.225	2	78.451	53.2/46.8
CoMSIA							
S/E	0.456	0.949	0.469	0.143	6	71.455	25.1/74.9
S/H	0.459	0.780	0.432	0.275	2	47.970	34.4/65.6
S/D	0.261	0.732	0.514	0.310	3	23.642	52.7/47.3
S/A	0.602	0.978	0.401	0.093	6	174.060	46.3/53.7
S/E/H	0.477	0.961	0.460	0.126	6	93.558	13.8/52.8/33.4
S/E/D	0.210	0.912	0.553	0.184	5	49.990	17.7/64.4/18.0
S/E/A	0.550	0.955	0.426	0.134	6	82.091	19.9/48.1/32.0
S/E/H/D	0.415	0.938	0.476	0.155	5	72.712	10.9/45.8/29.3/13.9
S/E/H/A[a]	**0.639**	**0.973**	**0.382**	**0.105**	**6**	**136.014**	**12.5/35.6/42.2/27.7**
S/E/H/D/A	0.494	0.961	0.442	0.123	5	118.951	9.3/33.4/22.8/10.4/24.2

[a] Bold values indicate the best CoMSIA model

N optimum number of components; *s* standard error of prediction; *SEE* standard error of estimate; *F* F-test value; *S* steric field; *E* electrostatic field; *H* hydrophobic field; *D* hydrogen donor field; *A* hydrogen acceptor field

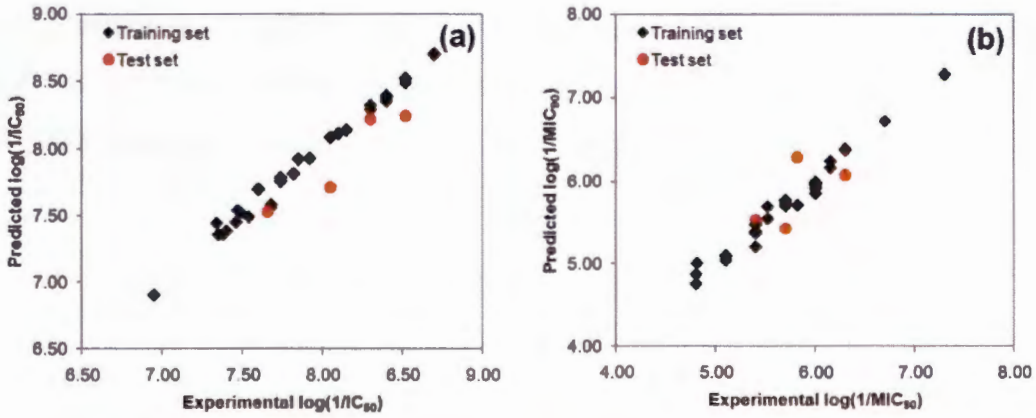


Figure 3.19 Plot of experimental and predicted activities of the training and test data sets derived from IC_{50} (a) and MIC_{90} (b) CoMSIA models

1.2) CoMSIA contour maps

To reveal the importance of molecular descriptor fields in both IC_{50} and MIC_{90} values of InhA inhibitors, CoMSIA contour maps were established. Compound **22** (Table 2.7 in chapter 2) presented the best MIC value. Graphical interpretation of its IC_{50} and MIC_{90} CoMSIA contour maps was done. Interpretation of its IC_{50} and MIC_{90} CoMSIA contour maps revealed structural requirements in terms of steric, electrostatic, hydrophobic and hydrogen donor and acceptor fields for IC_{50} and MIC_{90} values of InhA inhibitors.

1.2.1) Steric requirements for IC_{50} and MIC_{90} values

Figure 3.20 shows the CoMSIA steric contour maps obtained from selected IC_{50} and MIC_{90} CoMSIA models. These contours highlight the steric requirements for IC_{50} and MIC_{99} values of benzofuran pyrrolidine pyrazole derivatives. Both IC_{50} and MIC_{90} CoMSIA models show a green contour at the $\mathbf{R_3}$ substituent. These results indicated that a bulky $\mathbf{R_3}$ substituent is favourable for both IC_{50} and MIC_{90} values. Accordingly, an ethyl group is more preferred for the steric requirement of the $\mathbf{R_3}$ substituent than a methyl group. This is consistent with the MD simulations since an ethyl group can form more interactions with InhA. At the $\mathbf{R_2}$ position, IC_{50} and MIC_{90} CoMSIA models present a large yellow contour. However, IC_{50} CoMSIA model shows a favorable green steric contour at the terminal of the $\mathbf{R_2}$ substituent (Figure 3.20a). Based on MD simulations results, the $\mathbf{R_2}$ substituent had weak interaction with the InhA pocket leading to less influence on the IC_{50} value. Therefore, the steric requirement of $\mathbf{R_2}$ substituent should be based on the MIC_{90} CoMSIA steric contour that presented only a yellow contour near this substituent (Figure 3.20b).

2.3) Binding mode of compound 28

The binding mode of compound **28** complexed with InhA obtained from MD simulations is shown in Figure 3.25. Residues located near each substituent and the core structure are listed in Figure 3.26. A hydrogen atom (the $\mathbf{R_1}$ substituent) is near the carbonyl backbone of Met103. 2-pyridinyl methyl (the $\mathbf{R_2}$ substituent) protrudes from the InhA pocket and interacts with the solvent (Figure 3.25). The ethyl moiety (the $\mathbf{R_3}$ substituent) is located near backbones of Gly96, Phe97 and pyrophosphate and ribose groups of NAD^-. The ethyl group (the $\mathbf{R_4}$ substituent) was located in the hydrophobic side chains of Phe149, Tyr158, Met199 and nicotinamide of NAD^-. With regard to the core structure, the pyrazole ring in the core structure was sandwiched between two hydrophobic side chains of Met161 and Ala198. CO and NH of pyrazole amide formed hydrogen bonds with the backbones of Met98 and Ala198, respectively. The benzofuran core was buried in the hydrophobic side chains of Ile215, Ala157, Ile202 and Ala201, and was sandwiched between the hydrophobic side chains of Leu207 and Met103. The carbonyl of benzofuran core formed a hydrogen bond with the NH backbone of Ala201. NH of pyrrolidine amide formed a hydrogen bond with the CO backbone of Leu197.

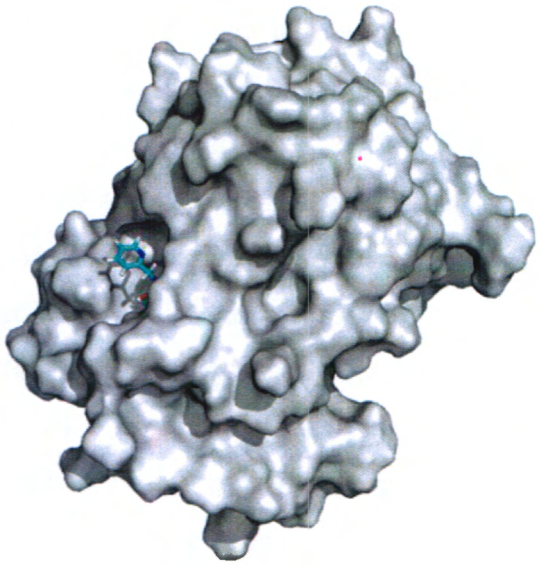


Figure 3.25 Compound 28 (cyan) in its complex with whole InhA (grey) obtained from MD simulations

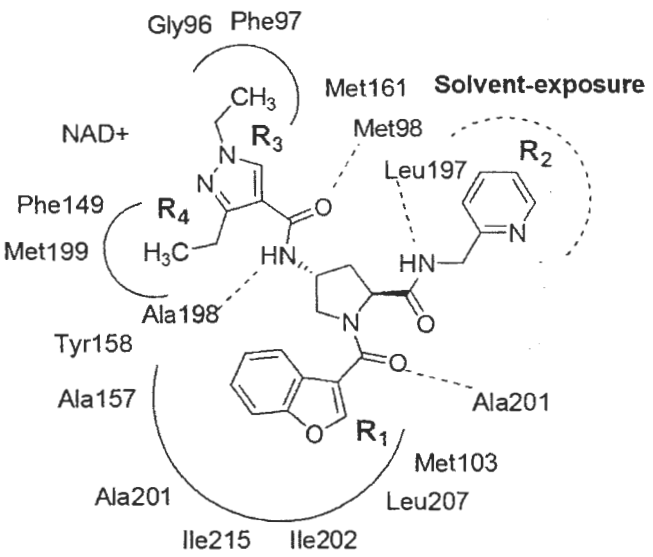


Figure 3.26 List of residues surrounding within 4 Å from compound 28

2.4) Interaction energy

Free-energy decomposition calculations were used to investigate the interaction energies between compound **28** and each residue in the InhA pocket. Figure 3.27 shows these interaction energies obtained from free-energy decomposition calculations. The lowest interaction energy (-7.42 kcal/mol) was observed for Met103, indicating that this residue had the largest contribution to binding of compound **28** in the InhA pocket. As previously mentioned, Met103 and Leu207 were sandwiched in the benzofuran core. Another remarkable interaction energy (-7.06 kcal/mol) was found for NAD^+. This was responsible for van der Waal and electrostatic interactions with the $\mathbf{R_3}$ and $\mathbf{R_4}$ substituents of compound **28** (Figure 3.27). Ala198 showed an interaction energy (-6.16 kcal/mol), comparable with those of Met103 and NAD^+. This residue formed hydrogen bonds with the NH of pyrazole amide and sandwiched the pyrazole ring (Figure 3.26). Met98, Leu197 and Ala201 formed other hydrogen bonds with the core structure with interaction energies of -2.94, -3.27 and -5.33 kcal/mol, respectively. Based on interaction energy profile of compound **28**, the core structure formed more attractive interactive energies with surrounding residues than **R** substituents (Figure 3.27). This result indicates that the core structure is the key fragment for binding of this compound in the InhA pocket.

significantly change IC_{50} values, but rather produced a tenfold increase in MIC_{90} values (compounds **22** and **23**, Table 2.7 in chapter 2). Accordingly, the $\mathbf{R_2}$ substituent is a key group that can be used to adjust the MIC_{90} value without negative contribution to the IC_{50} value. Based on the results obtained from our MD simulations and CoMSIA studies, the structural concept to correctly balance IC_{50} and MIC_{90} values of benzofuran pyrrolidine pyrazole derivatives is summarized in Figure 3.32. New compounds designed based on this concept should show better IC_{50} and MIC_{90} values.

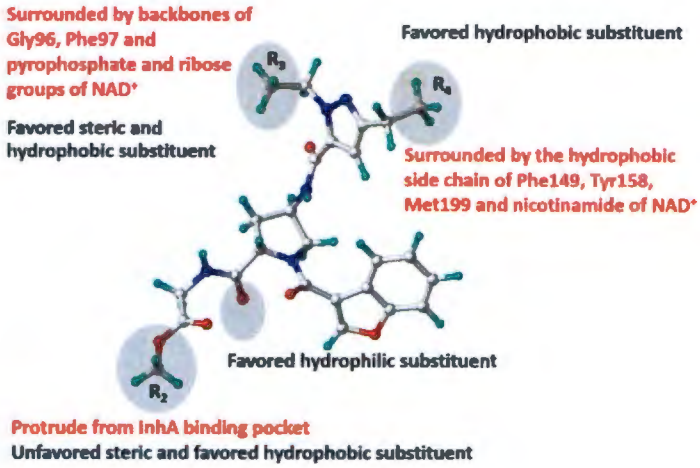


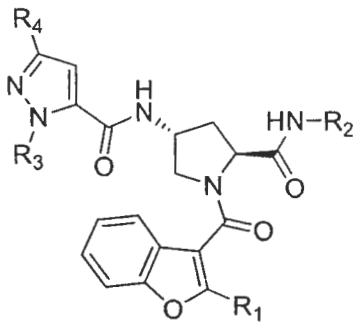
Figure 3.32 Structural concept for good IC_{50} and MIC_{90} correlation summarized from MD simulations and CoMSIA results. Red and black letters indicate the results obtained from MD simulations and CoMSIA results, respectively

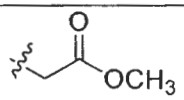
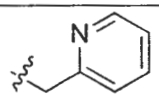
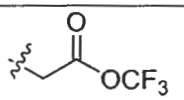
3) Design new and more potent benzofuran pyrrolidine pyrazole derivatives as direct InhA inhibitors

Based on the integration of 3D-QSAR CoMSIA contour maps, the structural basis to improve inhibitory activity against InhA and *M. tuberculosis* whole cell was proposed. 123 designed compounds of benzofuran pyrrolidine pyrazole derivatives (BD series, **BD1-BD123**) were summarized in Table 3.13. No contour located at R_1 position. High predicted biological activities both InhA inhibition and *M. tuberculosis* whole cell inhibition were required. Therefore, designed compounds were started from compound **22** as template structure. Hydrogen atom was kept at R_1 position of new designed compounds due to no contour maps available at R_1 position.

At R_2 position, unfavorable steric and favorable hydrophobic substituents were designed (**BD1-BD81**) to examine the optimal R_2 substituents. At R_3 position, favorable steric and hydrophobic substituents were modified. At R_4 position, favorable hydrophobic substituents were designed. The obtained results show that 17 compounds of new designed compounds shown higher predicted inhibitory against *M. tuberculosis* whole cell and 2 compounds (**BD101** and **BD102**) from of high predicted biological activity against *M. tuberculosis* whole cell were shown high predicted activity in InhA inhibition.

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives

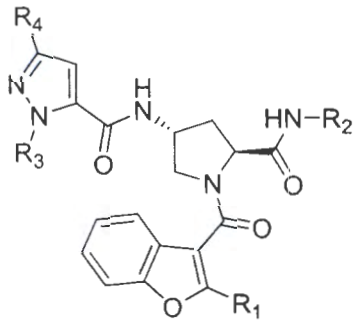


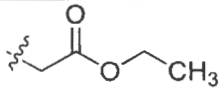
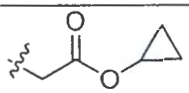
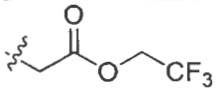
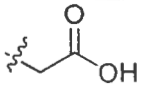
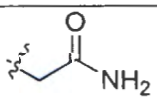
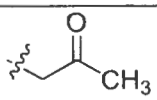
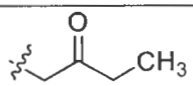
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC50)	log(1/MIC90)
22		Et	Et	8.52	7.30
28		Et	Et	8.70	6.15
BD1		Et	Et	8.23	6.87

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

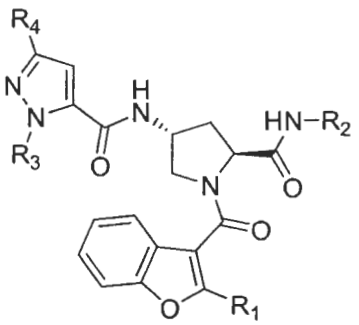


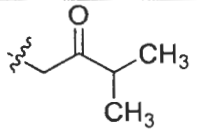
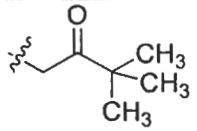
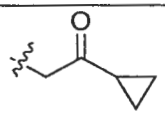
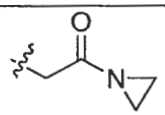
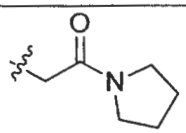
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC50)	log(1/MIC90)
BD2		Et	Et	8.50	7.06
BD3		Et	Et	8.29	7.06
BD4		Et	Et	8.24	6.97
BD5		Et	Et	8.09	7.06
BD6		Et	Et	8.13	7.21
BD7		Et	Et	8.05	6.78
BD8		Et	Et	8.04	6.66

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)



Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC_{50})	log(1/MIC_{90})
BD9		Et	Et	7.91	6.60
BD10		Et	Et	7.79	6.51
BD11		Et	Et	8.07	6.70
BD12		Et	Et	8.41	7.16
BD13		Et	Et	8.24	6.98

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

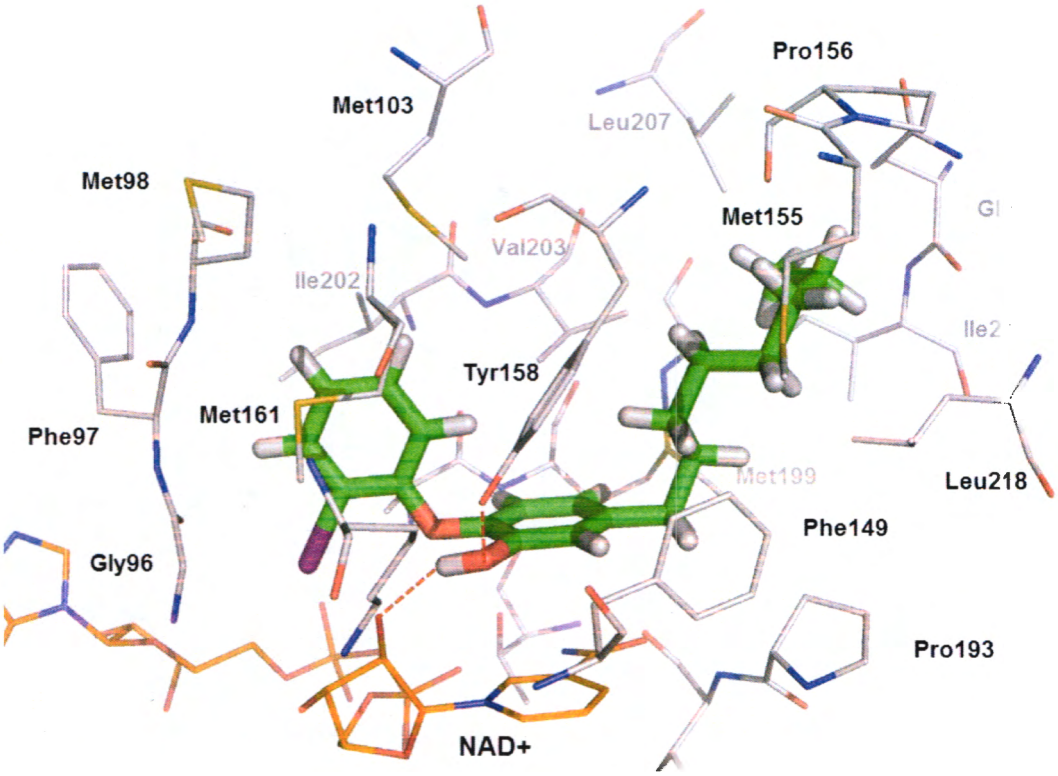


Figure 3.16 Binding mode of OD04 in InhA binding site derived from molecular docking

The binding modes of **HD19** and **HD27**, the highest predicted activity in HD series (Table 3.8 and Table 3.9) were analyzed. Two of these compounds bound with InhA binding pocket with different binding mode as compared to the X-ray binding mode of diphenyl ether compounds (Figure 3.17 and Figure 3.18). Three hydrogen bond interactions between HD19 with Pro156, Gln214 and Arg225 were found. First hydrogen bond interaction was observed between hydroxyl (OH) of diphenyl ether core structure with an oxygen atom of carbonyl group of Pro156. Second hydrogen bond interaction was observed between an oxygen carbonyl (C=O) of primary amide at R_3 substituent with NH sidechain of Arg225. The latest hydrogen bond interactions of **HD19**, the Br substituent at R_2 position cloud formed hydrogen bond interaction with NH side chain of Gln214. Hexyl R_1 substituent formed hydrophobic interaction with Phe149, Tyr158, Ala198, Met199 and Val203 side chain.

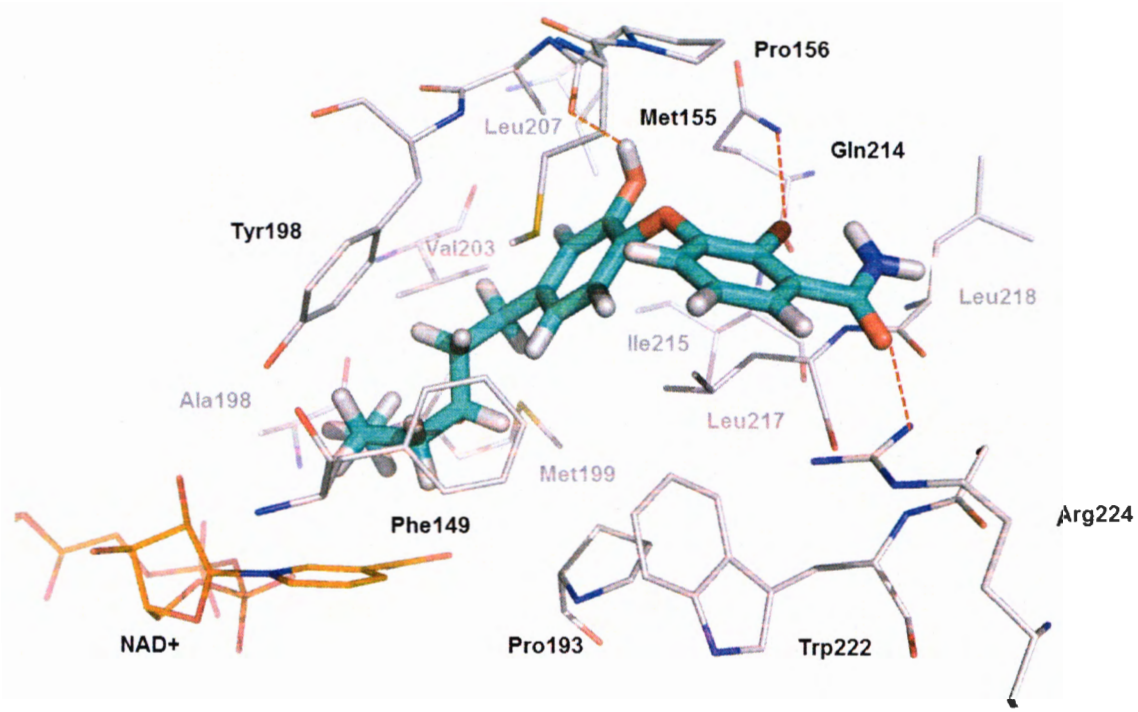


Figure 3.17 Binding mode of HD19 in InhA binding site derived from molecular docking

The binding mode derived from molecular docking of HD27 was shown in Figure 3.18. Two hydrogen bond interactions were observed between HD27 and amino acids in InhA binding pocket. Hydroxyl group at ring A of diphenyl ether core structure formed hydrogen bond interaction with an oxygen atom of carbonyl Pro156. Carbonyl of secondary amide R_3 substituent interacted with NH side chain of Gln214 via hydrogen bond interaction. Hexyl (R_1) and Br (R_2) substituents formed hydrophobic interactions with amino acid in InhA binding site.

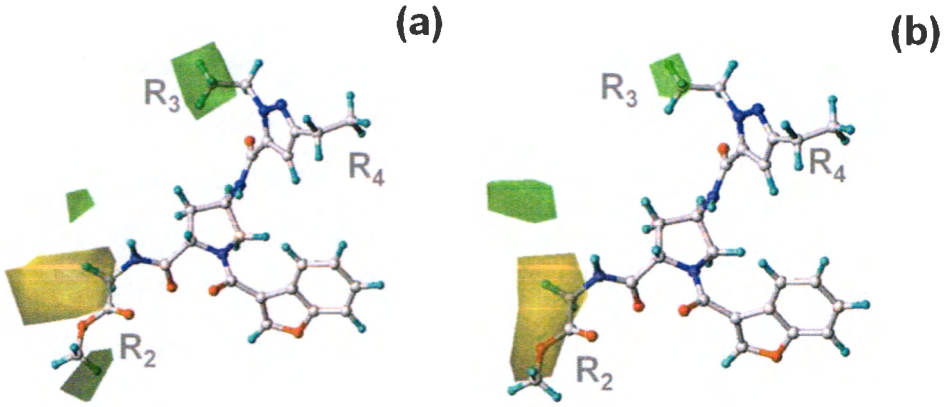


Figure 3.20 Steric contour maps of IC_{50} (a) and MIC_{90} (b) CoMSIA models in combination with compound 22

1.2.2) Electrostatic requirements for IC_{50} and MIC_{90} values

Electrostatic requirements for IC_{50} and MIC_{90} values of benzofuran pyrrolidine pyrazole derivatives are visualized in Figure 3.21. Both IC_{50} and MIC_{90} CoMSIA contours show only an electrostatic requirement at the $\mathbf{R_2}$ substituent. The IC_{50} CoMSIA shows a red contour at the ester moiety of $\mathbf{R_2}$ substituent, whereas MIC_{90} CoMSIA presents a blue contour at this position. These results show different electrostatic requirements for IC_{50} and MIC_{90} values of benzofuran pyrrolidin pyrazole derivatives. However, the $\mathbf{R_2}$ substituent has weak influence on the IC_{50} value. Therefore, the electrostatic requirement of $\mathbf{R_2}$ substituent for MIC_{90} values should take more priority.

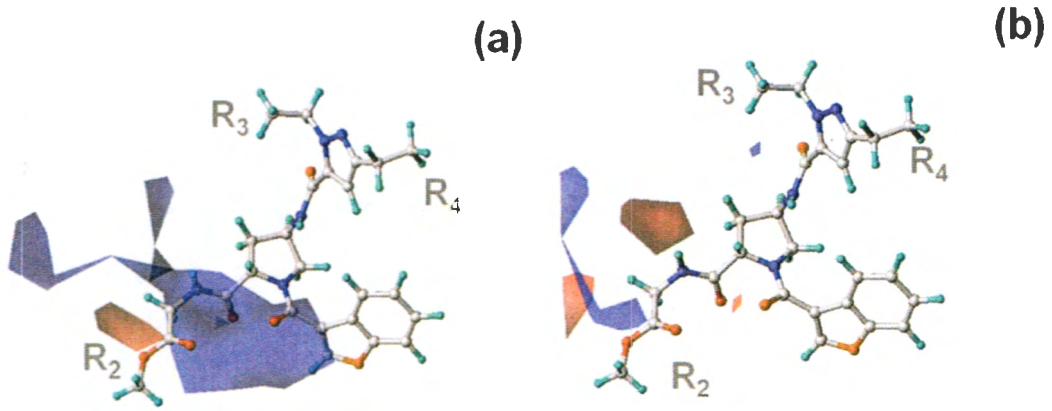


Figure 3.21 Electrostatic contour maps of IC_{50} (a) and MIC_{90} (b) CoMSIA models in combination with compound 22

1.2.3) Hydrophobic requirements for IC_{50} and MIC_{90} values

Both IC_{50} and MIC_{90} CoMSIA contours show a purple contour at the $\mathbf{R_3}$ substituent of compound **22** (Figure 3.22). This shows that the hydrophobic requirements of the $\mathbf{R_3}$ substituent for both IC_{50} and MIC values were similar. The $\mathbf{R_3}$ substituent was either a methyl or ethyl group. As seen in Figure 3.22, the terminal of ethyl group was buried in a purple $\mathbf{R_3}$ contour. Therefore, the ethyl group was preferable for the hydrophobic requirement of the substituent. IC_{50} and MIC_{90} values of compound **2** with the methyl group at the $\mathbf{R_3}$ substituent were weaker than those of compound **22** containing an ethyl group. At the $\mathbf{R_2}$ substituent, both IC_{50} and MIC_{90} CoMSIA contours display a purple contour at this position (Figure 3.22). Therefore, the presence of a hydrophobic substituent at this purple region should enhance both IC_{50} and MIC_{90} values. The grey contour located at the carbonyl moiety of the $\mathbf{R_2}$ substituent in both IC_{50} and MIC_{90} CoMSIA contours indicated that this moiety is important for both IC_{50} and MIC_{90} values. Another important hydrophobic contour is located at the $\mathbf{R_4}$ substituent. The MIC_{90} CoMSIA shows a purple region near the $\mathbf{R_4}$ substituent (Figure 3.22b), but this contour disappeared in the IC_{50} CoMSIA contour (Figure 3.22a). Therefore, a hydrophobic moiety could be presented at purple region to enhance the MIC_{90} value without a negative contribution to the IC_{50} value.

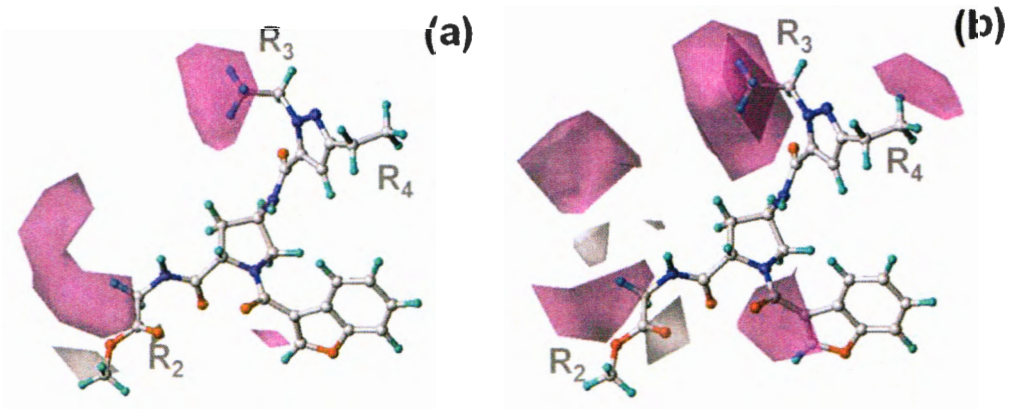


Figure 3.22 Hydrophobic contour maps of IC_{50} (a) and MIC_{90} (b) CoMSIA models in combination with compound 22

1.2.4) Hydrogen donor and acceptor requirements for IC_{50} and MIC_{90} values

The hydrogen donor field was included in the selected IC_{50} CoMSIA model, but this molecular descriptor was instead changed to a hydrogen acceptor field in the selected MIC_{90} CoMSIA model. The IC_{50} CoMSIA model did not show any hydrogen donor contour near any **R** substituents. However, this model showed a favourable hydrogen donor contour at the amide moiety of the core structure. The amide moiety appears to impact the IC_{50} value. Consistent with the MD simulations results, this moiety can form hydrogen bonds with Leu197. The MIC_{90} CoMSIA model shows a favourable hydrogen acceptor contour at the carbonyl moiety of $\mathbf{R_2}$ substituent, indicating that this moiety is essential to a good MIC_{90} value.

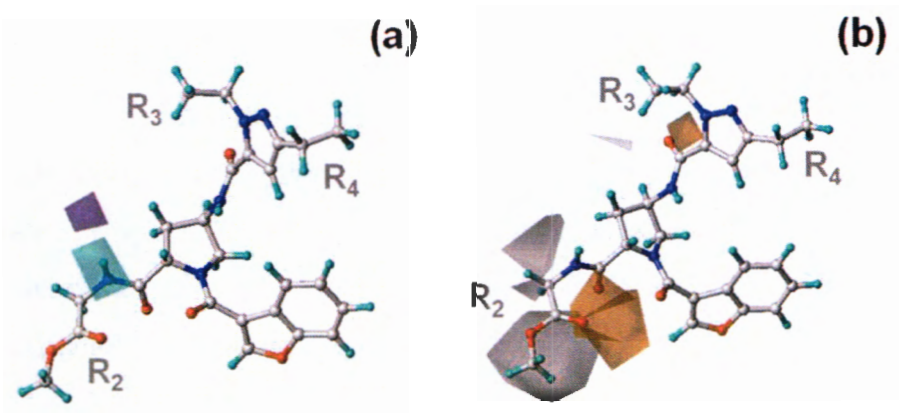


Figure 3.23 Hydrogen donor contour of IC_{50} CoMSIA model (a) and hydrogen acceptor contour MIC_{90} CoMSIA model (b) in combination with compound 22

2) MD simulations

2.1 Stability of the complex models

To reveal the structural stability of simulation system, the RMSD values for the position of all solute species were separately analyzed. The RMSD plots for the four simulation systems over 10 ns are shown in Figure 3.24. Convergent RMSD plots indicated that the equilibrium state was reached for each system during this simulation period. As shown, the RMSDs for compounds **2**, **22**, **23** and **28** in InhA converged after approximately 2 ns.

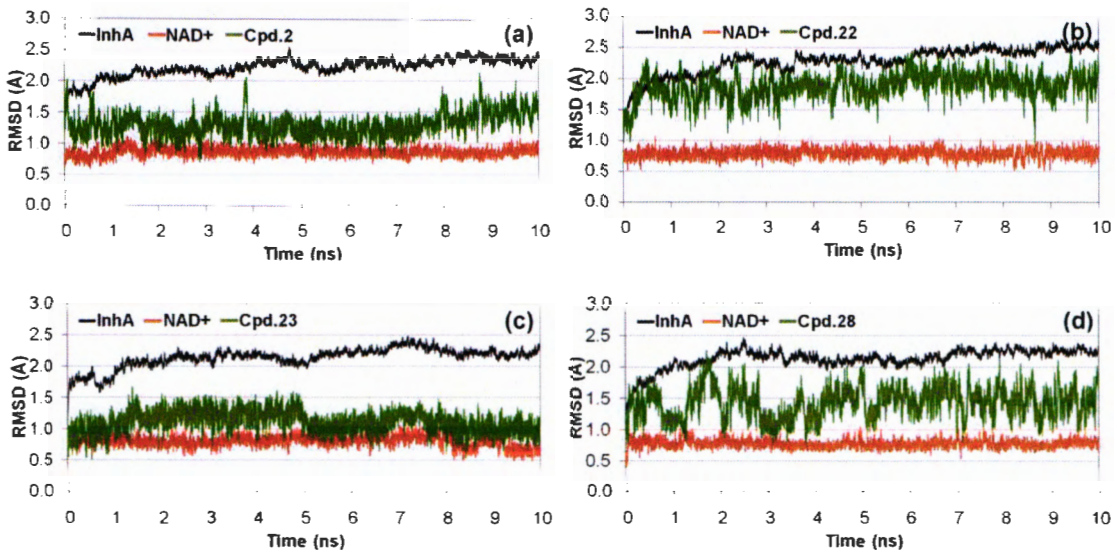


Figure 3.24 RMSD plots of compounds 2 (a), 22 (b), 23 (c), and 28 (d) complexed with InhA.

2.2) Reliability of the calculation methods

MD simulations were employed to model the binding modes of compounds **2**, **22**, **23** and **28** in the InhA pocket. The experimental binding free energy (ΔG_{exp}) lying within the experimental error of the calculated values (ΔG_{bind}) considered as the correlation between the experimental binding free energy and the calculated values was used to indicate the reliability of the modelled binding modes of these compounds. ΔG_{bind} values of compounds **2**, **22**, **23** and **28** were close to their ΔG_{exp} values (Table 3.12). Therefore, we concluded that MD simulations reliably modelled binding modes of compounds **2**, **22**, **23** and **28** in the InhA pocket.

Table 3.12 ΔG_{bind} and ΔG_{exp} of compounds 2, 22, 23 and 28 in InhA (kcal/mol)

Cpd.	ΔH	-TΔS	ΔG_{bind}	ΔG_{exp}
2	-46.91±5.08	-31.03±6.06	-15.88±5.14	-15.52
22	-49.69±3.87	-33.15±6.41	-16.54±4.80	-15.82
23	-49.61±3.71	-32.79±5.57	-16.82±4.79	-15.65
28	-49.26±4.45	-32.52±6.58	-16.74±5.34	-16.07

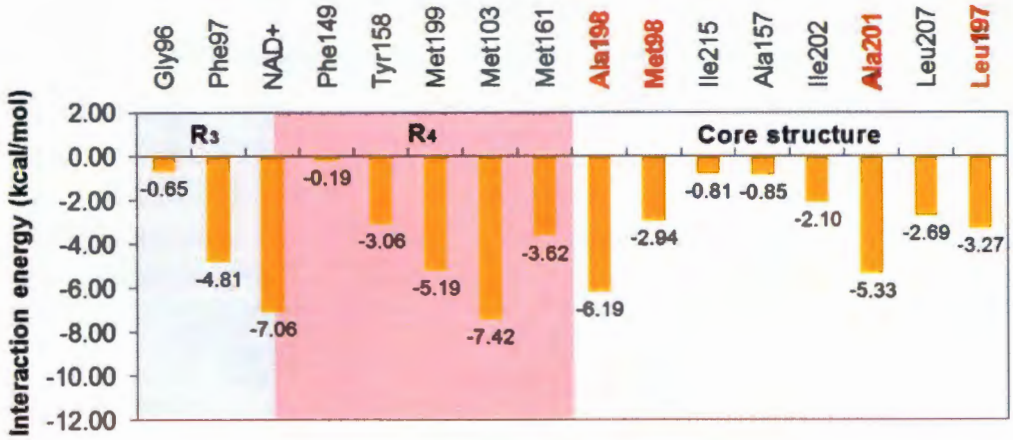


Figure 3.27 Interaction energy profile of compound 28 and surrounding residues within 4 Å

2.5) The effect of the R_2 substituent on IC_{50} and MIC_{90} values

As compared with the positions of other **R** substituents, the $\mathbf{R_2}$ position had the most varied substituents (Table 2.7 in chapter 2). Compound **28** exposing the 2-pyridylmethyl at the $\mathbf{R_2}$ position showed the best activity for InhA inhibition with an IC_{50} of 0.002 μM. When the $\mathbf{R_2}$ substituent of this compound was replaced by CH_2COOMe (compound **22**), the IC_{50} value was slightly changed to 0.003 μM. In contrast, the MIC_{90} value against whole *M. tuberculosis* cell was greatly changed from 0.7 μM to 0.05 μM (Table 2.7 in chapter 2). To reveal the effect of the $\mathbf{R_2}$ substituent on the IC_{50} value, the binding modes of compounds **28** and **22** were compared (Figure 3.28). The binding modes of these compounds in the InhA pocket were similar, and the $\mathbf{R_2}$ substituents occupied in the same positions. Moreover, the interaction energy profiles of compounds **28** and **22** with residues in InhA pocket were similar (Figure 3.29). As discussed above, the $\mathbf{R_2}$ substituent of compound **28** protruded from the InhA pocket leading to weak interaction of this substituent with the pocket. Therefore, the IC_{50} value against InhA was not significantly changed when the $\mathbf{R_2}$ substituent was varied. When the $\mathbf{R_2}$ substituent was replaced by a hydrogen atom (compound **23**), the binding mode and interaction energy profile of this compound were similar to those of compounds **22** and **28** (Figures 3.28 and Figure 3.29). With regard to IC_{50} values, compound **23** showed a comparable IC_{50} value with those of compounds **22** and **28**. However, the MIC_{90} value of this compound (0.5 μM) was

largely increased over that that of compound **22** (0.05 μM). These results indicate that the $\mathbf{R_2}$ substituent had a small effect on the IC_{50} value against InhA due to its weak interaction with the InhA pocket. Alternatively, this substituent is crucial to controlling the MIC_{90} against intact *M. tuberculosis* cells.

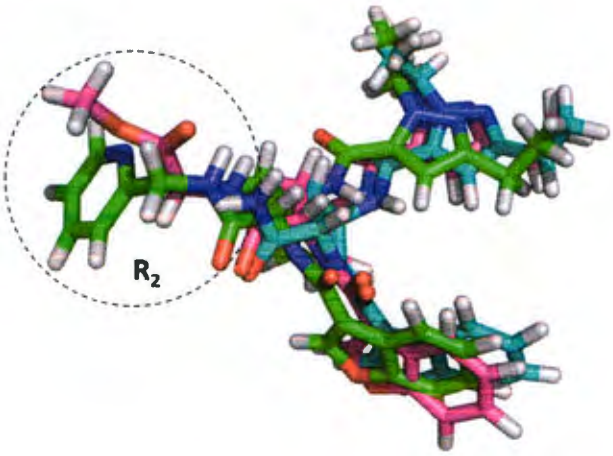


Figure 3.28 Superimposition of binding modes of compounds 22 (pink), 23 (cyan) and 28 (green)

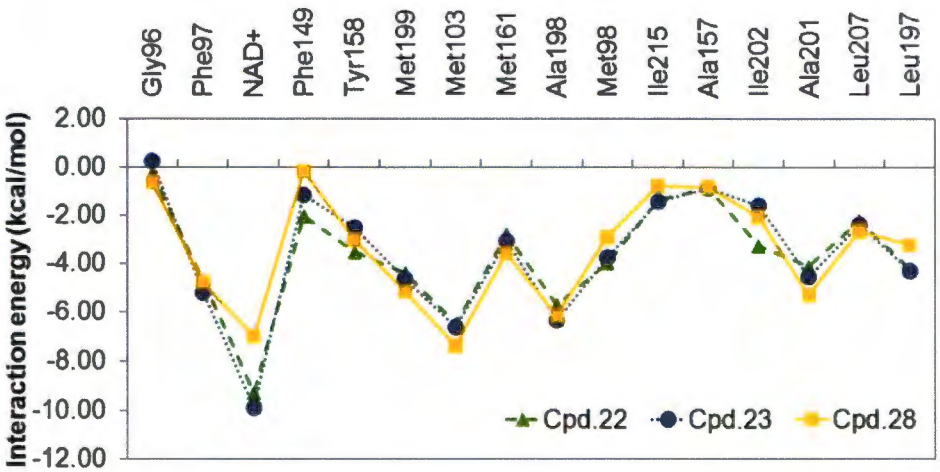


Figure 3.29 Comparison of the interaction energy profiles of compounds 22 (green), 23 (blue) and 28 (yellow) with surrounding pocket within 4 Å

2.6) The effect of the R_3 substituent on IC_{50} and MIC_{90} values

The $\mathbf{R_3}$ substituent of compounds in the data set was varied as ethyl (Et) or methyl (Me) groups (Table 2.7 in chapter 2). Compounds **2** and **22** with structural differences at the $\mathbf{R_3}$ substituent were selected to show the effect of the $\mathbf{R_3}$ substituent on IC_{50} and MIC_{90} values. IC_{50} values of these compounds (0.005 and 0.003 μM, respectively) were not significant different, but their MIC_{90} values were tenfold different (0.5 and 0.05 μM, respectively). Figure 3.30 shows the binding modes of compounds **2** and **22** in InhA obtained from MD simulations. The $\mathbf{R_3}$ substituents of these compounds were located in the same position and surrounded by backbones of Gly96, Phe97 as well as pyrophosphate and ribose groups of NAD^+. The ethyl group (The $\mathbf{R_3}$ substituent) of compound **22** is close to Phe97 and pyrophosphate and ribose groups of NAD^+ more than the methyl group of compound **2**. Therefore, interaction energies of compound **22** with Phe97 and NAD^+ had greater attraction than those of compound **2** (Figure 3.31). Moreover, the presence of a methyl group at the $\mathbf{R_3}$ position of compound **2** shifted the position of benzofuran core surrounded by Met103 and Ile202, and disrupted hydrogen bond interaction with Met98. Accordingly, interaction energies of compound **2** with Met98, Met103 and Ile202 showed less attraction than those of compound **22** (Figure 3.31). These results indicate that compound **22** should have a better IC_{50} against InhA compared to compound **2**. However, other than the interaction energies of Met98, Met103, Ile202, Phe97 and NAD^+, compounds **2** and **22** are comparable. The IC_{50} value for InhA inhibition by compound **22** was slightly better than that of compound **2**. However, its MIC_{90} value was tenfold better than that of compound **2**. The results indicated that the ethyl group at the $\mathbf{R_3}$ position is more conducive to favorable IC_{50} and MIC_{90} values than the methyl group.

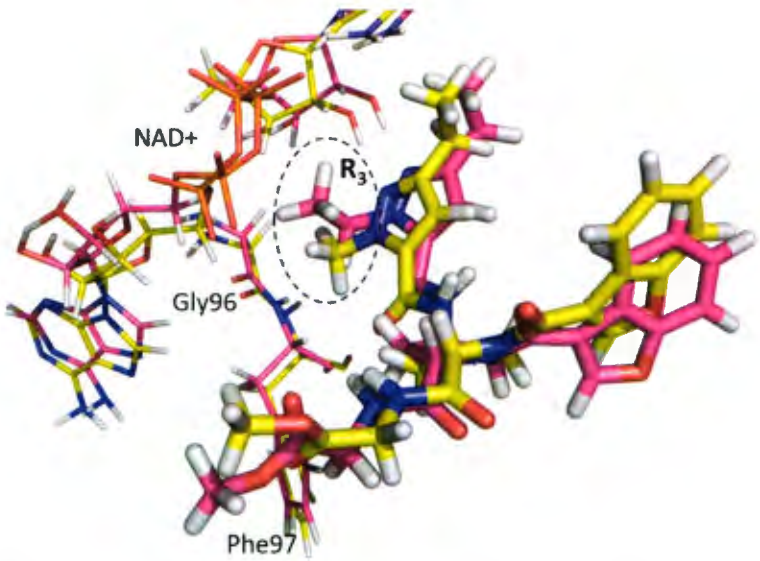


Figure 3.30 Superimposition of binding modes of compounds 2(yellow) and 22 (pink)

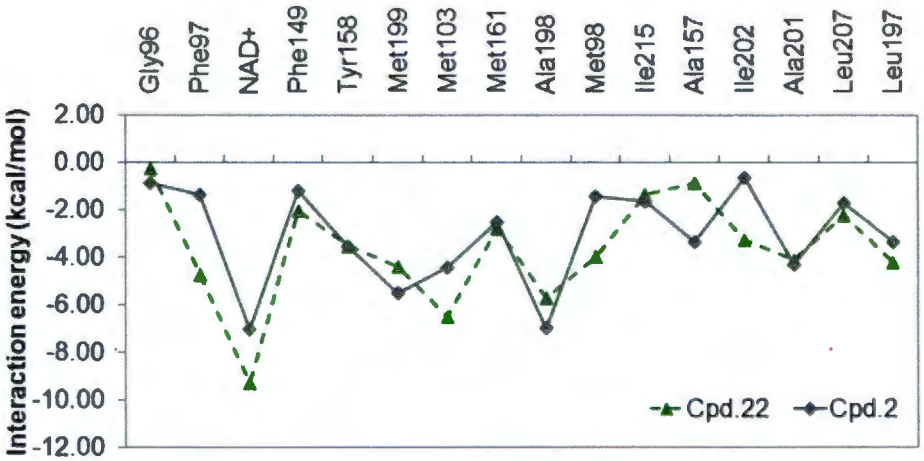
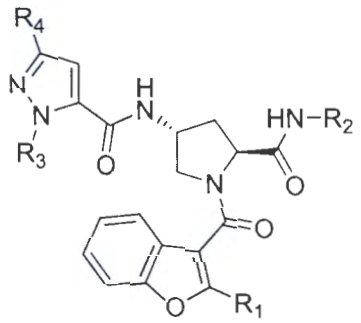


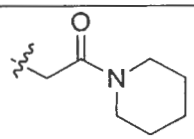
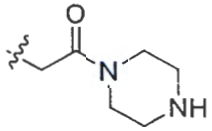
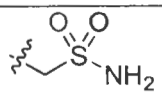
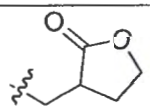
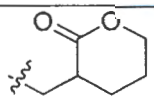
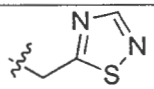
Figure 3.31 Comparison of the interaction energy profiles of compounds 2 (gray) and 22 (green) with surrounding pocket within 4 Å

2.7) The structural concept for good IC_{50} and MIC_{90} correlation

Based on the MD simulations results, the core structure of benzofuran pyrrolidine pyrazole derivatives is of key importance for binding in the InhA pocket. Therefore, this fragment is crucial for favorable IC_{50} values. Among all **R** substituents, the $\mathbf{R_2}$ substituent has the least interaction with the InhA pocket because it protrudes from the pocket. Modifications of the $\mathbf{R_2}$ substituent did not

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

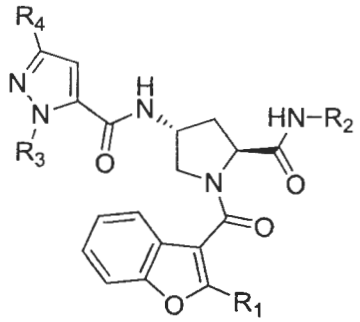


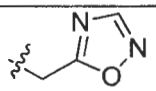
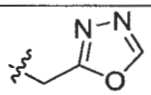
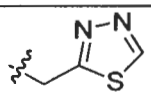
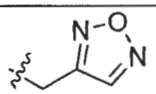
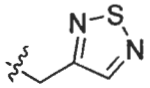
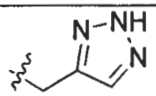
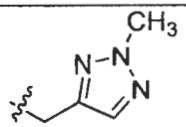
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC50)	log(1/MIC90)
BD14		Et	Et	8.02	6.89
BD15		Et	Et	8.42	7.00
BD16		Et	Et	8.11	5.80
BD17		Et	Et	8.35	7.21
BD18		Et	Et	8.17	7.13
BD19		Et	Et	8.42	6.09

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

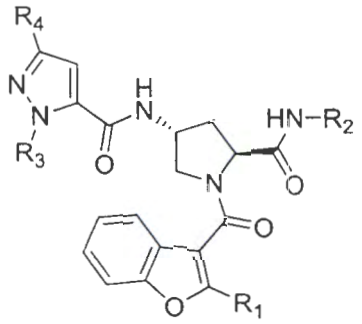


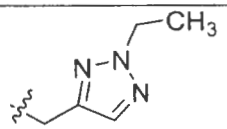
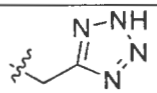
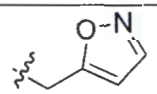
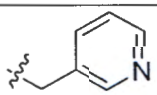
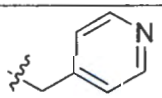
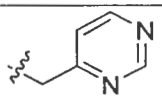
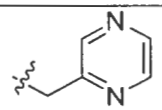
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC50)	log(1/MIC90)
BD20		Et	Et	8.31	5.98
BD21		Et	Et	8.30	5.92
BD22		Et	Et	8.43	5.71
BD23		Et	Et	8.32	5.94
BD24		Et	Et	8.26	5.93
BD25		Et	Et	8.07	6.11
BD26		Et	Et	8.16	6.02

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

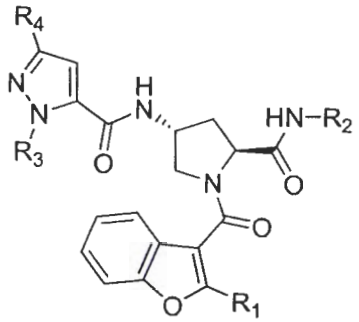


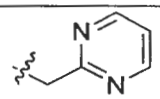
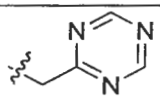
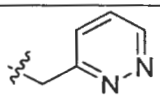
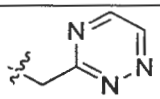
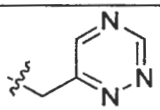
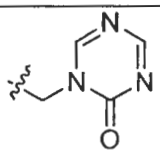
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC50)	log(1/MIC90)
BD27		Et	Et	7.96	5.82
BD28		Et	Et	**8.07**	5.46
BD29		Et	Et	8.23	5.76
BD30		Et	Et	8.04	6.32
BD31		Et	Et	8.12	5.80
BD32		Et	Et	7.75	5.56
BD33		Et	Et	8.18	5.89

[a] Bold letter represents the better predicted biological activity than template compound **22**.

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

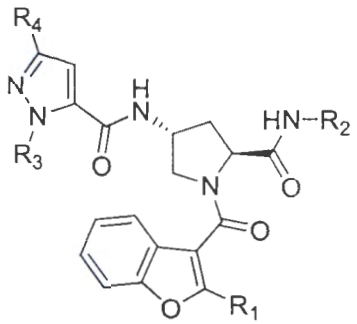


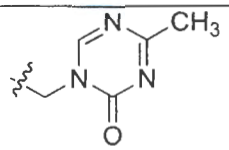
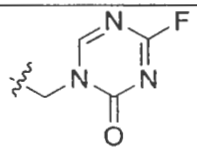
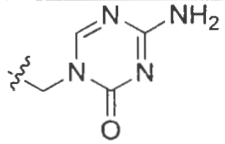
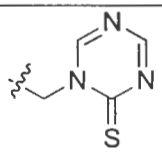
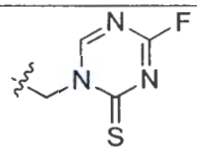
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC50)	log(1/MIC90)
BD34		Et	Et	7.98	5.38
BD35		Et	Et	8.02	5.34
BD36		Et	Et	8.21	5.36
BD37		Et	Et	8.12	5.38
BD38		Et	Et	8.11	6.34
BD39		Et	Et	8.23	5.85

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

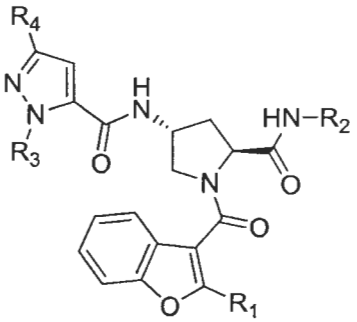


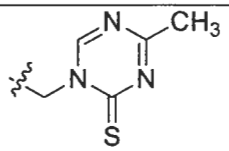
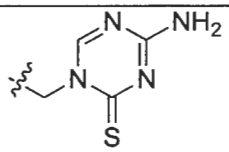
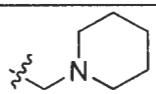
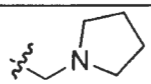
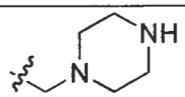
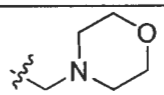
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC50)	log(1/MIC90)
BD40		Et	Et	7.85	5.99
BD41		Et	Et	8.19	5.71
BD42		Et	Et	7.14	5.47
BD43		Et	Et	8.01	5.93
BD44		Et	Et	8.14	5.63

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

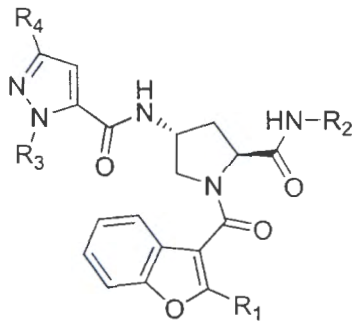


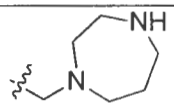
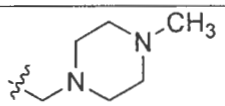
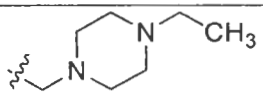
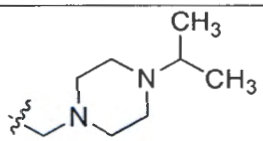
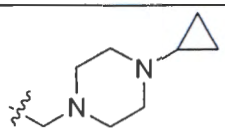
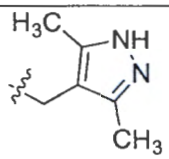
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC50)	log(1/MIC90)
BD45		Et	Et	7.70	5.00
BD46		Et	Et	7.86	6.11
BD47		Et	Et	7.90	5.21
BD48		Et	Et	8.07	5.37
BD49		Et	Et	7.54	5.45
BD50		Et	Et	7.67	5.45

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

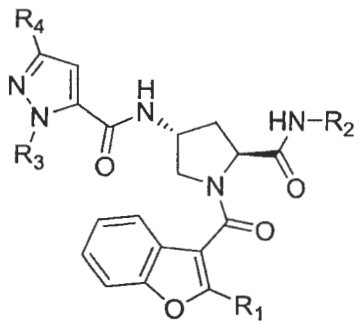


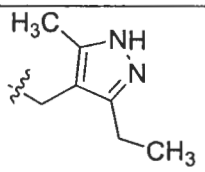
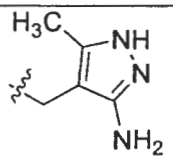
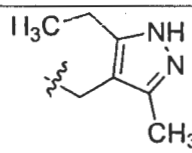
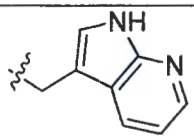
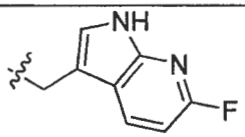
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC50)	log(1/MIC90)
BD51		Et	Et	7.38	5.53
BD52		Et	Et	7.64	5.78
BD53		Et	Et	8.70	5.77
BD54		Et	Et	7.48	5.49
BD55		Et	Et	7.78	6.36
BD56		Et	Et	7.76	5.65

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

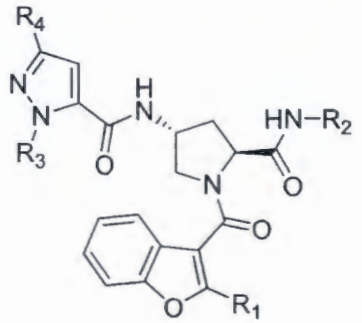


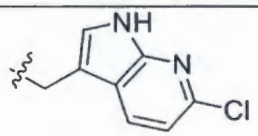
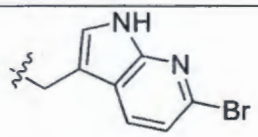
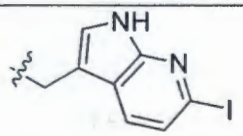
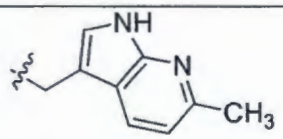
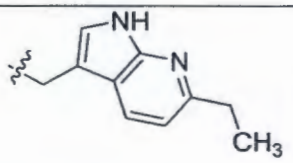
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC50)	log(1/MIC90)
BD57		Et	Et	7.75	5.57
BD58		Et	Et	7.92	6.03
BD59		Et	Et	7.67	5.80
BD60		Et	Et	8.11	5.80
BD61		Et	Et	8.10	5.79

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

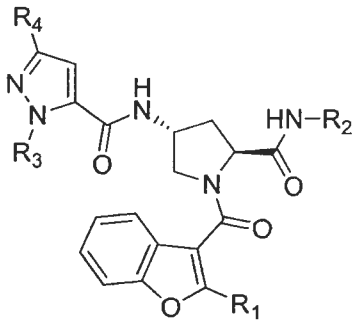


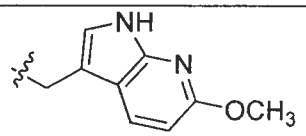
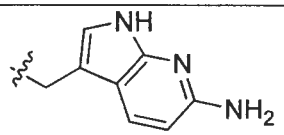
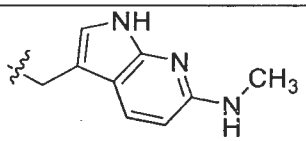
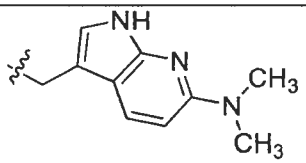
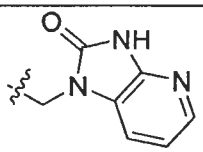
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC50)	log(1/MIC90)
BD62		Et	Et	8.09	6.17
BD63		Et	Et	7.83	5.88
BD64		Et	Et	7.98	6.01
BD65		Et	Et	7.91	6.07
BD66		Et	Et	7.96	6.01

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

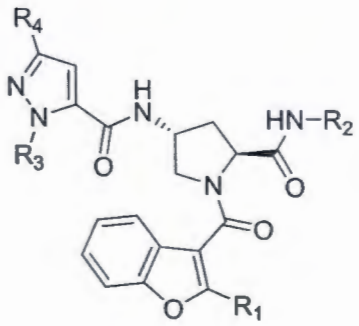


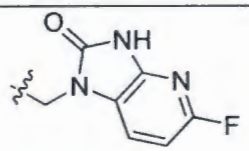
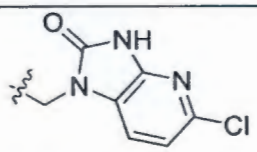
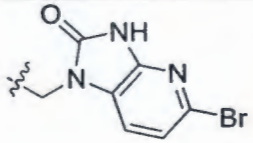
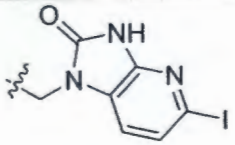
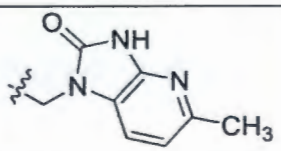
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC50)	log(1/MIC90)
BD67		Et	Et	8.01	5.80
BD68		Et	Et	7.92	5.55
BD69		Et	Et	8.07	5.83
BD70		Et	Et	7.97	5.92
BD71		Et	Et	7.99	5.56

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

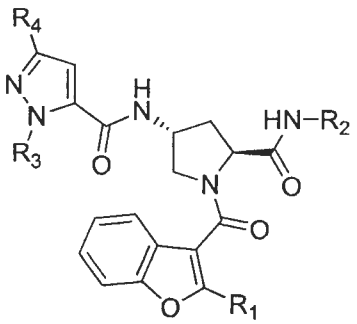


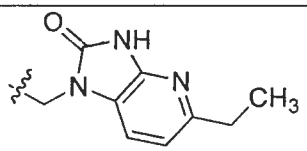
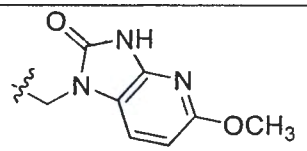
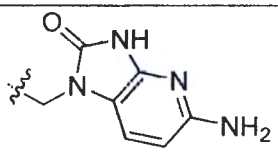
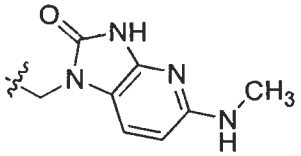
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC50)	log(1/MIC90)
BD72		Et	Et	8.12	5.90
BD73		Et	Et	8.15	5.88
BD74		Et	Et	8.08	5.94
BD75		Et	Et	8.26	5.99
BD76		Et	Et	7.30	4.74

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

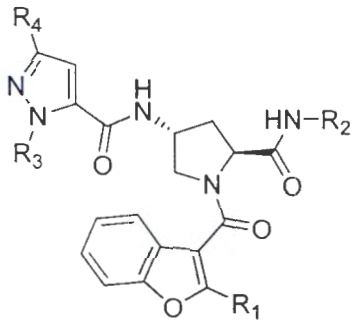


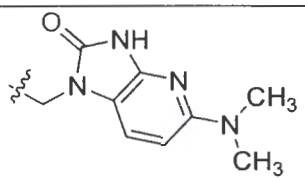
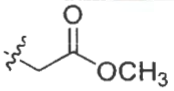
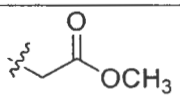
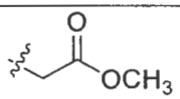
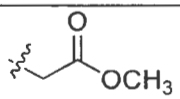
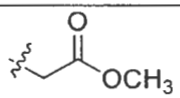
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC50)	log(1/MIC90)
BD77		Et	Et	7.18	5.21
BD78		Et	Et	7.83	5.17
BD79		Et	Et	7.86	7.02
BD80		Et	Et	7.79	5.95

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

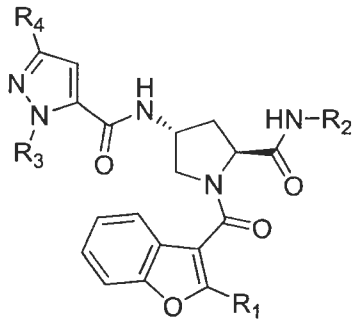


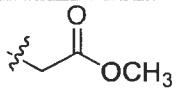
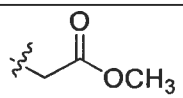
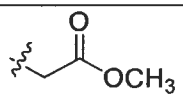
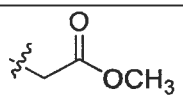
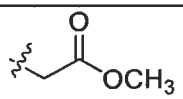
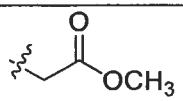
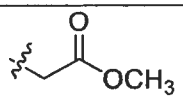
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC50)	log(1/MIC90)
BD81		Et	Et	7.77	5.97
BD82		CH_2CF_3	Et	8.62	7.26
BD83		CF_3	Et	8.33	6.97
BD84		OCH_3	Et	8.37	6.98
BD85		OCF_3	Et	8.46	7.14
BD86		CH_2CF_3	Me	8.55	7.20

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

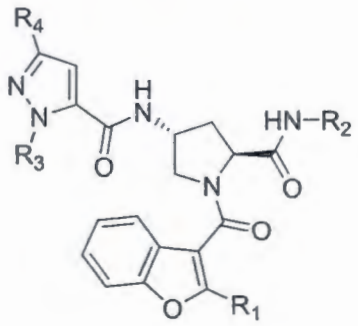


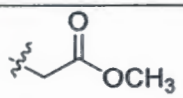
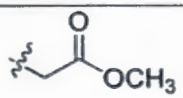
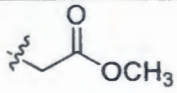
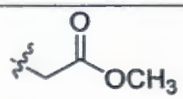
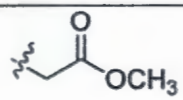
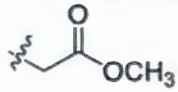
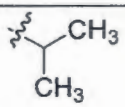
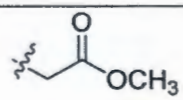
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC_{50})	log(1/MIC_{90})
BD87		CF_3	Me	8.26	6.90
BD88		OCH_3	Me	8.30	6.91
BD89		OCF_3	Me	8.39	7.08
BD90		Et	CH_2CF_3	8.56	7.13
BD91		Et	CF_3	8.44	7.03
BD92		Et	OCH_3	8.45	7.09
BD93		Et	OCF_3	8.51	7.07

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

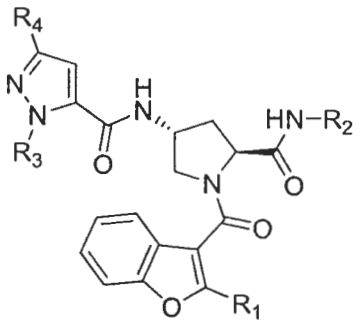


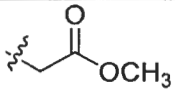
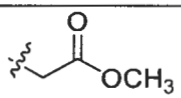
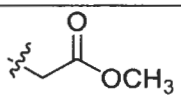
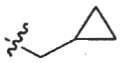
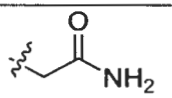
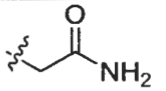
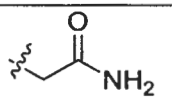
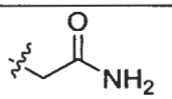
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC$_{50}$)	log(1/MIC$_{90}$)
BD94		Me	CH_2CF_3	8.34	6.92
BD95		Me	CF_3	8.22	6.82
BD96		Me	OCH_3	8.23	6.87
BD97		Me	OCF_3	8.30	6.86
BD98			**Et**	**8.24**	**7.31**
BD99			**Et**	**8.33**	**7.34**
BD100		**CH_2CF_3**	**CH_2CF_3**	**8.64**	**7.45**

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC$_{90}$) and compound **28** (IC$_{50}$)

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

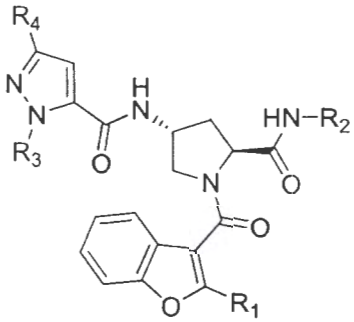


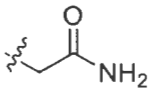
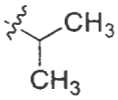
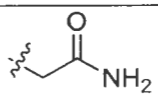
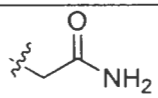
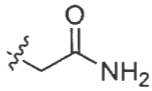
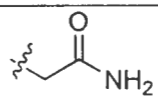
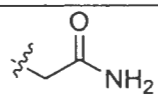
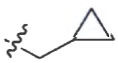
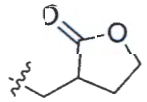
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC50)	log(1/MIC90)
BD101*		**$CH_2CH_2CH_3$**	**CH_2CF_3**	**8.72**	**7.48**
BD102*		**$CH_2CH(CH_3)_2$**	**CH_2CF_3**	**8.83**	**7.60**
BD103			**Et**	**8.43**	**7.46**
BD104		**CH_2CF_3**	**Et**	**8.22**	**7.60**
BD105		**CF_3**	**Et**	**7.93**	**7.32**
BD106		OCH_3	Et	7.97	7.26
BD107		**OCF_3**	**Et**	**8.06**	**7.41**

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

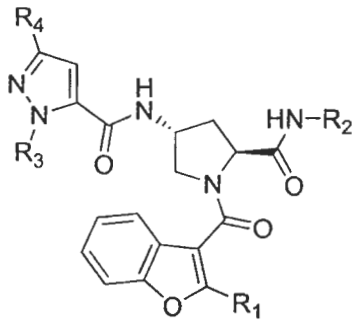


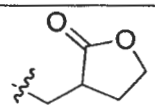
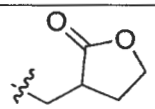
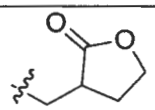
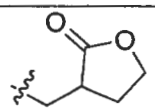
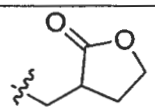
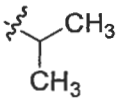
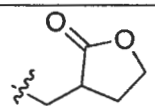
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC$_{50}$)	log(1/MIC$_{90}$)
BD108			Et	8.05	7.30
BD109			Et	7.98	7.18
BD110		**CH_2CF_3**	**CH_2CF_3**	**8.24**	**7.56**
BD111		**$CH_2CH_2CH_3$**	**Et**	**8.29**	**7.63**
BD112		**$CH_2CH(CH_3)_2$**	**Et**	**8.40**	**7.75**
BD113			**Et**	**8.26**	**7.62**
BD114		**Et**	**CH_2CF_3**	**8.44**	**7.34**

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC$_{90}$) and compound **28** (IC$_{50}$)

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)

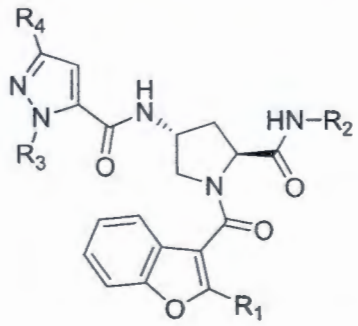


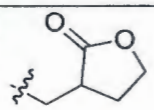
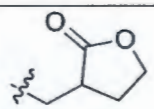
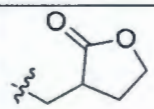
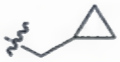
Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC_{50})	log(1/MIC_{90})
BD115		Et	CF_3	8.14	7.06
BD116		Et	OCH_3	8.19	6.99
BD117		Et	OCF_3	8.27	7.14
BD118		CH_2CF_3	CH_2CF_3	8.46	7.30
BD119			CH_2CF_3	8.42	7.12
BD120			CH_2CF_3	8.07	7.12

[a] Bold letter represents the better predicted biological activity than template compound **22**

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.13 Chemical structure and predicted activity of new designed benzofuran pyrrolidine pyrazole derivatives (continued)



Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC_{50})	log(1/MIC_{90})
BD121		$CH_2CH_2CH_3$	CH_2CF_3	**8.53**	**7.33**
BD122		$CH_2CH(CH_3)_2$	CH_2CF_3	**8.64**	**7.45**
BD123			CH_2CF_3	**8.24**	**7.31**

[a] Bold letter represents the better predicted biological activity than template compound **22**

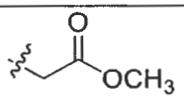
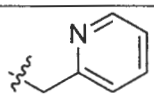
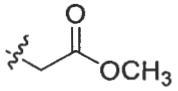
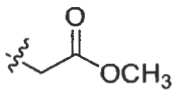
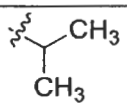
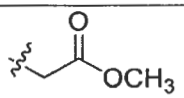
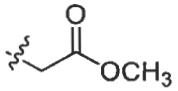
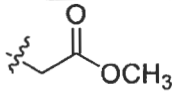
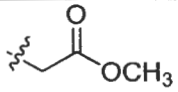
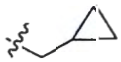
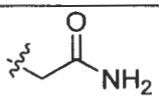
* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

4) Promising benzofuran pyrrolidine pyrazole derivatives based on rational design

Based on the structural basis that derived from 3D-QSAR CoMSIA and MD simulations, 123 novel benzofuran pyrrolidine pyrazole derivatives were designed. Among of these compounds, 17 designed compounds were obtained based on high predicted biological activity against *M. tuberculosis* whole cell as summarized in Table 3.14. Moreover, 2 compounds (**BD101** and **BD102**) are higher predicted activity in both InhA inhibition and *M. tuberculosis* whole cell inhibition.

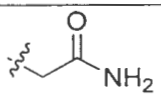
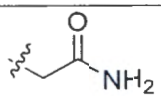
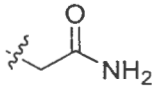
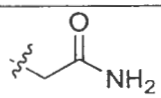
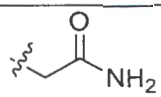
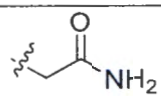
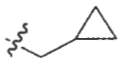
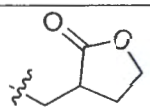
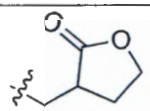
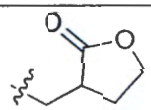
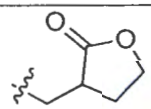
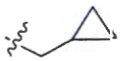
To confirm that high predicted compounds bound in InhA binding site, molecular docking calculations was used to predict the binding mode in InhA binding pocket (Figure 3.33).

Table 3.14 Highly predicted activity benzofuran pyrrolidine pyrazole compounds

Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC_{50})	log(1/MIC_{90})
22		Et	Et	8.52	7.30
28		Et	Et	8.70	6.15
BD98			Et	8.24	7.31
BD99			Et	8.33	7.34
BD100		CH_2CF_3	CH_2CF_3	8.64	7.45
BD101*		$CH_2CH_2CH_3$	CH_2CF_3	8.72	7.48
BD102*		$CH_2CH(CH_3)_2$	CH_2CF_3	8.83	7.60
BD103			Et	8.43	7.46
BD104		CH_2CF_3	Et	8.22	7.60

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

Table 3.14 Highly predicted activity benzofuran pyrrolidine pyrazole compounds (continued)

Cpd.	R_2	R_3	R_4	Predicted biological actives[a]	
				log(1/IC_{50})	log(1/MIC_{90})
BD105		CF_3	Et	7.93	7.32
BD107		OCF_3	Et	8.06	7.41
BD110		CH_2CF_3	CH_2CF_3	8.24	7.56
BD111		$CH_2CH_2CH_3$	Et	8.29	7.63
BD112		$CH_2CH(CH_3)_2$	Et	8.40	7.75
BD113			Et	8.26	7.62
BD114		Et	CH_2CF_3	8.44	7.34
BD121		$CH_2CH_2CH_3$	CH_2CF_3	8.53	7.33
BD122		$CH_2CH(CH_3)_2$	CH_2CF_3	8.64	7.45
BD123			CH_2CF_3	8.24	7.31

* Higher predicted activities (both InhA inhibition and *M. tuberculosis* whole cell) than compound **22** (MIC_{90}) and compound **28** (IC_{50})

The binding mode of all high predicted compounds was displayed in Figure 3.33. Based on the docking binding mode of designed compounds can be classified as two groups. The same binding mode with template compounds (set 1) and different binding mode (set 2) were obtained.

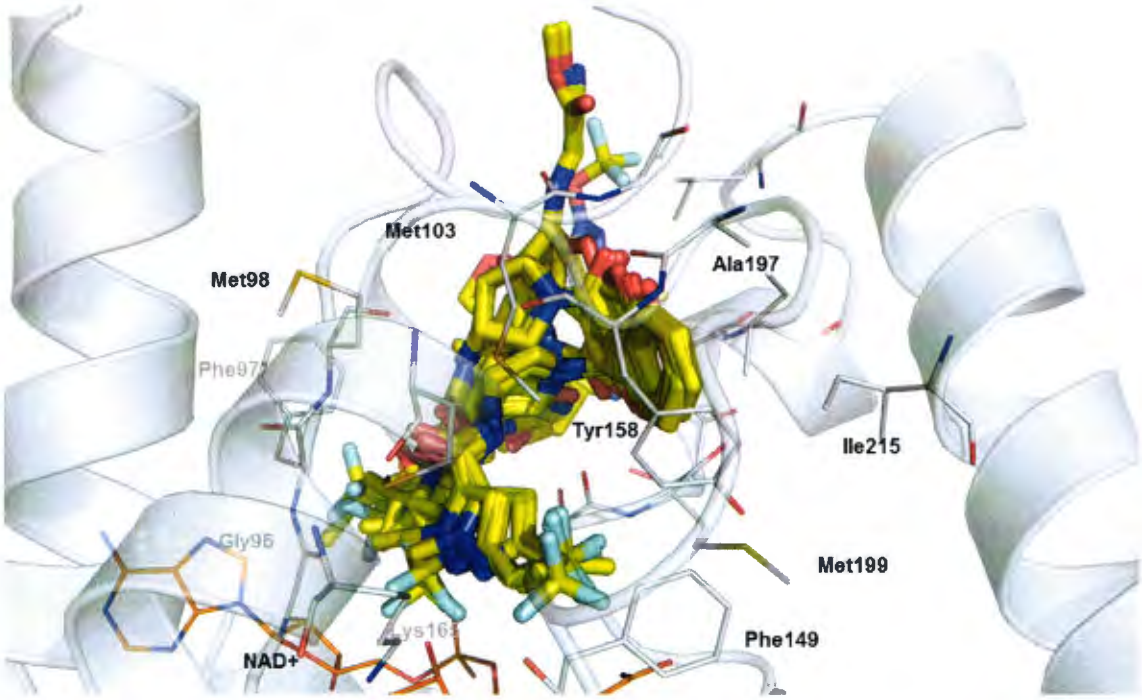


Figure 3.33 Superimposition of new designed benzofuran pyrrolidine pyrazole compounds

The binding mode of **BD102**, the highest InhA inhibition (Table 3.13 and 3.14) was classified as set 2 (different binding mode with template compound). Two hydrogen bond interactions were observed. NH of core structure formed hydrogen bond interaction with oxygen atom of carbonyl group of Met103 backbone. An oxygen atom of carbonyl group at R_2 position formed hydrogen bond interaction with NH of Ile105 backbone. A $CH_2CH(CH_3)_2$ substituent at R_3 position generated the steric effect in the binding pocket. Therefore, the pyrazole core was moved and lost two hydrogen bond interactions (hydrogen bond interaction with NAD^+ and Met98 backbone) in InhA binding pocket as compated to template compound **22**. A $CH_2CH(CH_3)_2$ form hydrophobic interactions with Ile16 and Phe97

side chains. For R_4 (CH_3CF_3) formed hydrophobic interaction with side chain of Ile16 and Ala201 side chain.

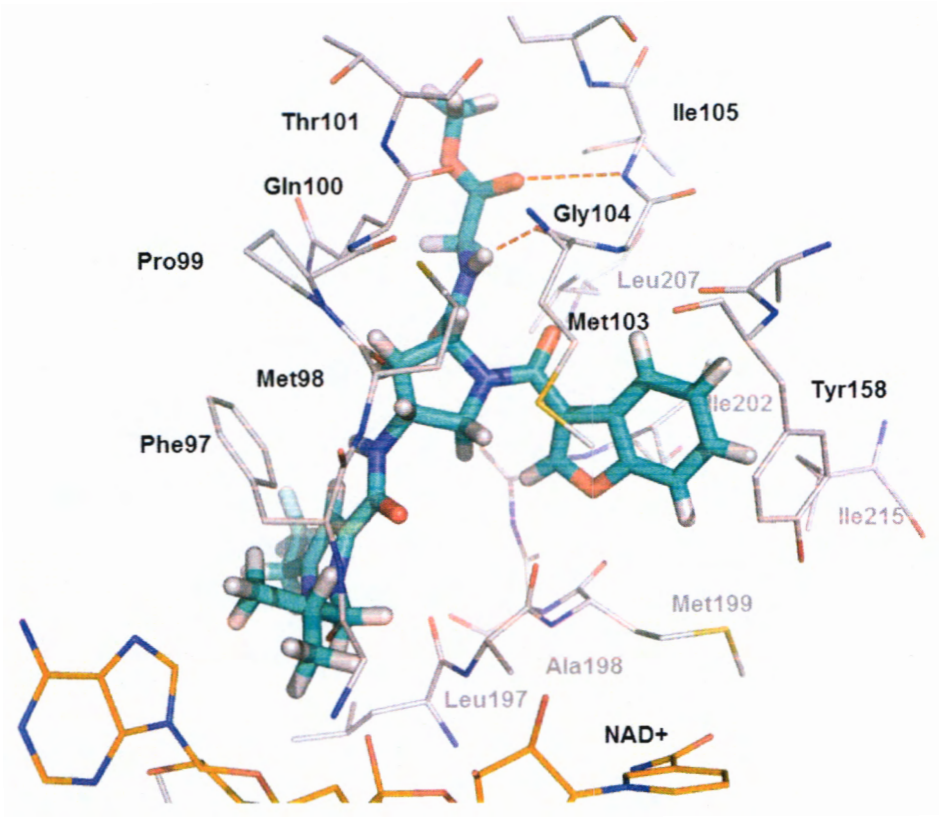


Figure 3.34 Binding mode of BD102 in InhA binding pocket derived from molecular docking

The binding mode of the 2nd high active compound, **BD101** (Table 3.13 and 3.14) was shown in Figure 3.35. The binding mode of this compound was similar to compound **22** (template compound). Five hydrogen bond interactions between core benzofuran pyrrolidine pyrazole were found. An oxygen atom of carbonyl group (C=O) closed to benzofuran ring formed two hydrogen bond interaction with NH backbone of Ala201 and Ile202. NH closed to R_2 position formed hydrogen bond interaction with an oxygen carbonyl (C=O) of Leu197. Three hydrogen bond interactions between pyrazole and amide linker with hydroxyl ribose of NAD^+ cofactor, NH backbone of Met98 and an oxygen carbonyl of Ala198 backbone. R_2 substituent formed hydrophobic interaction with Leu197 backbone. N-propyl R_3 substituent formed hydrophobic interaction with Phe97 and Leu197 backbone.

A CH_2CF_3 substituent at R_4 position formed hydrophobic interactions with Phe149, Tyr158 and Met199 side chain.

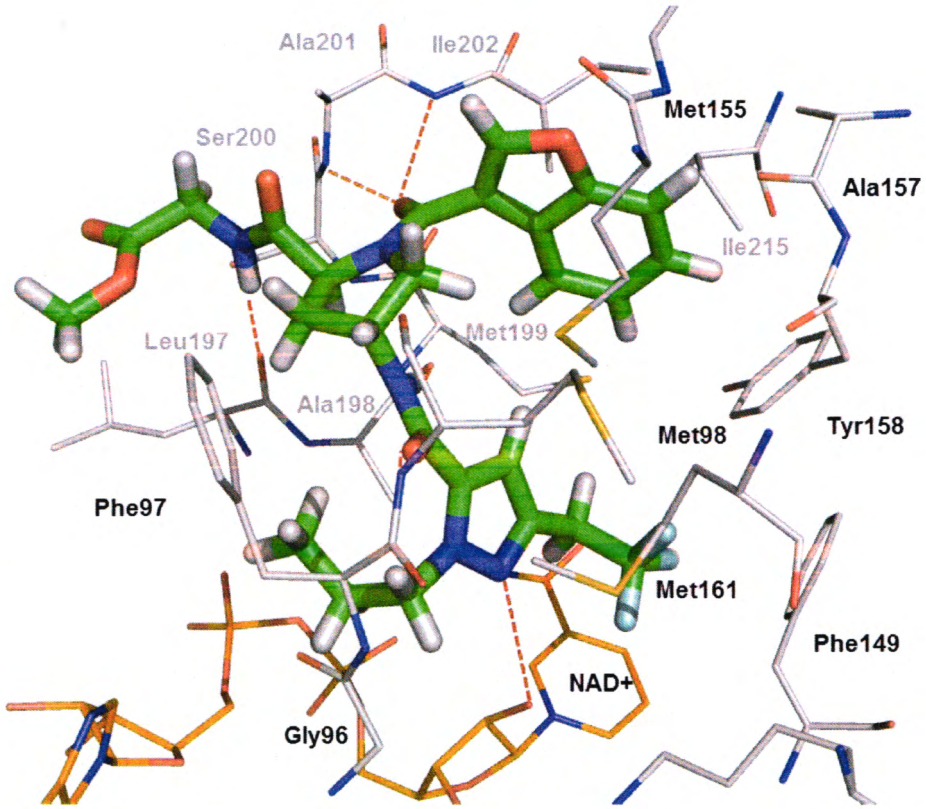


Figure 3.35 Binding mode of BD101 in InhA binding pocket derived from molecular docking

3.1.1.9 Virtual screening of novel InhA inhibitors

1) Receptors identification

Based on RMSD calculations of all atom calculations of 99 monomers compared with mean structure, RMSD values are ranked to be 0.20 Å to 1.43 Å, indicating that InhA enzyme is flexible in the substrate binding site. Therefore, two X-ray crystal structure of InhA complexed with InhA inhibitors were selected as receptor for virtual screening. InhA enzyme complexed with diphenyl ether derivative (PDB code: 2X23) was selected and represented as closed conformations of substrate binding site. InhA enzyme complexed with benzofuran pyrrolidine pyrazole derivative (PDB code: 4COD) was selected and represented as open conformations of substrate binding site. The different binding conformations of substrate binding site and binding

interactions were observed as shown in Figure 3.36. The crucial binding interaction of diphenyl ether (PDB code: 2X23) (Y158-in conformation) is hydrogen bond interaction with Tyr158 (Green color). However, the crucial binding interactions of benzofuran pyrrolidine pyrazole derivative (Y158-out conformation) with InhA are different. Hydrogen bond interaction with Tyr158 was lost. Benzofuran pyrrolidine pyrazole derivative complexed with Met98 is the hydrogen bond interaction.

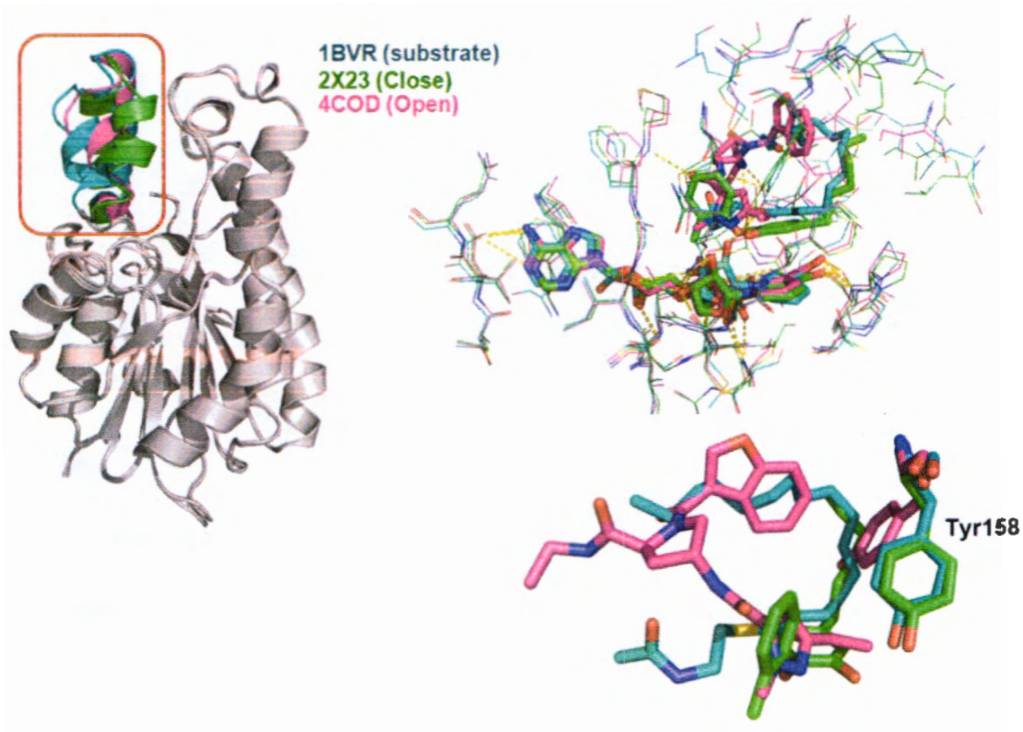


Figure 3.36 Structure of InhA enzyme

2) Validation of docking program

In this study, molecular docking calculations using Glide program was carried out to screen novel InhA inhibitors. Therefore, docking parameters were validated by RMSD values. RMSD of ligand from X-ray conformations and docking conformations was calculated. The RMSD values and docking score of InhA ligand PDB code 2X23 and 4COD were summarized in Table 3.15 and Figure 3.37.

Table 3.15 RMSD of ligand produced by Glide docking with XP scoring function

X-ray	PDB code	
	2X23	4COD
RMSD (Å)	0.49	0.33
Docking score (kcal/mol)	-10.52	-9.12

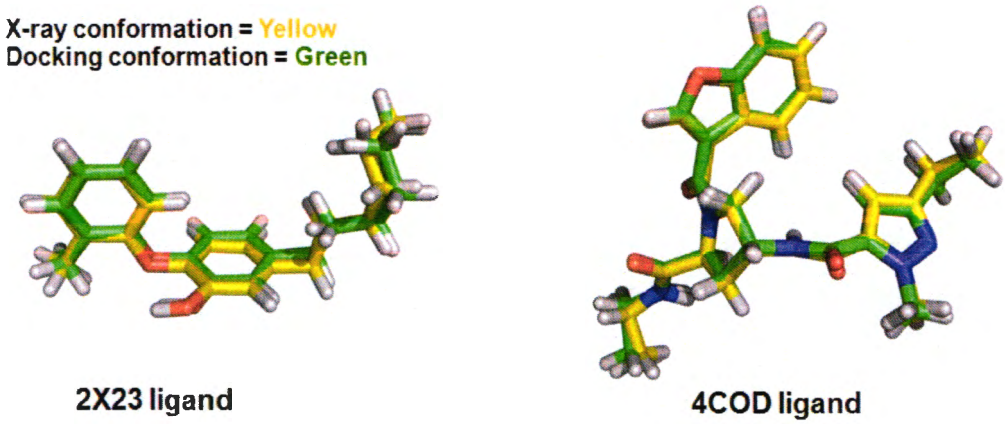


Figure 3.37 Binding mode of InhA ligand from X-ray crystal and docking conformation

Figure 3.37 shows the binding conformation of InhA ligand obtained from the X-ray crystal structure and docking conformations. Form docking calculations, docking parameters produced the binding mode of InhA ligand similar to the binding mode of X-ray crystal structure indicated that docking parameters of Glide docking program predicted the binding mode with high accuracy. Therefore, the docking parameters of Glide docking program can used to screen candidate compounds form chemical database.

3) Virtual screening of InhA inhibitors from Specs database

To identify novel potent InhA inhibitors, Specs database contained 207,369 compounds were selected. The flow diagram of the screening protocol is given in the Figure 3.32. Molecular docking using SP scoring function was performed to identify novel InhA inhibitors into each receptor. Top 2,000 compounds

with the highest docking score were docked into InhA binding site again using XP scoring function. Top 1,000 compounds with highest docking score from each receptor were selected for future analysis. Common compounds among the top 1,000 in each XP docking screened were selected as hit compounds. Thirty-one compounds found in both Y158-in and Y158-out conformations were selected as hit compounds as shown in Table 3.16. The hit compounds were classified as 5-oxo-3-pyrrolidine carboxamide (14 compounds), 1-(3-phenoxybenzyl)-4-(carbonyl)piperazine (3 compounds), 1-(benzimidazol-1-yl)-3- aryloxypropan-2-ol (2 compounds) , 2-(3-benzoylphenyl)-*N*-phenyl propanamide (3 compounds), 2-(3-methoxyphenyl)-2-oxoethyl 3-(benzoyl amino) benzoate (2 compounds) and other derivatives (7 compounds), respectively as shown in Table 3.17, Table 3.18, Table 3.19, Table 3.20, Table 3.21 and Table 3.22.

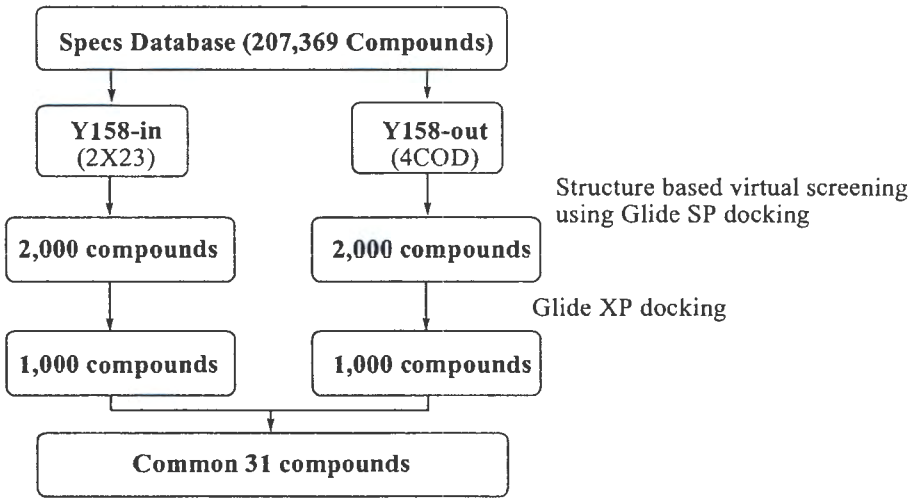


Figure 3.38 Schematic representation of the virtual screening workflow

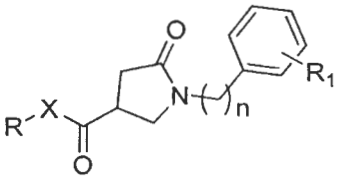
Table 3.16 Selected 31 hit compounds with Glide XP docking scores

Cpd.	Specs ID	Name	Docking score (kcal/mol)	
			Y158-in	Y158-out
1	AK-918/40909651	2-(3-methoxyphenyl)-2-oxoethyl 3-(benzoylamino)benzoate	-9.68	-8.54
2	AG-690/40719693	1-(3-methoxybenzoyl)-4-(3-phenoxy benzyl)piperazine	-10.37	-8.28
3	AN-652/41793331	2-(3-benzoylphenyl)-*N*-(3-methyl phenyl)propanamide	-9.55	-8.23
4	AT-057/43313860	1-(3-phenoxybenzyl)-4-(phenylacetyl) piperazine	-10.99	-8.16
5	AT-057/43313837	1-(phenoxyacetyl)-4-(3-phenoxybenzyl) piperazine	-10.17	-8.07
6	AK-918/42905679	1-benzoylpropyl 1-(4-ethoxyphenyl)-5-oxo-3-pyrrolidinecarboxylate	-10.08	-8.03
7	AK-778/43464987	3-{2-[5-(4-chlorophenyl)-2-furyl]-2-oxoethyl}-3-hydroxy-1,3-dihydro-2*H*-indol-2-one	-9.58	-8.01
8	AT-057/43314035	1-(4-methoxybenzoyl)-4-[(2-naphthyloxy) acetyl]piperazine	-10.97	-7.79
9	AK-087/42718313	1-[4-(benzyloxy)-2-hydroxyphenyl]-2-(hydroxymethyl)-1-butanone	-10.90	-7.78
10	AK-918/42688804	1-methyl-2-oxo-2-phenylethyl 1-(3-methoxyphenyl)-5-oxo-3-pyrrolidine carboxylate	-9.90	-7.78
11	AN-652/42190908	2-(3-benzoylphenyl)-N-(2-methoxyphenyl) propanamide	-9.04	-7.77
12	AN-652/41793320	2-(3-benzoylphenyl)-N-phenyl propanamide	-9.87	-7.73
13	AK-968/15606121	N-(2-carbamoylphenyl)-5-(2,3-dihydro-1H-inden-5-yloxymethyl) furan-2-carboxamide	-11.05	-7.70
14	AK-918/42028907	N-{3-[2-(3-methoxyphenyl)-2-oxoethoxy]phenyl}-2-methyl benzamide	-9.26	-7.66
15	AK-918/42028944	N-[3-(1-methyl-2-oxo-2-phenylethoxy) phenyl]-2-thiophenecarboxamide	-9.61	-7.63
16	AK-968/40358194	4-methyl-N-(3-{[(4-methylbenzoyl) amino]methyl}benzyl)benzamide	-10.34	-7.61
17	AK-918/41957257	2-(3-chloro-4-methylanilino)-2-oxoethyl 5-oxo-1-phenyl-3-pyrrolidinecarboxylate	-9.30	-7.50
18	AF-399/41615593	1-([1,1'-biphenyl]-4-yloxy)-3-(2-methyl-1H-benzimidazol-1-yl)-2-propanol	-10.16	-7.45
19	AK-918/42813987	N-(4-acetylphenyl)-1-benzyl-5-oxo-3-pyrrolidinecarboxamide	-10.11	-7.43
20	AK-918/42688743	1-methyl-2-oxo-2-phenylethyl 1-(2-ethylphenyl)-5-oxo-3-pyrrolidine carboxylate	-9.39	-7.42
21	AK-918/42814152	N-(4-acetylphenyl)-1-(3-chloro-4-methylphenyl)-5-oxo-3-pyrrolidine carboxamide	-10.19	-7.29
22	AK-918/43446442	3,4-dimethylphenyl 1-(3,4-dimethyl phenyl)-5-oxo-3-pyrrolidine carboxylate	-9.42	-7.28

Table 3.16 Selected 31 hit compounds with Glide XP docking scores (continued)

Cpd.	Specs ID	Name	Docking score (kcal/mol)	
			Y158-in	Y158-out
23	AK-918/42813982	1-benzyl-*N*-[4-(3-methyl phenoxy) phenyl]-5-oxo-3-pyrrolidine carboxamide	-10.53	-7.22
24	AK-918/42905452	4-acetylphenyl 1-(4-methylphenyl)-5-oxo-3-pyrrolidinecarboxylate	-9.72	-7.18
25	AP-124/43238118	1-(1*H*-benzimidazol-1-yl)-3-(2,3-dihydro-1*H*-inden-5-yloxy)-2-propanol	-10.16	-7.14
26	AN-988/41349156	3-[(1,1-dioxido-1,2-benzisothiazol-3-yl)amino]phenyl 1-(4-chlorophenyl)-5-oxopyrrolidine-3-carboxylate	-10.56	-6.99
27	AK-918/42813833	*N*-(4-acetylphenyl)-1-(3-methoxy phenyl)-5-oxo-3-pyrrolidine carboxamide	-9.54	-6.91
28	AN-919/13592066	3-{[2-(3-methoxyphenyl)ethyl]amino}-1-(4-methylphenyl)-2,5-pyrrolidine dione	-9.48	-6.90
29	AK-918/42905451	3,4-dimethylphenyl 1-(4-methyl phenyl)-5-oxo-3-pyrrolidine carboxylate	-9.19	-6.86
30	AG-205/36264055	2-[1,1'-biphenyl]-4-yl-2-oxoethyl 5-oxo-1-phenyl-3-pyrrolidinecarboxylate	-10.11	-6.84
31	AG-205/36264046	2-oxo-2-phenylethyl 1-(4-fluoro phenyl)-5-oxo-3-pyrrolidine carboxylate	-9.17	-6.84

Table 3.17 Structure of hit compounds in the series of 5-oxo-3-pyrrolidine carboxamide derivatives



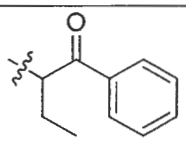
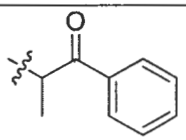
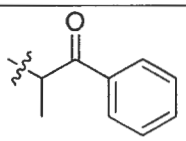
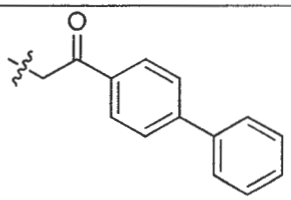
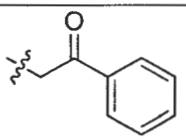
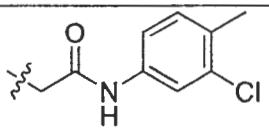
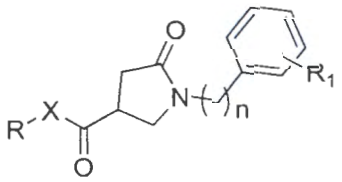
Cpd.	Specs ID	R	X	n	R_1
1	AK-918/42905679		O	0	4-OEt
2	AK-918/42688804		O	0	3-OMe
3	AK-918/42688743		O	0	2-Et
4	AG-205/36264055		O	0	H
5	AG-205/36264046		O	0	4-F
6	AK-918/41957257		O	0	H

Table 3.17 Structure of hit compounds in the series of 5-oxo-3-pyrrolidine carboxamide derivatives (continued)



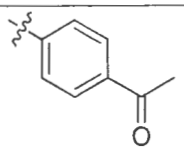
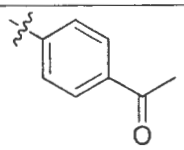
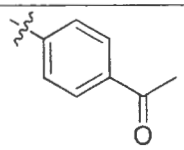
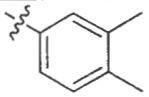
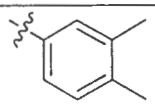
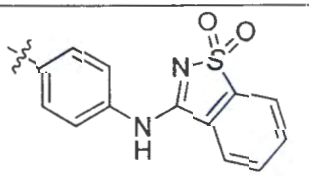
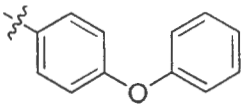
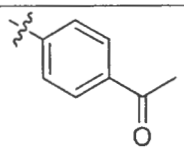
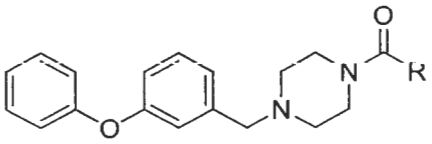
Cpd.	Specs ID	R	X	n	R_1
7	AK-918/42814152		NH	0	3-Cl, 4-Me
8	AK-918/42813833		NH	0	3-OMe
9	AK-918/42905452		O	0	4-Me
10	AK-918/43446442		O	0	3,4-diMe
11	AK-918/42905451		O	0	4-Me
12	AN-988/41349156		O	0	4-Cl
13	AK-918/42813982		NH	1	H
14	AK-918/42813987		NH	1	H

Table 3.18 Structure of hit compounds in the series of 1-(3-phenoxybenzyl)-4-(carbonyl)piperazine derivatives



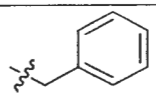
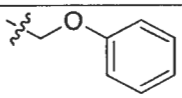
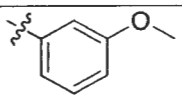
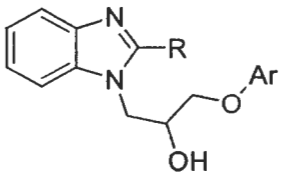
Cpd.	Specs ID	R
1	AT-057/43313860	
2	AT-057/43313837	
3	AG-690/40719693	

Table 3.19 Structure of hit compounds in the series of 1-(benzimidazol-1-yl)-3-aryloxypropan-2-ol derivatives



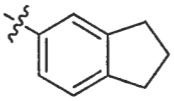
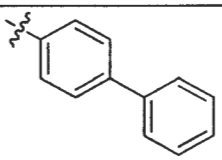
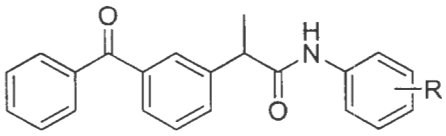
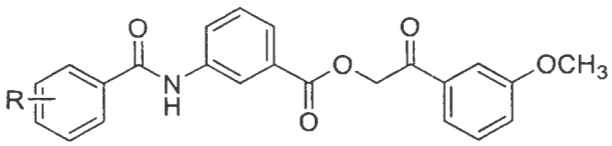
Cpd.	Specs ID	Ar	R
1	AP-124/43238118		H
2	AF-399/41615593		Me

Table 3.20 Structure of hit compounds in the series of 2-(3-benzoylphenyl)-*N*-phenylpropanamide derivatives



Cpd.	Specs ID	R
1	AN-652/41793331	3-Me
2	AN-652/42190908	2-OMe
3	AN-652/41793320	H

Table 3.21 Structure of hit compounds in the series of 2-(3-methoxyphenyl)-2-oxoethyl 3-(benzoylamino)benzoate derivatives



Cpd.	Specs ID	R
1	AK-918/40909651	H
2	AK-918/42028907	2-Me

Table 3.22 Structure of other hit compounds

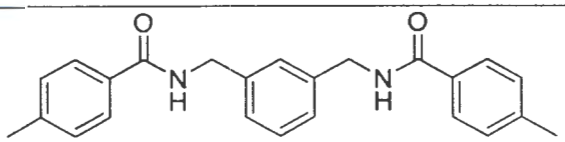
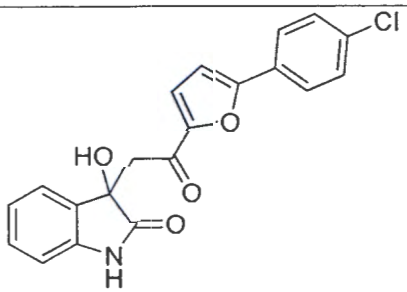
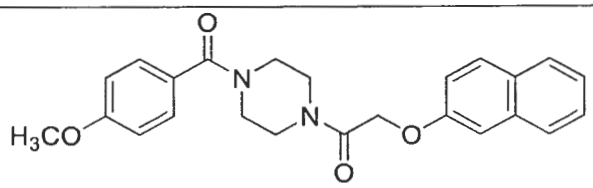
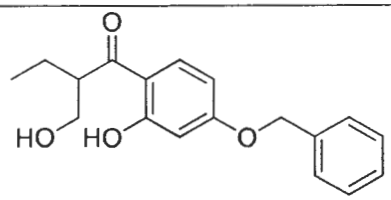
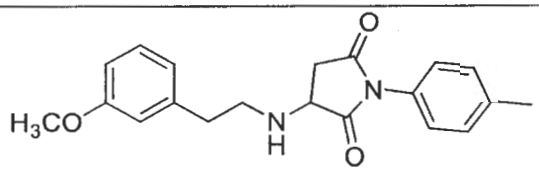
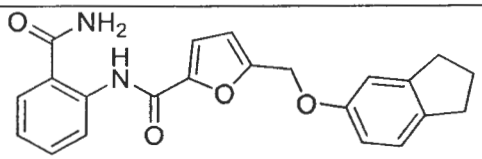
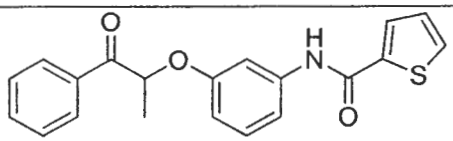
Cpd.	Specs ID	Structure
1	AK-968/40358194	
2	AK-778/43464987	

Table 3.22 Structure of other hit compounds (continued)

Cpd.	Specs ID	Structure
3	AT-057/43314035	
4	AK-087/42718313	
5	AN-919/13592066	
6	AK-968/15606121	
7	AK-918/42028944	

Based on Glide XP docking score, we found that doking score of all hit compounds in Y158-in are higher than docking score in in Y158-out binding pocket. Previously, 1-(3-phenoxybenzyl)-4-(carbonyl)piperazine derivatives were reported as active compounds against InhA and mycobacterial whole cell (Kinjo et al., 2013; Kanetaka et al., 2015) indicated that the high performant to identify active InhA inhibitors of structure based virtual screening in this work was obtained. In addition, -(3-phenoxybenzyl)-4-(carbonyl)piperazine derivatives were found as high Glide XP docking score in Y158-in binding pocket (Table 3.16).

N-(2-carbamoylphenyl)-5-(2,3-dihydro-1*H*-inden-5-yloxy methyl) furan-2-carboxamide (AK-968/15606121) was found as highest Glide XP docking score in Y158-in binding site. This compound contained 2-carbamoylphenyl (black color), furan-2-carboxamide (blue color) and 5-methoxy-2,3-dihydro-1*H*-

indene (green color) fragments, respectively. The binding mode of this compound was displayed in Figure 3.39. Three hydrogen bond interactions were observed. Carbonyl of furan-2-carboxamide fragment formed hydrogen bond interaction with hydroxyl group (OH) of Tyr158 and nicotinamide ribose. Other one hydrogen bond interaction is the interaction between NH of primary amide of 2-carbamoylphenyl with carbonyl backbone of Gly96. Two pi-pi interactions were found between furan ring of furan-2-carboxamide fragment with nicotinamide ring of NAD^+ cofactor and phenyl ring of 5-methoxy-2,3-dihydro-1*H*-indene fragment with phenyl ring of Tyr158 as shown in Figure 3.39. As compared hit compound (AK-968/15606121) conformation in InhA binding pocket with diphenyl ether derivatives (PDB 2X23) this compound was well superimposed. 2-Carbamoylphenyl fragment of hit compound overlapped with phenyl B ring of diphenyl ether InhA inhibitors. Primary amide substituent on phenyl ring at *ortho* position (2-carbamoylphenyl) overlapped with methyl substituent of diphenyl ether derivative. Additionally hydrogen bond interaction between NH (primary amide) and carbonyl backbone of Gly96 was found. As considered furan-2-carboxamide fragment with diphenyl ether compound, the carbonyl functional located close to the location of hydroxyl (OH) group of diphenyl ether compound and formed two hydrogen bond interactions with Tyr158 and nicotinamide ribose that we found from OH of diphenyl ether compound. NH of secondary amide acts as linker between furan ring with phenyl ring of 2-carbamoylphenyl like oxygen ether of diphenyl ether compound. For furan ring, this ring formed pi-pi interaction with nicotinamide ring of NAD^+ like phenyl A ring of diphenyl ether derivative. 5-Methoxy-2,3-dihydro-1*H*-indene fragment of hit compound overlapped with hexyl part of diphenyl ether that located in hydrophobic binding pocket. This fragment formed hydrophobic interactions with amino acids in InhA binding pocket, i.e. Phe149, Met155, Pro156, Ala157, Tyr158, Pro193, Ile194, Met199, Val03, Leu207, Gln214, Ile215, Leu218 and Trp222. In addition, pi-pi interaction of 2,3-dihydro-1*H*-indene ring with Tyr158 was observed. Therefore, AK-968/15606121 hit compound (Glide XP docking score = -11.05 kcal/mol) in Y158-in binding site showed strongly binding affinity than diphenyl ether compound (Glide XP docking score = -10.52 kcal/mol).

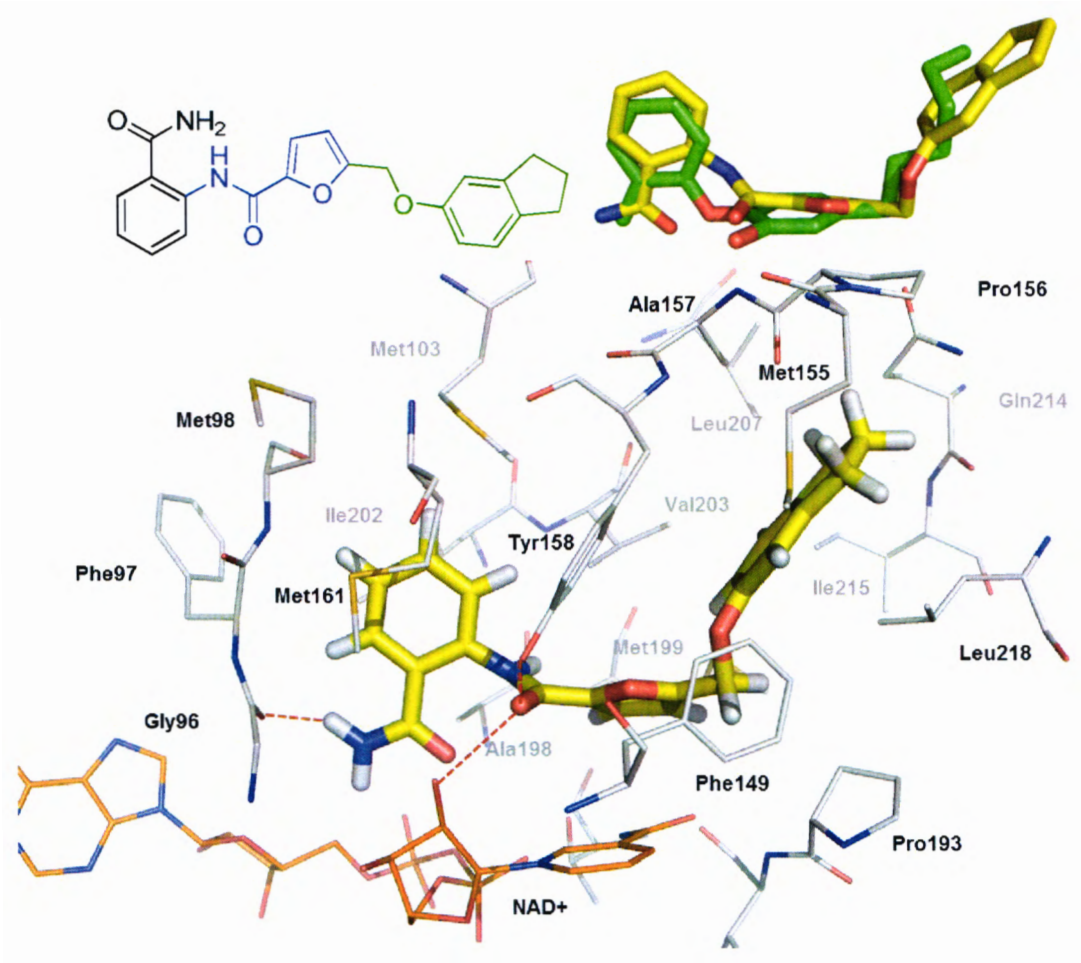


Figure 3.39 Binding mode of AK-968/15606121

The binding interactions of highest Glide XP docking score of Y158-out hit compounds, 2-(3-methoxyphenyl)-2-oxoethyl-3-(benzoylamino)benzoate (AK-918/ 40909651) in InhA binding site were analyzed as shown in Figure 3.40. This compound consists of three parts, benzamide (black color), benzoate (blue color) and 3'-methoxyacetophenone (green color) fragments, respectively. Carbonyl of benzamide fragment overlapped with pyrazole ring of benzofuran pyrrolidine pyrazole derivatives, whereas benzene ring overlapped with ethyl substituent on pyrazole ring. This carbonyl formed hydrogen bond interaction with hydroxyl (OH) functional of nicotinamide ribose of NAD^+. Benzene ring formed pi-pi and sigma-pi interactions with nicotinamide of NAD^+ and Phe149, respectively. Benzoate fragment of hit compound covered the amide linker and pyrrolidine ring of benzofuran pyrrolidine pyrazole derivatives. This fragment formed hydrophobic interaction with Gly96,

Phe97, Met103, Met161 and Ala198. 3'-Methoxyacetophenone fragment overlapped with benzofuran and their inker. Carbonyl functional acts as hydrogen bond acceptor to form hydrogen bond interaction with NH backbone of Ile202 in InhA binding pocket. Hydrophobic interactions of 3'-methoxyacetophenone fragment with Met103, Gly104, Ala157, Tyr158, Met199, Ile202, Leu207 and Ile215 were observed. There results were confirmed from Glide XP docking score. Glide XP docking score of hit compound (Glide XP docking score = -8.54 kcal/mol) is lower than benzofuran pyrrolidine pyrazole compound (Glide XP docking score = -9.12 kcal/mol).

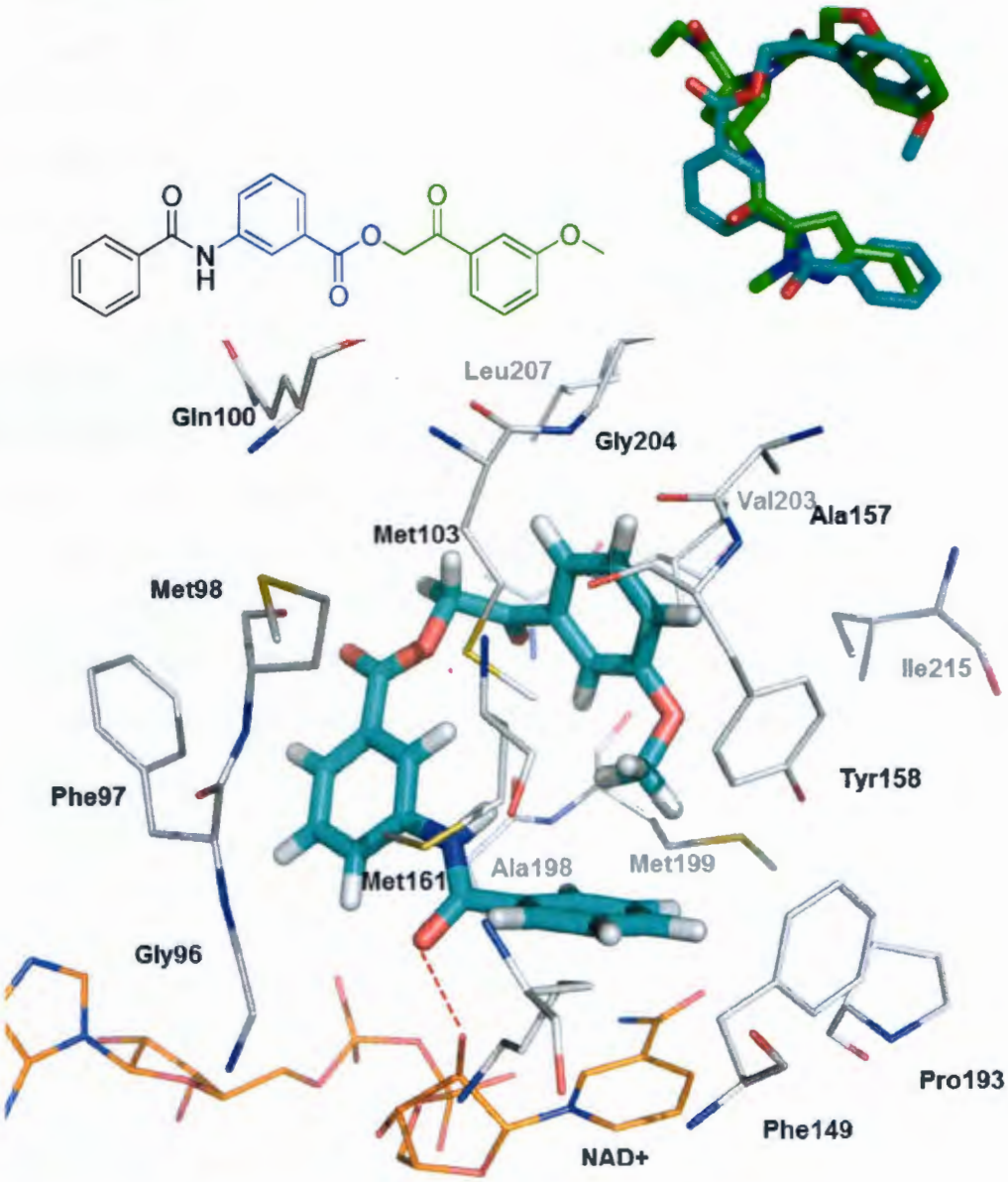


Figure 3.40 Binding mode of AK-918/40909651

3.1.2 Serine/Threonine kinase G (PknG) inhibitors

3.1.2.1 3D-QSAR study

1) CoMSIA models

The statistical parameters of CoMSIA models were summarized in Table 3.23. Based on the internal statistical results shown in Table 3.23, two models, model 8 and 9 have predictive ability (q^2) higher than 0.6. The highest q^2 (0.72) of CoMSIA model was model 9. However, this model provided information only steric, electrostatic, hydrogen bond donor and hydrogen bond acceptor fields. This model is not enough informative data to investigate the structural requirement of PknG inhibitors in term of hydrophobic field. Model 9 provided steric, electrostatic, hydrogen donor and hydrogen bond acceptor information for rational design. Whereas, model 8 provided steric, electrostatic, hydrophobic, hydrogen bond donor information to enhance the biological activity. To derive in more detail of structural informative of PknG inhibitors, CoMSIA model that provided more information were considered. Therefore, model 8 was selected as the best CoMSIA model in this study. Cross-validated q^2 of 0.63 with five components and non-cross-validated PLS analysis resulted in a correlation coefficient r^2 of 0.96, F = 110.16, and an estimated standard error of 0.27. To access the power of the selected CoMSIA model, test set compounds were used as external validation. Predicted activity derived from CoMSIA model closed to experimental activity (Figure 3.41). Residue between experimental and predicted activity of all test set compounds were lower than one logarithm unit indicating that best CoMSIA model was reliable based on external validations.

Table 3.23 Summary of statistical results of CoMSIA models

Model	q^2	r^2	s	SEE	F	N	Fractions
1.S/E	0.61	0.97	0.89	0.23	120.88	6	29.8/70.2
2.S/H	0.54	0.70	0.88	0.72	27.53	2	16.0/84.0
3.S/D	0.72	0.96	0.75	0.27	89.61	6	31.3/68.7
4.S/A	0.64	0.93	0.84	0.37	56.69	5	47.1/52.9
5.S/E/D	0.77	0.96	0.66	0.29	92.92	5	14.0/38.3/47.7
6.S/H/D	0.61	0.95	0.87	0.30	83.88	5	12.0/48.9/39.1
7.S/D/A	0.68	0.96	0.81	0.29	80.03	6	21.6/52.1/26.4
8.S/E/H/D	**0.63**	**0.96**	**0.85**	**0.27**	**110.16**	**5**	**8.8/24.7/35.7/30.8**
9.S/E/D/A	0.72	0.96	0.76	0.28	84.73	6	13.7/33.3/38.0/15.0
10.S/E/H/A	0.43	0.98	1.08	0.22	143.57	6	12.7/28.1/43.1/16.0
11.S/E/H/D/A	0.55	0.97	0.94	0.25	123.21	5	8.8/20.4/31.5/26.0/13.2

Bold values indicate the best CoMSIA model

N, optimum number of components; s, standard error of prediction; SEE, standard error of estimate; F, F test value; S, steric field; E, electrostatic field; H, hydrophobic; D, hydrogen donor field; A, hydrogen acceptor field

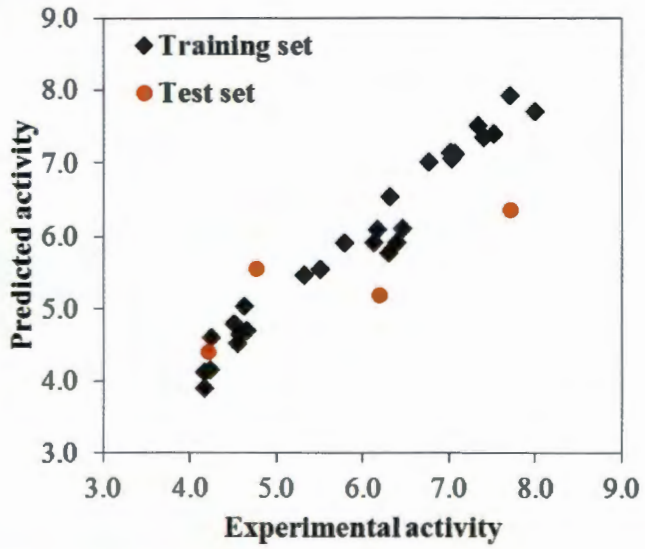


Figure 3.41 Plots between the experimental and predicted activities of the training and test sets derived from the CoMSIA model

2) CoMSIA contour maps

To visualize the structural requirement of PknG inhibitors, CoMSIA contour maps were established. CoMSIA contour maps that reveal the influence of steric, electrostatic, hydrophobic, and hydrogen donor fields to the activity of benzothiophene derivatives are shown in Figure 3.42. Compound **27c** as the most active compound was combined with a graphic interpretation of CoMSIA contour maps (Table 2.8 in chapter 2). Favorable and unfavorable steric regions to enhance the activities of benzothiophene derivative are represented by green and yellow contours, respectively, whereas blue and red contours indicate regions where electropositive and electronegative groups lead to an increase of the PknG inhibitory activity, respectively. Magenta and white contours represent areas, where the hydrophobic and the hydrophilic group are predicted to favor the biological activities against PknG enzyme. The cyan and purple contours indicate regions that favor the hydrogen bond donor group and hydrogen bond acceptor group, respectively.

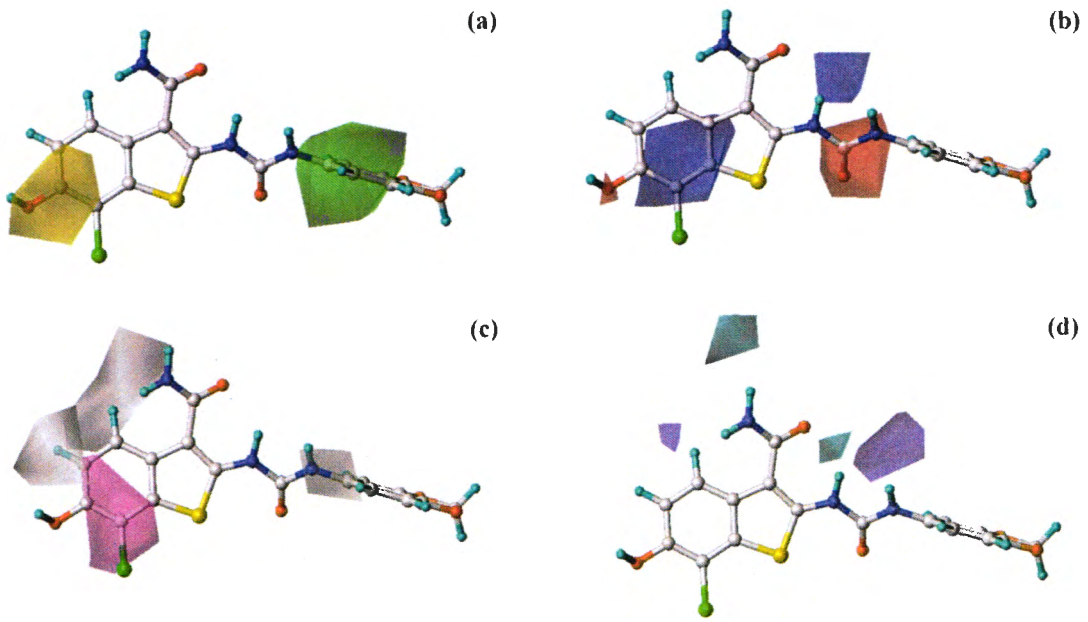


Figure 3.42 Steric (a), electrostatic (b), hydrophobic (c) and hydrogen bond donor (d) contours of the best CoMSIA model in combination with compound 27c

3) Structural requirement at R_1 position

White and cyan contours located near R_1 position of compound **27c** indicated that this position required hydrophilic and hydrogen bond donor properties of substituent to improve the biological activity. This could be explained by the fact that compounds **3b** ($\log(1/IC_{50}) = 6.30$) having NH_2 substitution at R_1 favored the activity while in compound **13b** ($\log(1/IC_{50}) = 4.66$) and compound **22c** ($\log(1/IC_{50}) = 4.24$) that contained NCH_3 and OCH_2CH_3, more hydrophobic groups at R_1 position showed less inhibitory activity. Therefore, hydrophilic and hydrogen bond donor properties such as NH_2 should be contained in this position.

4) Structural requirement at R_2 position

Green and white contours closed to R_2 position of compounds **30b** indicated that that bulky substituent with hydrophilic property of substituents will enhance the activity at these positions. This requirement can be explain by compounds **3b-12b** ($\log(1/IC_{50}) = 4.21$-6.39). The biological activity of tetrahydrobenzothiophene derivatives were increased by the size of R_2 substituent. However, size of R_2 substituent should be optimal. If contained large R_2 substituent, the biological activity against PknG was decreased like compounds **12b**. In case of hydrophilic property requirement at R_2 position was confirmed by compounds **26c** ($\log(1/IC_{50}) = 7.33$), **29c** ($\log(1/IC_{50}) = 7.70$) and **27c**($\log(1/IC_{50}) = 8.00$). Bulky substituent with hydrophilic property like compounds **27c** showed higher biological activity than compounds **26c** and **29c**. These results indicated that R_2 position is important part to modify and improve biological active of tetrahydrobenzothiophene derivatives. The big changed of experimental biological data as compared to the substituents of R_2 position were found. Therefore, the size of substituents on this position with hydrophilic property should be optimized.

5) Structural requirement at X position

Yellow and magenta contours located at X position indicated that this position required small group with hydrophobic property to improve biological activity of tetrahydrobenzothiophene derivatives. This can explain why compound **1a** ($\log(1/IC_{50}) = 6.46$) is potent than compound **3b** ($\log(1/IC_{50}) = 6.30$) and **2a** ($\log(1/IC_{50}) = 4.17$) in case of steric effect. Compounds **3b** ($\log(1/IC_{50}) = 6.30$) , **14b-19b** ($\log(1/IC_{50}) = 4.17$-6.20) were used as example to explain the effect of

hydrophobic property on X position. At X position of compounds **14b-19b** contained hydrophilic compared to compounds **3b** indicated that compounds **3b** was high potent than compounds **14b-19b**. Therefore, this position should be contained small substituent with hydrophobic property to improve biological activity.

3.1.2.2 Molecular docking calculations

The best docking score of each compound the derived from Glide XP were summarized in Table 3.24. The docked binding mode of all compounds is similarly with tetrahydrobenzothiophene compound (AX20017) bound in X-ray crystal structure (Scherr et al., 2007). Docking score can be classified between active and inactive compounds (Compounds showing IC_{50} against PknG higher than 100 μM as shown in Table 2.8 in chapter 2) of PknG inhibitors as shown in box plot (Figure 3.43a). The medium values of docking score of active and inactive compounds derived from box plot are -6.70 and -5.83 kcal/mol, respectively. It can be seen that the Glide XP docking scores of active and inactive compounds were well separated. Moreover, maximum docking score of inactive compounds (-6.44 kcal/mol) is low binding affinity than medium value of active compounds (-6.70 kcal/mol). These results revealed that docking results were reliable to estimate the binding affinity of active and inactive compounds in this study. The predicted binding mode of PknG inhibitors in this studied could be correctly binding active conformation in PknG binding pocket. The superimposition of all compounds bound to PknG binding pocket derived from molecular docking is in the similarly fashion as compared to the X-ray structure of tetrahydrobenzothiophene (AX20017) compounds (Scherr et al., 2007). Carbonyl backbone of Glu233 and NH backbone of Val235 conserved hydrogen bond interactions with primary amide of these compounds, except compound **22c**. To elucidate in more details of crucial binding interactions and effect of substituent of inhibitors on biological activity against PknG, two selected compounds, compounds **21c** and **27c** were subjected to future investigate using MD simulations. Poorly correlation (q^2=0.32 and RMSE = 1.04 as shown in Table 3.25 and Figure 3.43b) of docking score with biological data was obtained. This result is the main problem of docking calculations. Docking calculations is fast and simple method. Whereas, the correlation of estimated binding free energy with inhibitory activity is low as we found in this study (q^2=0.32). Therefore, post docking methods to estimate the binding

affinity were performed. This study, different binding free energy calculations methods, MM-PBSA and MM-GBSA approaches were applied to estimate binding affinity of tetrahydrobenzothiophene derivatives in PknG binding pocket based on the docking binding mode derived from this docking study. Furthermore, the effect of different MM-GBSA solvation models (igb1, igb2, igb5, igb7 and igb8) were also calculated to determine the binding affinity of tetrahydrobenzothiophene derivatives in PknG binding pocket and to increase the correlations of biological activity and calculated binding free energy.

Table 3.24 Predicted binding energy of PknG inhibitors derived from Glide XP

Cpd.	IC_{50} (μM)	$log(1/IC_{50})$	XP score (kcal/mol)
1a	0.35	6.46	-6.62
2a	68	4.17	-6.19
3b	0.500	6.30	-6.76
4b	31	4.51	-6.12
5b	23	4.64	-6.69
6b	17	4.77	-6.64
7b	4.79	5.32	-6.77
8b	1.66	5.78	-6.83
9b	56	4.25	-6.60
10b	62	4.21	-6.54
11b	0.41	6.39	-6.84
12b	26	4.59	-6.70
13b	22	4.66	-5.24
14b	0.63	5.20	-7.87
15b	68	4.17	-5.99
16b	0.74	6.13	-6.79
17b	28	4.55	-6.63
18b	0.68	6.17	-6.30
19b	0.085	7.07	-6.74

Table 4.24 Predicted binding energy of PknG inhibitors derived from Glide XP (continued)

Cpd.	IC_{50} (μM)	$log(1/IC_{50})$	XP score (kcal/mol)
20b	58	4.24	-6.27
21c	0.095	7.02	-6.40
22c	0.093	7.03	-6.01
23c	0.17	6.77	-6.49
24c	3.16	5.50	-7.31
25c	0.49	6.31	-7.18
26c	0.047	7.33	-7.28
27c	0.01	8.00	-7.79
28c	0.05	7.30	-8.31
29c	0.02	7.70	-7.18
30c	0.04	7.40	-7.04
31c	0.03	7.52	-7.32
In1	>100	<5.49	-5.02
In2	>100	<5.49	-2.65
In3	>100	<5.49	-5.82
In4	>100	<5.49	-6.28
In5	>100	<5.49	-6.44
In6	>100	<5.49	-5.87
In7	>100	<5.49	-5.41
In8	>100	<5.49	-5.84

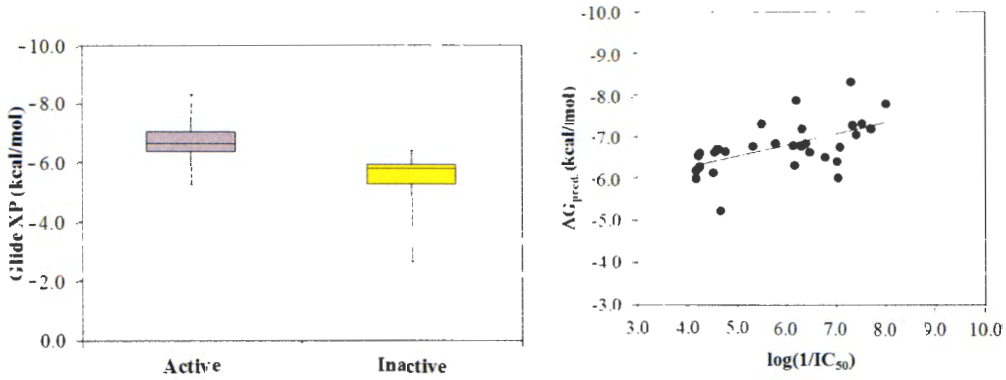


Figure 3.43 Box plot of Glide XP scoring distribution obtained from active and inactive compounds

Table 3.25 Summary of the medium value and statistical parameters obtained from Glide XP for PknG

Method	Medium value of active	Medium value of inactive	RMSE	q^2
Glide XP	-6.70 kcal/mol	-5.83 kcal/mol	1.04	0.32

3.1.2.3 MD simulations

1) Structural stability during MD simulations

Two complexes of PknG inhibitors (compounds **21c** and **27c**) were modeled using MD simulations. The RMSDs for all atoms of PknG and inhibitor in each complex over the 15 ns of simulation times were plotted in order to determine the structural stability of each complex (Figure 3.44). Convergent RMSD plots indicated that the equilibrium states of compounds **21c** and **27c** were reached for each system after 3 ns and 12 ns, respectively. Average RMSD after reached equilibrium state of compound **21c** was small fluctuation between 2.5 and 3.5 Å, whereas, average RMSD of compound **27c** in PknG binding pocket was large fluctuation between 3.0 and 4.0 Å. These results can be explained using the information of size of two compounds. A total polar surface area (TPSA) and total volume were calculated using Molinspiration Cheminformatics Software (Ertl. Rohde and Selzer, 2000) and considered. Compound **27c** (Volume =309.70 $Å^3$, TPSA=122.92) showed higher

volume than compound **21c** (Volume = 220.00 Å^3, TPSA =72.19) that required bigger binding pocket to bind in PknG binding site as compared to compound **21c**. Therefore, more details in terms of binding free energy, structural complex and binding interactions were mainly analyzed over the last 2 ns.

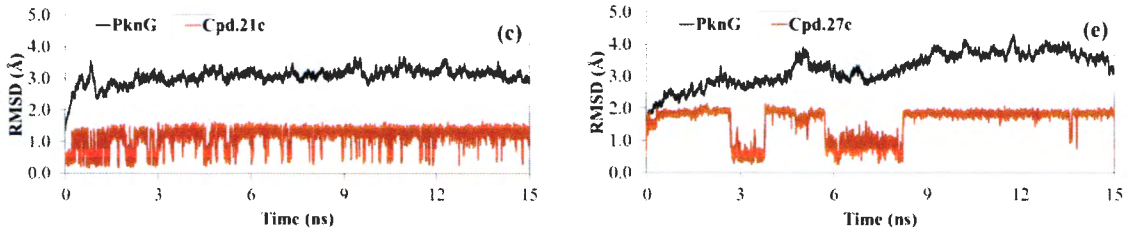


Figure 3.44 RMSDs for all atoms of PknG and inhibitor in each complex over the 15 ns of simulation time

2) Binding mode of tetrahydrobenzothiophene and benzothiophene derivatives

In the X-ray crystal structure of PknG-AX20017 (**3b**) complex, Glu233 (O) and Val235 (NH) of PknG form hydrogen bonds to the primary amide group of compound **3b**. Additionally, a hydrogen bond is made between Val235 (NH) and the secondary amide group of compound **3b** (Scherr et al., 2007). This hydrogen bonding network is conserved for binding of compound **27c** (the most active compound) in the PknG pocket (Figure 3.45). This implies that the conserved hydrogen bonding network is the crucial interaction for binding of benzothiophene derivatives. Apart from the conserved hydrogen bonds, additional hydrogen bonds are formed between compound **27c** and Ile292(O), Asp293(OD), Gly237(NH) and Lys181 (Figure 3.45). These interactions are responsible for the better activity of compound **27c** as compared with compound **3b**. With regard to the benzothiophene moiety, it is made van der Waals interactions with residues Ile157, Ala158, Ile165, Val179, Met232, Val258 and Ile292 sidechains. The benzo dioxole (the $\mathbf{R_2}$ substituent) interacts with residues Ile157, Tyr234, Gly236, Gly237, Gln238 and Met283 to form van der Waals interactions. However, some moiety of the $\mathbf{R_2}$ substituent protrudes from the PknG pocket (Figure 3.46).

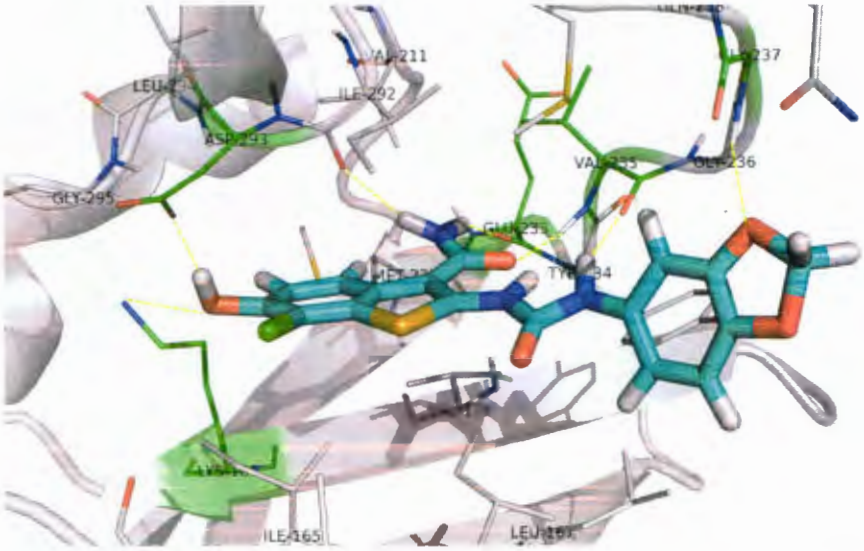


Figure 3.45 Binding modes of compound 27c in the PknG pocket obtained from MD simulations

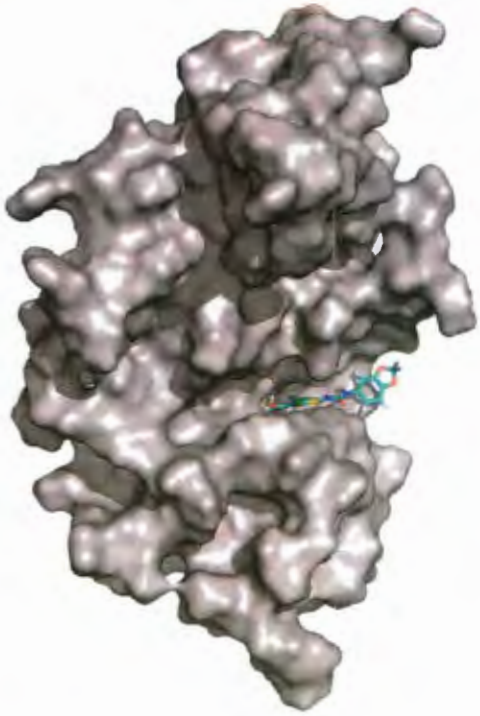


Figure 3.46 Compound 27c (cyan) in its complex with whole PknG (grey) obtained from MD simulations

3) The structural requirement of benzothiophene derivatives for PknG inhibition

A graphic interpretation of CoMSIA contour maps (Figure 3.42) combined with the analysis of PknG-inhibitor complexes obtained from MD simulations were utilized to reveal the structural requirement of thiophene derivatives for PknG inhibition. Compound **27c** presented the best IC_{50} value for PknG inhibition was used as the template for a graphic interpretation of CoMSIA contour maps (Table 2.8 in chapter 2). Chemical structures of data set were divided into three common structures. They have different substituents of $\mathbf{R_1}$, $\mathbf{R_2}$ and membered ring. The structural requirement of benzothiophene derivatives were discussed as below.

3.1) Membered ring substituent

Membered ring substituents were classified as cycloalkane and benzene ring. CoMSIA contour maps show yellow and purple contours at membered benzene ring of compound **27c**. This result means that the small substituent with hydrophobic property is preferable at membered ring substituent position. Most of compounds that contain the membered benzene ring (compounds **21c**, **23-31c**) show IC_{50} values better than those bearing the membered benzene ring (compounds **1-2a**, **3-20b**). This implies that benzene ring is suitable for steric and hydrophobic requirement of membered ring substituent. To investigate the influence of membered cycloalkane and benzene ring on an inhibitory activity against PknG, the complex structure of PknG-**21c** was predicted using MD simulation. Its binding modes found in the PknG pocket is compared with that of compound **3b** (Figure 3.47).

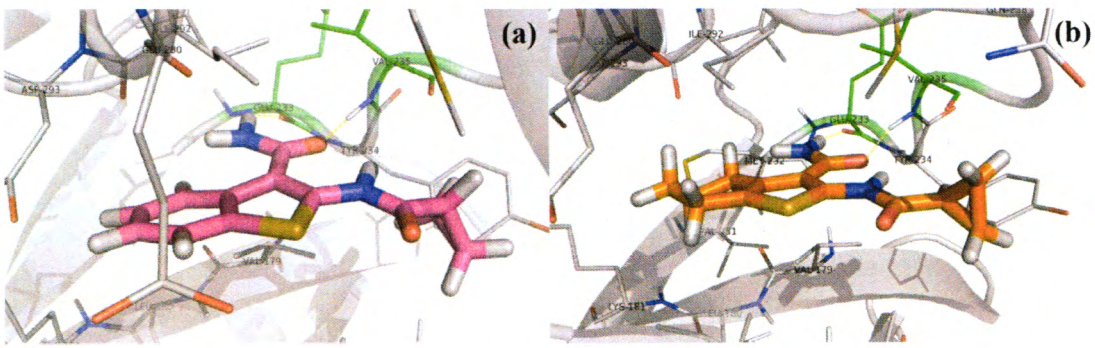


Figure 3.47 Binding modes of compounds 21c (a) and 3b (b) found in the PknG pocket

3.2) R_1 substituent

Except of compounds **13b** and **22c**, all compounds in the data set present the NH_2 at the $\mathbf{R_1}$ substituent. Hydrogen bond donor contour (Figure 3.42d) shows cyan contour near this substituent. It means that hydrogen bond donor group at the $\mathbf{R_1}$ substituent enhances the PknG inhibitory activity. Obviously, compounds **3b** and **22c** containing structural differences at the $\mathbf{R_1}$ substituent (NH_2 and OCH_2CH_3, respectively) show the large difference of IC_{50} values (0.5 μM and 58 μM, respectively). The rule of NH_2 at the $\mathbf{R_1}$ substituent on the PknG inhibitory activity is revealed in (Figures 3.42). This moiety is important to form a conserved hydrogen bond network with Glu233 (O) and an additional hydrogen bond with Ile292 (O).

3.3) R_2 substituent

The size of $\mathbf{R_2}$ substituent was varied from the smallest size of CH_3 group (compound **4b**) to the bulky size of NH-6-benzo[d][1,3]dioxole (compound **27c**). The steric requirement of $\mathbf{R_2}$ substituent was shown by the CoMSIA steric contour (Figure 3.42a). A green contour presents at the $\mathbf{R_2}$ position near the amido group indicating that this region prefers the bulky substituent for PknG inhibitory activity. It clearly shows by the modification of the $\mathbf{R_2}$ substituent size from the smaller size of CH_3 group to the larger size of cyclobutyl (compounds **4b** – **8b**). IC_{50} of these compounds were decreased from 31 μM to 1.66 μM. Interestingly, when the the bulkier NH-6-benzo[d][1,3]dioxole of compound **27c** was replaced by the smaller cyclopropyl of compound **29c**, the IC_{50} value was slightly changed from 0.01 to 0.02 μM. To clear this point, MD simulations were employed to model binding modes of compounds **27c** and **29c** in the PknG pocket. Binding modes of these compounds are similar (Figures 3.45). Because some moiety of the NH-6-benzo[d][1,3]dioxole of compound **27c** protrudes from the PknG pocket (Figure 3.43), its interactions with surrounding residues are comparable with those of the cyclopropyl group of compound **29c**. They form van der Waals interactions with Ile20, Tyr97, Gly99, Gly100 and Met146. Therefore, IC_{50} values of compounds **27c** and **29c** are slightly changed, although they have the large different size of $\mathbf{R_2}$ substituent. Accordingly, the short and bulky $\mathbf{R_2}$ substituent such as the cyclopropyl, cyclobutyl,

$CHC(CH_3)_2$ would enhance the activity for PknG inhibition of tetrahydrobenzothiophene derivatives.

4) Structural concept to design new and more potent PknG inhibitors

The results derived from 3D-QSAR and MD simulations can be suggested the crucial structural requirement of tetrahydrobenzothiophene derivatives as PknG inhibitors. Benzothiophene core structure played important rule for biological activity suggested from CoMSIA contour maps. An oxygen atom of carbonyl core structure formed hydrogen bond interaction with Val235 (NH backbone). For R_1 position, hydrogen bond donor property of primary amine (NH_2) substituent to from hydrogen bond interactions with Glu233 (O) and Ile292 (O) was obtained. Therefore, primary amine substituent at R_1 position should be used to keep primary amide functional of tetrahydrobenzothiophene derivatives. Bulky substituent suggested by CoMSIA contour map was required to bind hydrogen bond interaction and hydrophobic interactions with PknG at R_2 position. Small hydrogen bond donor substituent should be modified at X position to get more hydrogen bond interactions with Lys181 and Asp293.

3.1.2.4 Binding energy calculations

1) Enthalpy binding free energy

Docking scoring functions, the approximated scoring functions often produce results that poorly correlates with biological data (Witchapong et al., 2010(a); 2014(b); Bermin and Sippl, 201). This study, we obtained poorly correlation between docking scoring with biological activity (q^2_{LOO}=0.25 and RMSE = 1.06, Table 3.25). To increase the correlations between predicted binding affinity and biological activity of tetrahydrobenzothiophene derivatives in PknG binding pocket, post docking procedure based on binding free energy calculations was performed. The best regression of bioactivity versus binding free energy is the basically idea to select good reliable post docking procedure. Best docking conformations derived from Glide XP docking were subjected to energy minimization using AMBER12 program followed by single snapshot binding free energy calculations using MM-PBSA and MM-GBSA methods. The enthalpy binding free energy (ΔH_{pred}) of tetrahydrobenzothiophene derivatives bound in PknG binding pocket were

Table 3.29 Predicted binding free energy of PknG inhibitors derived from MM-PBSA and MM-GBSA methods (continued)

Cpd.	IC_{50} (μM)	$log(1/IC_{50})$	XP	Nrot	ΔG_{exp}	ΔG_{pred}					
						MM-PBSA	igb1	igb2	igb5	igb7	igb8
7b	4.79	5.32	-6.77	4	-5.84	-15.35	-47.57	-44.45	-46.53	-44.24	-42.47
8b	1.66	5.78	-6.83	4	-5.78	-14.36	-47.92	-44.85	-46.57	-43.77	-41.84
9b	56	4.25	-6.60	4	-8.77	-9.81	-44.04	-40.65	-42.65	-39.76	-38.70
10b	62	4.21	-6.54	4	-6.30	-12.96	-45.91	-43.00	-45.29	-42.71	-40.90
11b	0.41	6.39	-6.84	4	-6.40	-12.67	-47.40	-44.11	-46.21	-42.98	-41.30
12b	26	4.59	-6.70	4	-8.52	-12.94	-46.46	-42.82	-45.19	-42.65	-41.60
13b	22	4.66	-5.24	5	-5.72	-4.56	-43.32	-39.95	-41.67	-37.08	-35.23
14b	0.63	6.20	-7.87	4	-8.42	-19.10	-50.66	-47.74	-51.05	-44.79	-44.93
15b	68	4.17	-5.99	4	-6.25	-11.09	-46.98	-44.67	-49.83	-41.37	-39.64
16b	0.74	6.13	-6.79	4	-7.55	-15.98	-48.21	-44.40	-46.08	-46.08	-44.57
17b	28	4.55	-6.63	4	-8.67	-21.30	-51.80	-47.40	-49.25	-46.25	-44.68
18b	0.68	6.17	-6.30	5	-8.47	-17.43	-53.13	-49.32	-51.49	-47.46	-45.90
19b	0.085	7.07	-6.74	4	-9.71	-17.92	-47.39	-44.23	-46.69	-44.93	-42.20
20b	58	4.24	-6.27	6	-5.82	0.89	-39.68	-35.96	-37.34	-34.94	-34.23
21c	0.095	7.02	-6.40	4	-9.64	-11.83	-49.21	-46.39	-48.19	-47.52	-45.08
22c	0.093	7.03	-6.01	5	-9.66	-17.44	-51.04	-47.52	-49.59	-47.16	-46.95
23c	0.17	6.77	-6.49	4	-9.30	-22.17	-50.91	-47.50	-50.60	-47.87	-44.31
24c	3.16	5.50	-7.31	4	-10.06	-15.54	-56.86	-52.05	-54.64	-46.16	-46.05
25c	0.49	6.31	-7.18	4	-10.99	-20.82	-51.27	-46.85	-48.09	-45.02	-46.77
26c	0.047	7.33	-7.28	4	-10.03	-24.76	-53.50	-48.98	-51.55	-46.92	-46.75
27c	0.01	8.00	-7.79	5	-10.57	-18.76	-55.24	-50.49	-53.49	-52.03	-51.15
28c	0.05	7.30	-8.31	6	-10.16	-26.81	-56.95	-51.93	-54.74	-49.68	-49.85
29c	0.02	7.70	-7.18	4	-10.33	-19.38	-56.12	-51.97	-53.98	-51.27	-50.74
30c	0.04	7.40	-7.04	5	-5.08	-18.57	-60.63	-55.93	-57.26	-52.46	-51.78
31c	0.03	7.52	-7.32	5	-5.08	-25.32	-56.81	-52.24	-54.93	-51.50	-50.81
In1	>100	<5.49	-5.02	6	-5.08	11.89	-34.59	-31.09	-32.80	-28.44	-39.21
In2	>100	<5.49	-2.65	4	-5.08	21.60	-27.72	-25.31	-29.03	-22.31	-25.16
In3	>100	<5.49	-5.82	4	-5.08	-7.21	-39.70	-36.45	-38.23	-35.66	-42.08
In4	>100	<5.49	-6.28	4	-5.08	-3.18	-37.08	-33.36	-35.30	-32.70	-39.96
In5	>100	<5.49	-6.44	4	-5.08	-9.10	-41.02	-36.76	-38.99	-35.30	-42.01
In6	>100	<5.49	-5.87	5	-5.08	-6.87	-39.99	-37.03	-40.04	-38.81	-44.21
In7	>100	<5.49	-5.41	7	-8.65	8.04	-33.58	-30.15	-31.50	-27.94	-40.87
In8	>100	<5.49	-5.84	4	-6.19	8.49	-27.85	-26.05	-27.43	-25.73	-30.19

Table 3.30 Summary of the statistical data obtained for the binding free energy calculations

Method	igb	R^2	RMSE	q^2_{LOO}	Medium value of active	Medium value of inactive
MM-GBSA	1	0.45	0.99	0.40	-48.21	-35.84
	2	0.45	0.99	0.41	-45.89	-32.23
	5	0.44	1.00	0.39	-47.82	-34.05
	7	0.52	0.93	0.47	-44.93	-30.57
	8	0.52	0.93	0.46	-42.49	-40.42
MM-PBSA	-	0.40	1.04	0.33	-	-

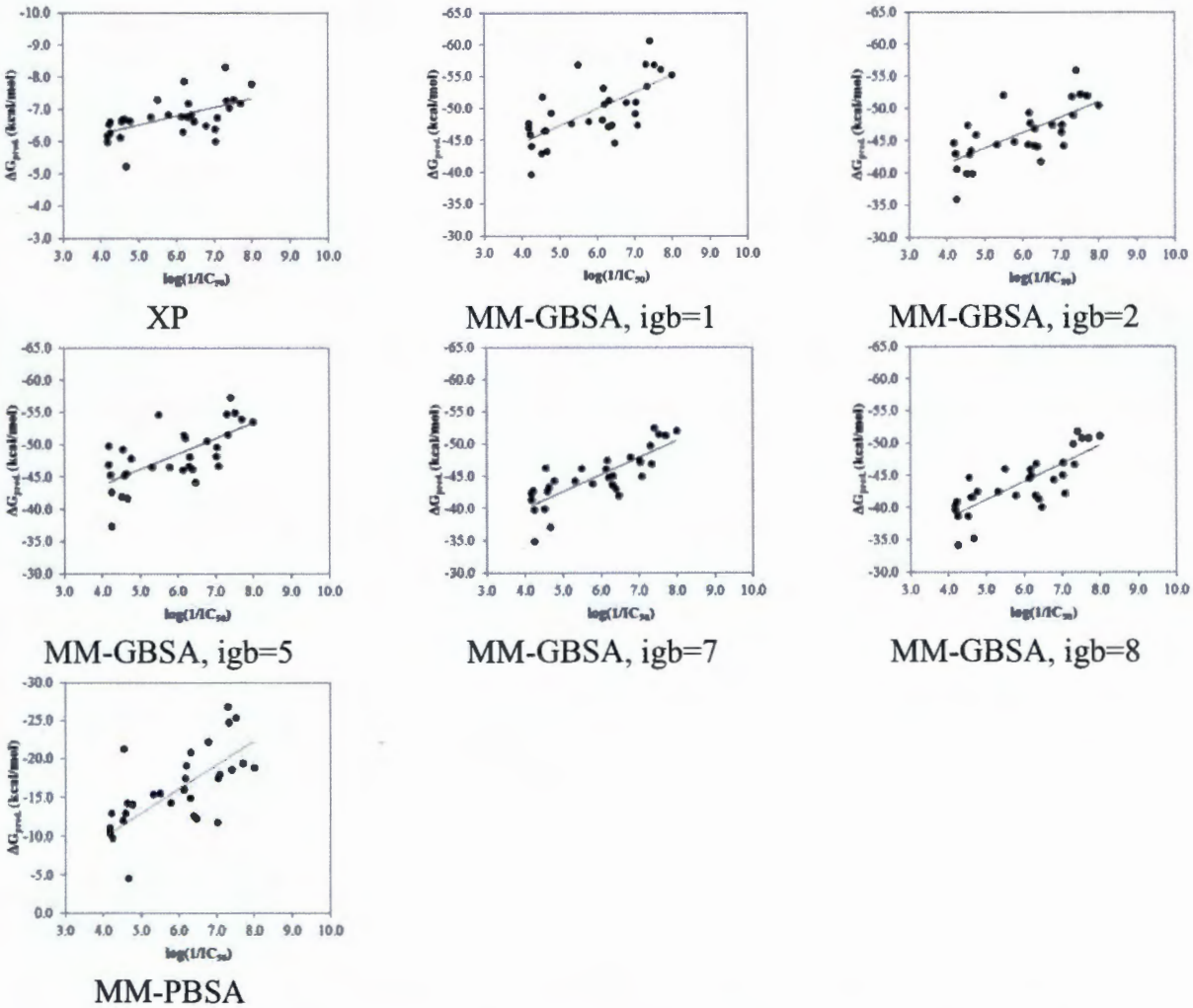


Figure 3.50 Correlation between experimental biological activity values and binding free energy (ΔG_{pred}) derived from MM-PBSA and MM-GBSA methods

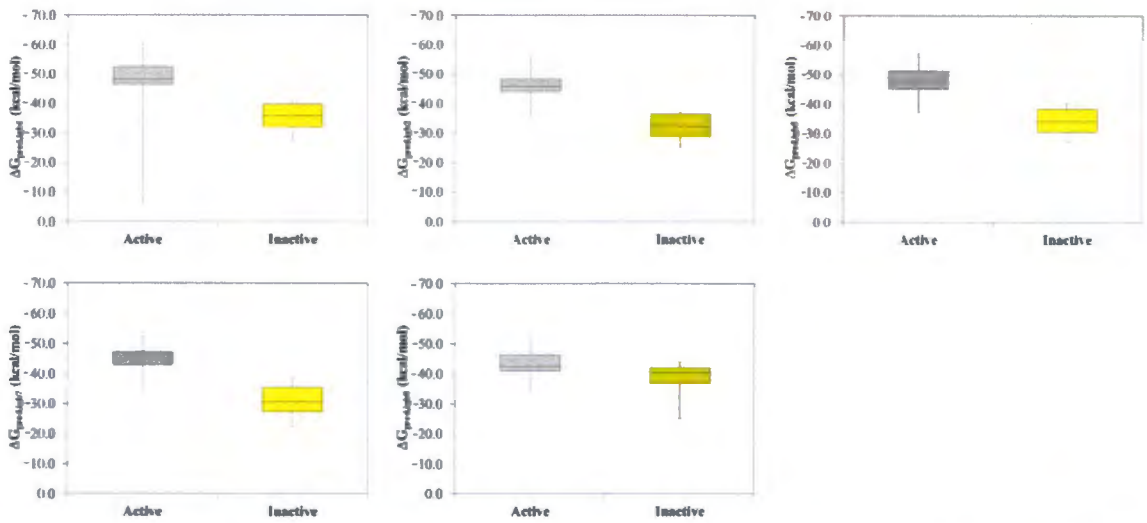
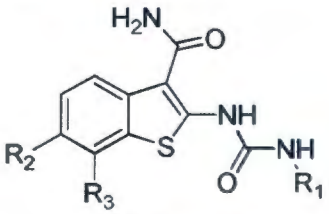


Figure 3.51 Box plot of the binding free energy calculations distribution obtained from active and inactive compounds

3.1.2.5 Designing of benzothiophene derivatives as PknG inhibitors

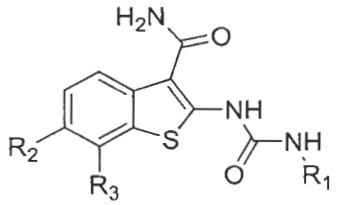
Based on the structural requirement derived from 3D-QSAR CoMSIA and crucial interaction derived from MD simulations, 479 benzothiophene compounds were designed. The structure of **27c** was selected as template. R_1 position was modified to search the optimal bulky substituent. R_2 was modified by small hydrophilic and electron withdrawing groups. Small hydrophobic substituents were selected to modify at R_3 position.

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives



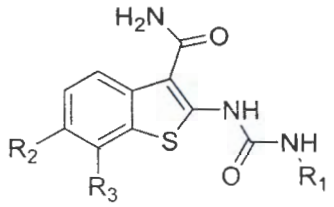
Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD001	5-indanyl	OH	H	3.61
PD002	**5-indanyl**	**OH**	**Cl**	**9.18**
PD003	**3,4-diMe-Ph**	**OH**	**Cl**	**7.98**
PD004	2,4-diMe-Ph	OH	Cl	6.53

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



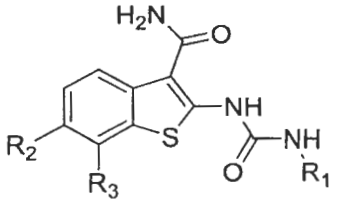
Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD005	2,4,6-triMe-Ph	OH	Cl	7.63
PD006	2,6-diMe-Ph	OH	Cl	7.62
PD007	2-Me-Ph	OH	Cl	6.94
PD008	2,3-diMe-Ph	OH	Cl	6.85
PD009	**3-Me-Ph**	**OH**	**Cl**	**8.03**
PD010	3,5-diMe-Ph	OH	Cl	5.44
PD011	**2-Me-Ph**	**OH**	**Cl**	**8.31**
PD012	**Ph**	**OH**	**Cl**	**8.03**
PD013	2-OMe-Ph	OH	Cl	6.48
PD014	**3-OMe-Ph**	**OH**	**Cl**	**8.16**
PD015	4-OMe-Ph	OH	Cl	7.93
PD016	2,3-diOMe-Ph	OH	Cl	6.72
PD017	2,4-diOMe-Ph	OH	Cl	6.43
PD018	2,5-diOMePh	OH	Cl	5.88
PD019	**2,6-diOMe-Ph**	**OH**	**Cl**	**8.74**
PD020	**3,4-diOMe-Ph**	**OH**	**Cl**	**8.32**
PD021	**3,5-diOMe-Ph**	**OH**	**Cl**	**8.87**
PD022	3-OMe,4-OH-Ph	OH	Cl	7.07
PD023	3,4,5-triOMe-Ph	OH	Cl	7.79
PD024	2,5-diMe-Ph	OH	Cl	7.64
PD025	2-OH-Ph	OH	Cl	7.76
PD026	3-OH-Ph	OH	Cl	7.89
PD027	4-OH-Ph	OH	Cl	7.97
PD028	2-NH_2-Ph	OH	Cl	6.78
PD029	3-NH_2-Ph	OH	Cl	7.79
PD030	**4-NH_2-Ph**	**OH**	**Cl**	**8.11**

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



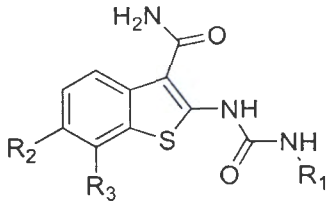
Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD031	2-F-Ph	OH	Cl	7.52
PD032	3-F-Ph	OH	Cl	7.88
PD033	4-F-Ph	OH	Cl	7.61
PD034	2-Cl-Ph	OH	Cl	7.26
PD035	**3-Cl-Ph**	**OH**	**Cl**	**8.20**
PD036	4-Cl-Ph	OH	Cl	7.93
PD037	2-Br-Ph	OH	Cl	6.96
PD038	3-Br-Ph	OH	Cl	7.79
PD039	4-Br-Ph	OH	Cl	7.92
PD040	2-I-Ph	OH	Cl	6.83
PD041	**3-I-Ph**	**OH**	**Cl**	**8.53**
PD042	4-I-Ph	OH	Cl	7.87
PD043	2-CN-Ph	OH	Cl	6.98
PD044	3-CN-Ph	OH	Cl	7.88
PD045	4-CN-Ph	OH	Cl	7.89
PD046	2-Et-Ph	OH	Cl	7.44
PD047	**3-Et-Ph**	**OH**	**Cl**	**8.05**
PD048	4-Et-Ph	OH	Cl	7.98
PD049	2-CF_3-Ph	OH	Cl	7.35
PD050	**3-CF_3-Ph**	**OH**	**Cl**	**8.30**
PD051	4-CF_3-Ph	OH	Cl	7.79
PD052	2-Cyclopropyl-Ph	**OH**	**Cl**	**8.23**
PD053	3-Cyclopropyl-Ph	OH	Cl	3.65
PD054	4-Cyclopropyl-Ph	OH	Cl	7.91
PD055	**2-NO_2-Ph**	**OH**	**Cl**	**8.27**
PD056	3-NO_2-Ph	OH	Cl	7.77

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



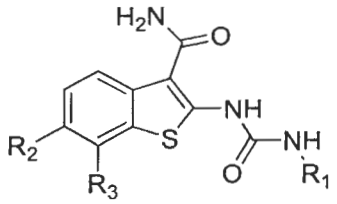
Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD057	4-NO_2-Ph	OH	Cl	7.16
PD058	2-SO_2NH_2-Ph	OH	Cl	7.58
PD059	3-SO_2NH_2-Ph	OH	Cl	7.97
PD060	4-SO_2NH_2-Ph	OH	Cl	7.90
PD061	**2,3-diF-Ph**	**OH**	**Cl**	**8.19**
PD062	2,4-diF-Ph	OH	Cl	7.69
PD063	2,5-diF-Ph	OH	Cl	6.94
PD064	2,6-diF-Ph	OH	Cl	7.81
PD065	3,4-diF-Ph	OH	Cl	7.91
PD066	**3,5-diF-Ph**	**OH**	**Cl**	**8.22**
PD067	2,3-diCl-Ph	OH	Cl	7.32
PD068	2,4-diCl-Ph	OH	Cl	6.95
PD069	2,5-diCl-Ph	OH	Cl	7.09
PD070	2,6-diCl-Ph	OH	Cl	5.83
PD071	3,4-diCl-Ph	OH	Cl	7.85
PD072	**3,5-diCl-Ph**	**OH**	**Cl**	**8.12**
PD073	2,3-diBr-Ph	OH	Cl	6.81
PD074	2,4-diBr-Ph	OH	Cl	6.98
PD075	2,5-diBr-Ph	OH	Cl	7.06
PD076	2,6-diBr-Ph	OH	Cl	6.01
PD077	3,4-diBr-Ph	OH	Cl	7.88
PD078	3,5-diBr-Ph	**OH**	**Cl**	**8.25**
PD079	2,3-diI-Ph	OH	Cl	6.81
PD080	2,4-diI-Ph	OH	Cl	6.92
PD081	2,5-diI-Ph	OH	Cl	6.86
PD082	2,6-diI-Ph	OH	Cl	7.32

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



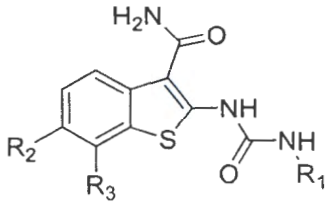
Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD083	**3,4-diI-Ph**	**OH**	**Cl**	**8.50**
PD084	**3,5-diI-Ph**	**OH**	**Cl**	**8.45**
PD085	2-F,3-Cl-Ph	OH	Cl	6.55
PD086	2-F,4-Cl-Ph	OH	Cl	7.03
PD087	2-F,5-Cl-Ph	OH	Cl	6.53
PD088	2-F,6-Cl-Ph	OH	Cl	6.87
PD089	2-CO_2H-Ph	OH	Cl	6.44
PD090	**3-CO_2H-Ph**	**OH**	**Cl**	**9.26**
PD091	**4-CO_2H-Ph**	**OH**	**Cl**	**8.41**
PD092	2-$CONH_2$-Ph	OH	Cl	7.42
PD093	3-$CONH_2$-Ph	OH	Cl	7.93
PD094	4-$CONH_2$-Ph	OH	Cl	7.80
PD095	**2,3-Dihydro-1,4-benzodioxin-6-yl**	**OH**	**Cl**	**9.30**
PD096	2-Pyridyl	OH	Cl	7.46
PD097	3-Pyridyl	OH	Cl	7.39
PD098	4-Pyridyl	OH	Cl	7.89
PD099	**Pyrimidin-4-yl**	**OH**	**Cl**	**8.84**
PD100	Pyrazin-2-yl	OH	Cl	7.19
PD101	Pyrimidin-2-yl	OH	Cl	7.69
PD102	1,3,5-Triazin-2-yl	OH	Cl	7.44
PD103	**2-furanyl**	**OH**	**Cl**	**8.01**
PD104	**2-thiophenyl**	**OH**	**Cl**	**8.02**
PD105	**2-pyrrolyl**	**OH**	**Cl**	**8.66**
PD106	**Oxazol-5-yl**	**OH**	**Cl**	**8.66**
PD107	Thiazol-5-yl	**OH**	**Cl**	**8.42**
PD108	1H-imidazol-5-yl	OH	Cl	7.93

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



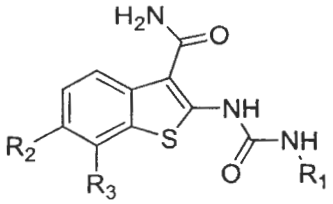
Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD109	1H-triazol-5-yl	OH	Cl	7.45
PD110	1H-tetrazol-5-yl	OH	Cl	7.60
PD111	Benzofuran-6-yl	OH	Cl	7.58
PD112	Benzothiophene-6-yl	OH	Cl	7.68
PD113	Indole-6-yl	OH	Cl	7.85
PD114	1,3-benzoxazole-6-yl	OH	Cl	7.74
PD115	1,3-benzothiazole-6-yl	OH	Cl	7.58
PD116	Benzimidazole-6-yl	OH	Cl	7.54
PD117	1H-benzotriazole-6-yl	OH	Cl	7.59
PD118	**3-OH-isoxazol-5-yl**	**OH**	**Cl**	**10.42**
PD119	3-OH-isothiazol-5-yl	OH	Cl	7.64
PD120	5-Me-isoxazol-4-yl	OH	Cl	7.04
PD121	1-Me, 3-OH-1*H*-pyrazol-5-yl	OH	Cl	6.59
PD122	**3-Me-isoxazol-5-yl**	**OH**	**Cl**	**8.61**
PD123	3-Me-isothiazol-5-yl	OH	Cl	7.77
PD124	5-Me-isoxazol-4-yl	OH	Cl	7.77
PD125	1,3-diMe -1*H*-pyrazol-5-yl	OH	Cl	7.93
PD126	Oxazolidine-2,4-dione-5-yl	OH	Cl	7.21
PD127	Imidazolidine-2,4-dione-5-yl	OH	Cl	7.82
PD128	2,4-Dioxo-1,3-thiazolidin-5-yl	OH	Cl	6.98
PD129	2-Thioxo-1,3-thiazolidin-4-one	OH	Cl	7.23
PD130	4-OH-1,2,5-oxadiazol-3-yl	OH	Cl	7.82
PD131	4-OH-1,2,5-thiadiazol-3-yl	OH	Cl	7.74
PD132	5-OH-2H-triazol-4-yl	OH	Cl	7.37
PD133	2-Me, 2H-tetrazol-5-yl	OH	Cl	7.46
PD134	4-OMe-1,2,5-oxadiazol-3-yl	OH	Cl	6.82

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



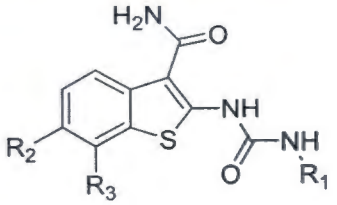
Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD135	4-OMe-1,2,5-thiadiazol-3-yl	OH	Cl	6.81
PD136	5-OMe-2H-triazol-4-yl	OH	Cl	7.66
PD137	2-Et, 2H-tetrazol-5-yl	OH	Cl	7.66
PD138	**4-OEt-1,2,5-oxadiazol-3-yl**	**OH**	**Cl**	**8.41**
PD139	4-OEt-1,2,5-thiadiazol-3-yl	OH	Cl	6.65
PD140	**5-OEt-2H-triazol-4-yl**	**OH**	**Cl**	**8.40**
PD141	5-OH,2-Me-2H-triazol-4-yl	OH	Cl	7.33
PD142	5-OMe,2-Me-2H-triazol-4-yl	OH	Cl	7.42
PD143	5-OEt,2-Me-2H-triazol-4-yl	OH	Cl	7.25
PD144	5-OH,2-Et-2H-triazol-4-yl	OH	Cl	6.83
PD145	5-OMe,2-Et-2H-triazol-4-yl	OH	Cl	7.00
PD146	5-OEt,2-Et-2H-triazol-4-yl	OH	Cl	7.50
PD147	Benzimidazol-1-yl	OH	Cl	6.55
PD148	6-F-Pyridin-2-yl	OH	Cl	7.42
PD149	6-Cl-Pyridin-2-yl	OH	Cl	7.86
PD150	6-Br-Pyridin-2-yl	OH	Cl	7.58
PD151	**6-I-Pyridin-2-yl**	**OH**	**Cl**	**8.16**
PD152	6-Me-Pyridin-2-yl	OH	Cl	7.50
PD153	**6-OH-Pyridin-2-yl**	**OH**	**Cl**	**8.01**
PD154	6-NH_2-Pyridin-2-yl	OH	Cl	7.45
PD155	6-OMe-Pyridin-2-yl	OH	Cl	7.59
PD156	6-Et-Pyridin-2-yl	OH	Cl	7.99
PD157	6-Cyclopropyl-Pyridin-2-yl	OH	Cl	7.68
PD158	6-CN-Pyridin-2-yl	OH	Cl	7.74
PD159	6-NO_2-Pyridin-2-yl	OH	Cl	7.49

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



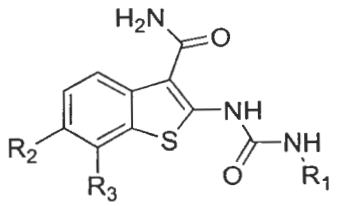
Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD160	6-CF_3-Pyridin-2-yl	**OH**	**Cl**	**8.10**
PD161	6-SO_2NH_2-Pyridin-2-yl	OH	Cl	7.58
PD162	6-CO_2H-Pyridin-2-yl	OH	Cl	6.72
PD163	6-$CONH_2$-Pyridin-2-yl	OH	Cl	7.86
PD164	2-F-Pyridin-3-yl	OH	Cl	7.46
PD165	2-Cl-Pyridin-3-yl	OH	Cl	7.06
PD166	2-Br-Pyridin-3-yl	OH	Cl	6.98
PD167	2-I-Pyridin-3-yl	OH	Cl	7.68
PD168	2-Me-Pyridin-3-yl	OH	Cl	7.65
PD169	2-OH-Pyridin-3-yl	OH	Cl	6.93
PD170	2-NH_2-Pyridin-3-yl	OH	Cl	5.99
PD171	2-OMe-Pyridin-3-yl	OH	Cl	6.68
PD172	2-Et-Pyridin-3-yl	OH	Cl	7.09
PD173	2-Cyclopropyl-Pyridin-3-yl	OH	Cl	7.32
PD174	2-CN-Pyridin-3-yl	OH	Cl	7.19
PD175	2-NO_2-Pyridin-3-yl	OH	Cl	7.94
PD176	2-CF_3-Pyridin-3-yl	OH	Cl	7.43
PD177	2-SO_2NH_2-Pyridin-3-yl	OH	Cl	7.16
PD178	2-CO_2H-Pyridin-3-yl	OH	Cl	7.11
PD179	2-$CONH_2$-Pyridin-3-yl	OH	Cl	7.21
PD180	**3-F-Pyridin-4-yl**	**OH**	**Cl**	**8.87**
PD181	3-Cl-Pyridin-4-yl	OH	Cl	7.17
PD182	3-Br-Pyridin-4-yl	OH	Cl	7.25
PD183	3-I-Pyridin-4-yl	OH	Cl	7.17
PD184	**3-Me-Pyridin-4-yl**	**OH**	**Cl**	**8.27**

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



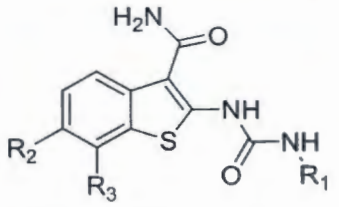
Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD185	3-OH-Pyridin-4-yl	OH	Cl	7.59
PD186	3-NH_2-Pyridin-4-yl	OH	Cl	7.06
PD187	3-OMe-Pyridin-4-yl	OH	Cl	7.00
PD188	3-Et-Pyridin-4-yl	OH	Cl	6.19
PD189	3-Cyclopropyl-Pyridin-4-yl	OH	Cl	7.45
PD190	3-CN-Pyridin-4-yl	**OH**	**Cl**	**8.58**
PD191	3-NO_2-Pyridin-4-yl	**OH**	**Cl**	**8.24**
PD192	3-CF_3-Pyridin-4-yl	OH	Cl	7.44
PD193	3-SO_2NH_2-Pyridin-4-yl	OH	Cl	7.42
PD194	3-CO_2H-Pyridin-4-yl	OH	Cl	7.82
PD195	3-$CONH_2$-Pyridin-4-yl	OH	Cl	7.73
PD196	5-F-Pyridin-2-yl	OH	Cl	7.92
PD197	5-Cl-Pyridin-2-yl	OH	Cl	7.52
PD198	5-Br-Pyridin-2-yl	OH	Cl	7.51
PD199	5-I-Pyridin-2-yl	OH	Cl	7.64
PD200	**5-Me-Pyridin-2-yl**	**OH**	**Cl**	**9.22**
PD201	**5-OH-Pyridin-2-yl**	**OH**	**Cl**	**8.94**
PD202	5-NH_2-Pyridin-2-yl	OH	Cl	7.96
PD203	5-OMe-Pyridin-2-yl	OH	Cl	7.67
PD204	5-Et-Pyridin-2-yl	OH	Cl	7.70
PD205	5-Cyclopropyl-Pyridin-2-yl	OH	Cl	7.66
PD206	5-CN-Pyridin-2-yl	OH	Cl	7.39
PD207	5-NO_2-Pyridin-2-yl	OH	Cl	6.88
PD208	5-CF_3-Pyridin-2-yl	OH	Cl	7.70
PD209	5-SO_2NH_2-Pyridin-2-yl	OH	Cl	7.47

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



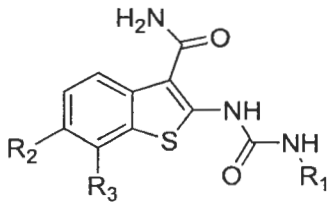
Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD210	**5-CO_2H-Pyridin-2-yl**	**OH**	**Cl**	**8.08**
PD211	5-$CONH_2$-Pyridin-2-yl	OH	Cl	7.55
PD212	6-F-Pyridin-2-yl	OH	Cl	7.82
PD213	6-Cl-Pyridin-2-yl	OH	Cl	7.31
PD214	6-Br-Pyridin-2-yl	OH	Cl	7.51
PD215	**6-I-Pyridin-2-yl**	**OH**	**Cl**	**8.39**
PD216	6-Me-Pyridin-2-yl	OH	Cl	6.53
PD217	6-OH-Pyridin-2-yl	OH	Cl	7.29
PD218	6-NH_2-Pyridin-2-yl	OH	Cl	6.55
PD219	6-OMe-Pyridin-2-yl	OH	Cl	7.72
PD220	6-Et-Pyridin-2-yl	OH	Cl	7.52
PD221	6-Cyclopropyl-Pyridin-2-yl	OH	Cl	7.78
PD222	6-CN-Pyridin-2-yl	OH	Cl	6.51
PD223	6-NO_2-Pyridin-2-yl	OH	Cl	7.38
PD224	**6-CF_3-Pyridin-2-yl**	**OH**	**Cl**	**8.28**
PD225	6-SO_2NH_2-Pyridin-2-yl	OH	Cl	7.61
PD226	6-CO2H-Pyridin-2-yl	OH	Cl	7.39
PD227	6-CONH2-Pyridin-2-yl	OH	Cl	7.55
PD228	3-F-Pyridin-2-yl	OH	Cl	7.40
PD229	3-Cl-Pyridin-2-yl	OH	Cl	7.47
PD230	3-Br-Pyridin-2-yl	OH	Cl	7.22
PD231	3-I-Pyridin-2-yl	OH	Cl	7.32
PD232	3-Me-Pyridin-2-yl	OH	Cl	7.27
PD233	3-OH-Pyridin-2-yl	OH	Cl	5.34
PD234	3-NH_2-Pyridin-2-yl	OH	Cl	7.27
PD235	3-OMe-Pyridin-2-yl	OH	Cl	7.05

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



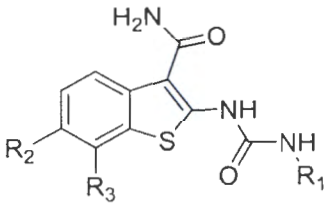
Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD236	3-Et-Pyridin-2-yl	OH	Cl	7.42
PD237	3-Cyclopropyl-Pyridin-2-yl	OH	Cl	6.22
PD238	3-CN-Pyridin-2-yl	OH	Cl	7.48
PD239	3-NO_2-Pyridin-2-yl	OH	Cl	4.25
PD240	**3-CF_3-Pyridin-2-yl**	**OH**	**Cl**	**10.39**
PD241	3-SO_2NH_2-Pyridin-2-yl	OH	Cl	7.67
PD242	3-CO_2H-Pyridin-2-yl	OH	Cl	7.67
PD243	3-$CONH_2$-Pyridin-2-yl	OH	Cl	7.79
PD244	**6-F-Pyridin-3-yl**	**OH**	**Cl**	**8.41**
PD245	**6-Cl-Pyridin-3-yl**	**OH**	**Cl**	**8.06**
PD246	6-Br-Pyridin-3-yl	OH	Cl	7.27
PD247	6-I-Pyridin-3-yl	OH	Cl	6.66
PD248	**6-Me-Pyridin-3-yl**	**OH**	**Cl**	**8.42**
PD249	**6-OH-Pyridin-3-yl**	**OH**	**Cl**	**6.77**
PD250	**6-NH_2-Pyridin-3-yl**	**OH**	**Cl**	**8.60**
PD251	6-OMe-Pyridin-3-yl	**OH**	**Cl**	**8.50**
PD252	**6-Et-Pyridin-3-yl**	**OH**	**Cl**	**8.48**
PD253	**6-Cyclopropyl-Pyridin-3-yl**	**OH**	**Cl**	**8.43**
PD254	**6-CN-Pyridin-3-yl**	**OH**	**Cl**	**8.73**
PD255	6-NO_2-Pyridin-3-yl	OH	Cl	7.08
PD256	**6-CF_3-Pyridin-3-yl**	**OH**	**Cl**	**8.27**
PD257	**6-SO_2NH_2-Pyridin-3-yl**	**OH**	**Cl**	**9.34**
PD258	**6-CO_2H-Pyridin-3-yl**	**OH**	**Cl**	**8.19**
PD259	**6-$CONH_2$-Pyridin-3-yl**	**OH**	**Cl**	**9.34**
PD260	5-F-Pyridin-3-yl	OH	Cl	7.34
PD261	5-Cl-Pyridin-3-yl	OH	Cl	7.34

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



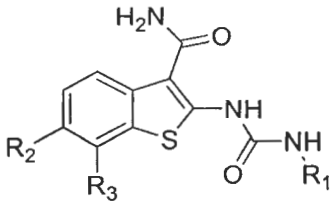
Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD262	**5-Br-Pyridin-3-yl**	**OH**	**Cl**	**8.82**
PD263	**5-I-Pyridin-3-yl**	**OH**	**Cl**	**9.00**
PD264	5-Me-Pyridin-3-yl	OH	Cl	7.72
PD265	5-OH-Pyridin-3-yl	OH	Cl	7.42
PD266	5-NH_2-Pyridin-3-yl	OH	Cl	7.47
PD267	**5-OMe-Pyridin-3-yl**	**OH**	**Cl**	**8.13**
PD268	**5-Et-Pyridin-3-yl**	**OH**	**Cl**	**8.54**
PD269	**5-Cyclopropyl-Pyridin-3-yl**	**OH**	**Cl**	**9.45**
PD270	**5-CN-Pyridin-3-yl**	**OH**	**Cl**	**8.35**
PD271	**5-NO_2-Pyridin-3-yl**	**OH**	**Cl**	**8.19**
PD272	5-CF_3-Pyridin-3-yl	OH	Cl	7.42
PD273	5-SO_2NH_2-Pyridin-3-yl	OH	Cl	7.82
PD274	**5-CO_2H-Pyridin-3-yl**	**OH**	**Cl**	**8.92**
PD275	5-$CONH_2$-Pyridin-3-yl	OH	Cl	7.82
PD276	**4-F-Pyridin-3-yl**	**OH**	**Cl**	**8.16**
PD277	4-Cl-Pyridin-3-yl	OH	Cl	7.31
PD278	4-Br-Pyridin-3-yl	OH	Cl	7.59
PD279	**4-I-Pyridin-3-yl**	**OH**	**Cl**	**8.96**
PD280	4-Me-Pyridin-3-yl	OH	Cl	7.07
PD281	4-OH-Pyridin-3-yl	OH	Cl	6.72
PD282	4-NH_2-Pyridin-3-yl	OH	Cl	7.32
PD283	4-OMe-Pyridin-3-yl	OH	Cl	7.39
PD284	4-Et-Pyridin-3-yl	OH	Cl	7.15
PD285	4-Cyclopropyl-Pyridin-3-yl	OH	Cl	7.48
PD286	4-CN-Pyridin-3-yl	OH	Cl	7.49
PD287	**4-NO_2-Pyridin-3-yl**	**OH**	**Cl**	**8.30**

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



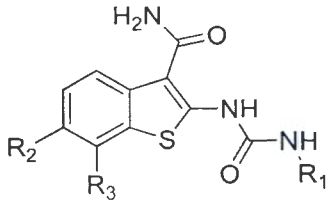
Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD288	**4-CF_3-Pyridin-3-yl**	**OH**	**Cl**	**9.52**
PD289	4-SO_2NH_2-Pyridin-3-yl	OH	Cl	5.25
PD290	4-CO_2H-Pyridin-3-yl	OH	Cl	7.86
PD291	4-$CONH_2$-Pyridin-3-yl	OH	Cl	7.57
PD292	**2-F-Pyridin-4-yl**	**OH**	**Cl**	**8.40**
PD293	2-Cl-Pyridin-4-yl	OH	Cl	7.67
PD294	**2-Br-Pyridin-4-yl**	**OH**	**Cl**	**8.48**
PD295	**2-I-Pyridin-4-yl**	**OH**	**Cl**	**8.43**
PD296	**2-Me-Pyridin-4-yl**	**OH**	**Cl**	**8.29**
PD297	2-OH-Pyridin-4-yl	OH	Cl	7.31
PD298	2-F-Pyridin-4-yl	OH	Cl	7.16
PD299	**2-OMe-Pyridin-4-yl**	**OH**	**Cl**	**8.03**
PD300	2-Et-Pyridin-4-yl	OH	Cl	7.72
PD301	2-Cyclopropyl-Pyridin-4-yl	OH	Cl	7.55
PD302	**2-CN-Pyridin-4-yl**	**OH**	**Cl**	**8.22**
PD303	**2-NO_2-Pyridin-4-yl**	**OH**	**Cl**	**8.15**
PD304	**2-CF_3-Pyridin-4-yl**	**OH**	**Cl**	**8.51**
PD305	2-SO_2NH_2-Pyridin-4-yl	OH	Cl	7.77
PD306	2-CO_2H-Pyridin-4-yl	OH	Cl	7.11
PD307	2-$CONH_2$-Pyridin-4-yl	OH	Cl	7.95
PD308	2-F-Pyrimidin-4-yl	OH	Cl	7.88
PD309	2-Cl-Pyrimidin-4-yl	OH	Cl	7.63
PD310	2-Br-Pyrimidin-4-yl	OH	Cl	7.56
PD311	**2-I-Pyrimidin-4-yl**	**OH**	**Cl**	**8.26**
PD312	**2-Me-Pyrimidin-4-yl**	**OH**	**Cl**	**8.32**
PD313	2-OH-Pyrimidin-4-yl	OH	Cl	7.77

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



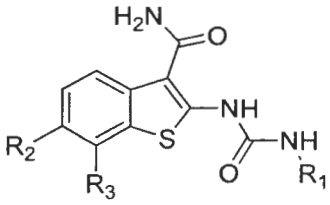
Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD314	2-NH_2-Pyrimidin-4-yl	OH	Cl	6.92
PD315	**2-OMe-Pyrimidin-4-yl**	**OH**	**Cl**	**8.06**
PD316	**2-Et-Pyrimidin-4-yl**	**OH**	**Cl**	**8.15**
PD317	**2-Cyclopropyl-Pyrimidin-4-yl**	**OH**	**Cl**	**8.02**
PD318	2-CN-Pyrimidin-4-yl	OH	Cl	7.86
PD319	2-NO_2-Pyrimidin-4-yl	OH	Cl	7.93
PD320	**2-CF_3-Pyrimidin-4-yl**	**OH**	**Cl**	**9.57**
PD321	2-SO_2NH_2-Pyrimidin-4-yl	OH	Cl	6.82
PD322	**2-CO_2H-Pyrimidin-4-yl**	**OH**	**Cl**	**8.30**
PD323	2-$CONH_2$-Pyrimidin-4-yl	OH	Cl	7.35
PD324	6-F-Pyrimidin-4-yl	OH	Cl	7.70
PD325	**6-Cl-Pyrimidin-4-yl**	**OH**	**Cl**	**8.08**
PD326	6-Br-Pyrimidin-4-yl	OH	Cl	7.51
PD327	6-I-Pyrimidin-4-yl	OH	Cl	7.66
PD328	**6-Me-Pyrimidin-4-yl**	**OH**	**Cl**	**8.00**
PD329	6-OH-Pyrimidin-4-yl	OH	Cl	7.35
PD330	6-NH_2-Pyrimidin-4-yl	OH	Cl	7.72
PD331	**6-OMe-Pyrimidin-4-yl**	**OH**	**Cl**	**8.86**
PD332	6-Et-Pyrimidin-4-yl	OH	Cl	7.45
PD333	**6-Cyclopropyl-Pyrimidin-4-yl**	**OH**	**Cl**	**9.29**
PD334	6-CN-Pyrimidin-4-yl	OH	Cl	7.63
PD335	**6-NO_2-Pyrimidin-4-yl**	**OH**	**Cl**	**8.38**
PD336	**6-CF_3-Pyrimidin-4-yl**	**OH**	**Cl**	**8.72**
PD337	6-SO_2NO_2-Pyrimidin-4-yl	OH	Cl	7.35
PD338	6-CO_2H-Pyrimidin-4-yl	OH	Cl	7.75
PD339	**6-$CONH_2$-Pyrimidin-4-yl**	**OH**	**Cl**	**8.91**

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



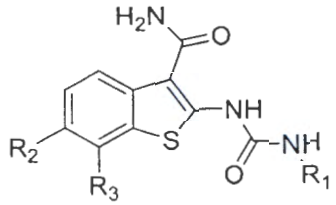
Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD340	5-F-Pyrimidin-4-yl	OH	Cl	7.22
PD341	5-Cl-Pyrimidin-4-yl	OH	Cl	7.12
PD342	5-Br-Pyrimidin-4-yl	OH	Cl	7.00
PD343	5-I-Pyrimidin-4-yl	OH	Cl	7.52
PD344	5-Me-Pyrimidin-4-yl	OH	Cl	7.07
PD345	5-OH-Pyrimidin-4-yl	OH	Cl	5.37
PD346	5-NH_2-Pyrimidin-4-yl	OH	Cl	7.38
PD347	5-OMe-Pyrimidin-4-yl	OH	Cl	7.79
PD348	5-Et-Pyrimidin-4-yl	OH	Cl	7.50
PD349	5-Cyclopropyl-Pyrimidin-4-yl	OH	Cl	7.34
PD350	5-CN-Pyrimidin-4-yl	OH	Cl	7.74
PD351	**5-NO_2-Pyrimidin-4-yl**	**OH**	**Cl**	**8.41**
PD352	**5-CF_3-Pyrimidin-4-yl**	**OH**	**Cl**	**9.46**
PD353	5-SO_2NH_2-Pyrimidin-4-yl	OH	Cl	7.62
PD354	5-CO_2H-Pyrimidin-4-yl	OH	Cl	6.84
PD355	5-$CONH_2$-Pyrimidin-4-yl	OH	Cl	7.46
PD356	6-F-pyrazin-2-yl	OH	Cl	7.81
PD357	**6-Cl-pyrazin-2-yl**	**OH**	**Cl**	**8.41**
PD358	**6-Br-pyrazin-2-yl**	**OH**	**Cl**	**8.68**
PD359	**6-I-pyrazin-2-yl**	**OH**	**Cl**	**8.54**
PD360	6-Me-pyrazin-2-yl	OH	Cl	7.82
PD361	6-OH-pyrazin-2-yl	OH	Cl	7.48
PD362	**6-NH_2-pyrazin-2-yl**	**OH**	**Cl**	**8.12**
PD363	6-OMe-pyrazin-2-yl	OH	Cl	7.56
PD364	**6-Et-pyrazin-2-yl**	**OH**	**Cl**	**8.16**
PD365	**6-Cyclopropyl-pyrazin-2-yl**	**OH**	**Cl**	**8.27**

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



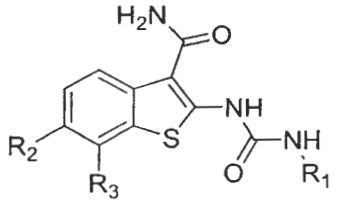
Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD366	**6-CN-pyrazin-2-yl**	**OH**	**Cl**	**8.16**
PD367	**6-NO_2-pyrazin-2-yl**	**OH**	**Cl**	**8.77**
PD368	**6-CF_3-pyrazin-2-yl**	**OH**	**Cl**	**8.58**
PD369	6-SO_2NH_2-pyrazin-2-yl	OH	Cl	6.63
PD370	6-CO_2H-pyrazin-2-yl	OH	Cl	6.89
PD371	6-$CONH_2$-pyrazin-2-yl	OH	Cl	7.50
PD372	5-F-pyrazin-2-yl	OH	Cl	7.64
PD373	5-Cl-pyrazin-2-yl	OH	Cl	7.31
PD374	5-Br-pyrazin-2-yl	OH	Cl	7.93
PD375	**5-I-pyrazin-2-yl**	**OH**	**Cl**	**8.02**
PD376	5-Me-pyrazin-2-yl	OH	Cl	7.91
PD377	5-OH-pyrazin-2-yl	OH	Cl	7.30
PD378	**5-NH_2-pyrazin-2-yl**	**OH**	**Cl**	**8.04**
PD379	**5-OMe-pyrazin-2-yl**	**OH**	**Cl**	**8.00**
PD380	5-Et-pyrazin-2-yl	OH	Cl	3.59
PD381	**5-Cyclopropyl-pyrazin-2-yl**	**OH**	**Cl**	**8.11**
PD382	**5-CN-pyrazin-2-yl**	**OH**	**Cl**	**9.13**
PD383	5-NO_2-pyrazin-2-yl	OH	Cl	7.19
PD384	5-CF_3-pyrazin-2-yl	OH	Cl	7.97
PD385	5-SO_2NH_2-pyrazin-2-yl	OH	Cl	7.97
PD386	5-CO_2H-pyrazin-2-yl	**OH**	**Cl**	**8.29**
PD387	5-$CONH_2$-pyrazin-2-yl	**OH**	**Cl**	**8.01**
PD388	3-F-pyrazin-2-yl	OH	Cl	5.32
PD389	3-Cl-pyrazin-2-yl	OH	Cl	7.41
PD390	3-Br-pyrazin-2-yl	OH	Cl	7.38
PD391	3-I-pyrazin-2-yl	OH	Cl	7.10

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



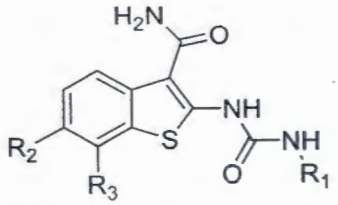
Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD392	3-Me-pyrazin-2-yl	OH	Cl	7.18
PD393	3-OH-pyrazin-2-yl	**OH**	**Cl**	**8.05**
PD394	3-NH_2-pyrazin-2-yl	OH	Cl	7.30
PD395	3-OMe-pyrazin-2-yl	OH	Cl	7.50
PD396	3-Et-pyrazin-2-yl	OH	Cl	7.74
PD397	3-Cyclopropyl-pyrazin-2-yl	OH	Cl	6.74
PD398	3-CN-pyrazin-2-yl	**OH**	**Cl**	**8.80**
PD399	3-NO_2-pyrazin-2-yl	**OH**	**Cl**	**8.46**
PD400	3-CF_3-pyrazin-2-yl	OH	Cl	7.30
PD401	3-SO_2NH_2-pyrazin-2-yl	OH	Cl	5.89
PD402	3-CO_2H-pyrazin-2-yl	OH	Cl	7.29
PD403	3-$CONH_2$-pyrazin-2-yl	OH	Cl	7.45
PD404	4-F-Pyrimidin-2-yl	OH	Cl	7.80
PD405	4-Cl-Pyrimidin-2-yl	OH	Cl	7.81
PD406	**4-Br-Pyrimidin-2-yl**	**OH**	**Cl**	**8.16**
PD407	4-I-Pyrimidin-2-yl	OH	Cl	7.17
PD408	**4-Me-Pyrimidin-2-yl**	**OH**	**Cl**	**8.03**
PD409	**4-OH-Pyrimidin-2-yl**	**OH**	**Cl**	**9.16**
PD410	4-NH_2-Pyrimidin-2-yl	OH	Cl	7.04
PD411	4-OMe-Pyrimidin-2-yl	OH	Cl	7.59
PD412	**4-Et-Pyrimidin-2-yl**	**OH**	**Cl**	**8.00**
PD413	4-Cyclopropyl-Pyrimidin-2-yl	OH	Cl	4.08
PD414	4-CN-Pyrimidin-2-yl	OH	Cl	7.55
PD415	4-NO_2-Pyrimidin-2-yl	OH	Cl	7.59
PD416	**4-CF_3-Pyrimidin-2-yl**	**OH**	**Cl**	**8.21**
PD417	4-SO_2NH_2-Pyrimidin-2-yl	OH	Cl	7.49

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



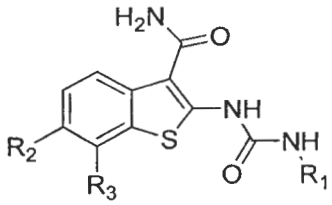
Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD418	4-CO_2H-Pyrimidin-2-yl	OH	Cl	6.71
PD419	4-$CONH_2$-Pyrimidin-2-yl	OH	Cl	7.04
PD420	5-F-Pyrimidin-2-yl	OH	Cl	7.09
PD421	5-Cl-Pyrimidin-2-yl	OH	Cl	6.97
PD422	5-Br-Pyrimidin-2-yl	OH	Cl	7.24
PD423	5-I-Pyrimidin-2-yl	OH	Cl	7.56
PD424	5-Me-Pyrimidin-2-yl	OH	Cl	7.65
PD425	**5-OH-Pyrimidin-2-yl**	**OH**	**Cl**	**8.21**
PD426	**5-NH_2-Pyrimidin-2-yl**	**OH**	**Cl**	**8.82**
PD427	5-OMe-Pyrimidin-2-yl	OH	Cl	7.33
PD428	5-Et-Pyrimidin-2-yl	OH	Cl	7.20
PD429	5-Cyclopropyl-Pyrimidin-2-yl	OH	Cl	4.05
PD430	**5-CN-Pyrimidin-2-yl**	**OH**	**Cl**	**8.66**
PD431	5-NO_2-Pyrimidin-2-yl	OH	Cl	6.94
PD432	5-CF_3-Pyrimidin-2-yl	OH	Cl	7.45
PD433	5-SO_2NH_2-Pyrimidin-2-yl	OH	Cl	6.32
PD434	5-CO_2H-Pyrimidin-2-yl	OH	Cl	7.65
PD435	5-$CONH_2$-Pyrimidin-2-yl	OH	Cl	7.87
PD436	4-F-1,3,5-triazin-2-yl	OH	Cl	7.59
PD437	4-Cl-1,3,5-triazin-2-yl	OH	Cl	7.59
PD438	**4-Br-1,3,5-triazin-2-yl**	**OH**	**Cl**	**8.61**
PD439	**4-I-1,3,5-triazin-2-yl**	**OH**	**Cl**	**9.24**
PD440	**4-Me-1,3,5-triazin-2-yl**	**OH**	**Cl**	**8.31**
PD441	4-OH-1,3,5-triazin-2-yl	OH	Cl	5.55
PD442	4-NH_2-1,3,5-triazin-2-yl	OH	Cl	7.06
PD443	4-OMe-1,3,5-triazin-2-yl	OH	Cl	7.70

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD444	**4-Et-1,3,5-triazin-2-yl**	**OH**	**Cl**	**8.28**
PD445	4-Cyclopropyl-1,3,5-triazin-2-yl	OH	Cl	7.93
PD446	4-CN-1,3,5-triazin-2-yl	**OH**	**Cl**	**8.05**
PD447	4-NO_2-1,3,5-triazin-2-yl	OH	Cl	7.91
PD448	4-CF_3-1,3,5-triazin-2-yl	OH	Cl	6.22
PD449	4-SO_2NH_2-1,3,5-triazin-2-yl	OH	Cl	7.24
PD450	4-CO_2H-1,3,5-triazin-2-yl	OH	Cl	6.70
PD451	4-$CONH_2$-1,3,5-triazin-2-yl	OH	Cl	7.50
PD452	1,3-Benzodioxol-5-yl	OH	H	7.25
PD453	1,3-Benzodioxol-5-yl	OH	F	7.88
PD454	**1,3-Benzodioxol-5-yl**	**OH**	**Br**	**9.06**
PD455	**1,3-Benzodioxol-5-yl**	**OH**	**I**	**8.23**
PD456	1,3-Benzodioxol-5-yl	OH	CH_3	7.06
PD457	1,3-Benzodioxol-5-yl	OH	OCH_3	7.07
PD458	**1,3-Benzodioxol-5-yl**	**OH**	**CH_2CH_3**	**8.48**
PD459	1,3-Benzodioxol-5-yl	OH		6.94
PD460	**1,3-Benzodioxol-5-yl**	**OH**	**CF_3**	**8.21**
PD461	**1,3-Benzodioxol-5-yl**	**F**	**Cl**	**9.18**
PD462	1,3-Benzodioxol-5-yl	Cl	Cl	7.34
PD463	1,3-Benzodioxol-5-yl	CO_2H	Cl	7.99
PD464	1,3-Benzodioxol-5-yl	CN	Cl	7.31
PD465	1,3-Benzodioxol-5-yl	NO_2	Cl	6.99
PD466	**3-OH-isoxazol-5-yl**	**OH**	**H**	**8.63**
PD467	**3-OH-isoxazol-5-yl**	**OH**	**F**	**8.42**
PD468	3-OH-isoxazol-5-yl	OH	Br	7.96
PD469	**3-OH-isoxazol-5-yl**	**OH**	**I**	**9.18**

Table 3.31 Chemical structure and predicted activity of new designed benzothiophene derivatives (continued)



Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD470	**3-OH-isoxazol-5-yl**	**OH**	**CH_3**	**8.01**
PD471	**3-OH-isoxazol-5-yl**	**OH**	**OCH_3**	**8.32**
PD472	**3-OH-isoxazol-5-yl**	**OH**	**CH_2CH_3**	**8.40**
PD473	3-OH-isoxazol-5-yl	OH		7.47
PD474	**3-OH-isoxazol-5-yl**	**OH**	**CF_3**	**9.84**
PD475	**3-OH-isoxazol-5-yl**	**F**	**Cl**	**9.46**
PD476	**3-OH-isoxazol-5-yl**	**Cl**	**Cl**	**9.44**
PD477	**3-OH-isoxazol-5-yl**	**CO_2H**	**Cl**	**8.93**
PD478	**3-OH-isoxazol-5-yl**	**CN**	**Cl**	**8.80**
PD479	**3-OH-isoxazol-5-yl**	**NO_2**	**Cl**	**9.28**

3.1.2.6 Promising benzothiophene derivatives based on rational design

479 compounds of benzothiophene derivatives were designed based on the structural requirements and crucial interaction derived from 3D-QSAR CoMSIA and MD simulations. Based on best CoMSIA model prediction (model 8 in Table 3.23), 145 compounds were obtained with higher predicted biological activity than compound **27c**, the template compounds. To ensure that design compounds bound with PknG binding pocket, molecular docking calculation was perform to predict the binding mode of new designed compounds. The docking results showed that all of new designed compounds can bound in PknG binding pocket. The binding modes of all compounds are same manner with compound **27c**, the template compounds as shown in Figure 3.52.

Table 3.32 Highly predicted activity benzothiophene compounds

Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD002	5-indanyl	OH	Cl	9.18
PD003	3,4-diMe-Ph	OH	Cl	7.98
PD009	3-Me-Ph	OH	Cl	8.03
PD011	2-Me-Ph	OH	Cl	8.31
PD012	Ph	OH	Cl	8.03
PD014	3-OMe-Ph	OH	Cl	8.16
PD019	2,6-diOMe-Ph	OH	Cl	8.74
PD020	3,4-diOMe-Ph	OH	Cl	8.32
PD021	3,5-diOMe-Ph	OH	Cl	8.87
PD030	4-NH_2-Ph	OH	Cl	8.11
PD035	3-Cl-Ph	OH	Cl	8.20
PD041	3-I-Ph	OH	Cl	8.53
PD047	3-Et-Ph	OH	Cl	8.05
PD050	3-CF_3-Ph	OH	Cl	8.30
PD052	2-Cyclopropyl-Ph	OH	Cl	8.23
PD055	2-NO_2-Ph	OH	Cl	8.27
PD061	2,3-diF-Ph	OH	Cl	8.19
PD066	3,5-diF-Ph	OH	Cl	8.22
PD072	3,5-diCl-Ph	OH	Cl	8.12
PD078	3,5-diBr-Ph	OH	Cl	8.25
PD083	3,4-diI-Ph	OH	Cl	8.50
PD084	3,5-diI-Ph	OH	Cl	8.45
PD090	3-CO_2H-Ph	OH	Cl	9.26
PD091	4-CO_2H-Ph	OH	Cl	8.41
PD095	2,3-Dihydro-1,4-benzodioxin-6-yl	OH	Cl	9.30
PD099	Pyrimidin-4-yl	OH	Cl	8.84
PD103	2-furanyl	OH	Cl	8.01
PD104	2-thiophenyl	OH	Cl	8.02
PD105	2-pyrrolyl	OH	Cl	8.66
PD106	Oxazol-5-yl	OH	Cl	8.66
PD107	Thiazol-5-yl	OH	Cl	8.42

Table 3.32 Highly predicted activity benzothiophene compounds (continued)

Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD118	3-OH-isoxazol-5-yl	OH	Cl	10.42
PD122	3-Me-isoxazol-5-yl	OH	Cl	8.61
PD138	4-OEt-1,2,5-oxadiazol-3-yl	OH	Cl	8.41
PD140	5-OEt-2H-triazol-4-yl	OH	Cl	8.40
PD151	6-I-Pyridin-2-yl	OH	Cl	8.16
PD153	6-OH-Pyridin-2-yl	OH	Cl	8.01
PD160	6-CF_3-Pyridin-2-yl	OH	Cl	8.10
PD180	3-F-Pyridin-4-yl	OH	Cl	8.87
PD184	3-Me-Pyridin-4-yl	OH	Cl	8.27
PD190	3-CN-Pyridin-4-yl	OH	Cl	8.58
PD191	3-NO_2-Pyridin-4-yl	OH	Cl	8.24
PD200	5-Me-Pyridin-2-yl	OH	Cl	9.22
PD201	5-OH-Pyridin-2-yl	OH	Cl	8.94
PD210	5-CO_2H-Pyridin-2-yl	OH	Cl	8.08
PD215	6-I-Pyridin-2-yl	OH	Cl	8.39
PD224	6-CF_3-Pyridin-2-yl	OH	Cl	8.28
PD240	3-CF_3-Pyridin-2-yl	OH	Cl	10.39
PD244	6-F-Pyridin-3-yl	OH	Cl	8.41
PD245	6-Cl-Pyridin-3-yl	OH	Cl	8.06
PD248	6-Me-Pyridin-3-yl	OH	Cl	8.42
PD249	6-OH-Pyridin-3-yl	OH	Cl	6.77
PD250	6-NH_2-Pyridin-3-yl	OH	Cl	8.60
PD251	6-OMe-Pyridin-3-yl	OH	Cl	8.50
PD252	6-Et-Pyridin-3-yl	OH	Cl	8.48
PD253	6-Cyclopropyl-Pyridin-3-yl	OH	Cl	8.43
PD254	6-CN-Pyridin-3-yl	OH	Cl	8.73
PD256	6-CF_3-Pyridin-3-yl	OH	Cl	8.27
PD257	6-SO_2NH_2-Pyridin-3-yl	OH	Cl	9.34
PD258	6-CO_2H-Pyridin-3-yl	OH	Cl	8.19
PD259	6-CONH2-Pyridin-3-yl	OH	Cl	9.34

Table 3.32 Highly predicted activity benzothiophene compounds (continued)

Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD262	5-Br-Pyridin-3-yl	OH	Cl	8.82
PD263	5-I-Pyridin-3-yl	OH	Cl	9.00
PD267	5-OMe-Pyridin-3-yl	OH	Cl	8.13
PD268	5-Et-Pyridin-3-yl	OH	Cl	8.54
PD269	5-Cyclopropyl-Pyridin-3-yl	OH	Cl	9.45
PD270	5-CN-Pyridin-3-yl	OH	Cl	8.35
PD271	5-NO_2-Pyridin-3-yl	OH	Cl	8.19
PD274	5-CO_2H-Pyridin-3-yl	OH	Cl	8.92
PD276	4-F-Pyridin-3-yl	OH	Cl	8.16
PD279	4-I-Pyridin-3-yl	OH	Cl	8.96
PD287	4-NO_2-Pyridin-3-yl	OH	Cl	8.30
PD288	4-CF_3-Pyridin-3-yl	OH	Cl	9.52
PD292	2-F-Pyridin-4-yl	OH	Cl	8.40
PD294	2-Br-Pyridin-4-yl	OH	Cl	8.48
PD295	2-I-Pyridin-4-yl	OH	Cl	8.43
PD296	2-Me-Pyridin-4-yl	OH	Cl	8.29
PD299	2-OMe-Pyridin-4-yl	OH	Cl	8.03
PD302	2-CN-Pyridin-4-yl	OH	Cl	8.22
PD303	2-NO_2-Pyridin-4-yl	OH	Cl	8.15
PD304	2-CF_3-Pyridin-4-yl	OH	Cl	8.51
PD311	2-I-Pyrimidin-4-yl	OH	Cl	8.26
PD312	2-Me-Pyrimidin-4-yl	OH	Cl	8.32
PD315	2-OMe-Pyrimidin-4-yl	OH	Cl	8.06
PD316	2-Et-Pyrimidin-4-yl	OH	Cl	8.15
PD317	2-Cyclopropyl-Pyrimidin-4-yl	OH	Cl	8.02
PD320	2-CF_3-Pyrimidin-4-yl	OH	Cl	9.57
PD322	2-CO_2H-Pyrimidin-4-yl	OH	Cl	8.30
PD325	6-Cl-Pyrimidin-4-yl	OH	Cl	8.08
PD328	6-Me-Pyrimidin-4-yl	OH	Cl	8.00
PD331	6-OMe-Pyrimidin-4-yl	OH	Cl	8.86

Table 3.32 Highly predicted activity benzothiophene compounds (continued)

Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD333	6-Cyclopropyl-Pyrimidin-4-yl	OH	Cl	9.29
PD335	6-NO_2-Pyrimidin-4-yl	OH	Cl	8.38
PD336	6-CF_3-Pyrimidin-4-yl	OH	Cl	8.72
PD339	6-$CONH_2$-Pyrimidin-4-yl	OH	Cl	8.91
PD351	5-NO_2-Pyrimidin-4-yl	OH	Cl	8.41
PD352	5-CF_3-Pyrimidin-4-yl	OH	Cl	9.46
PD357	6-Cl-pyrazin-2-yl	OH	Cl	8.41
PD358	6-Br-pyrazin-2-yl	OH	Cl	8.68
PD359	6-I-pyrazin-2-yl	OH	Cl	8.54
PD362	6-NH_2-pyrazin-2-yl	OH	Cl	8.12
PD364	6-Et-pyrazin-2-yl	OH	Cl	8.16
PD365	6-Cyclopropyl-pyrazin-2-yl	OH	Cl	8.27
PD366	6-CN-pyrazin-2-yl	OH	Cl	8.16
PD367	6-NO_2-pyrazin-2-yl	OH	Cl	8.77
PD368	6-CF_3-pyrazin-2-yl	OH	Cl	8.58
PD375	5-I-pyrazin-2-yl	OH	Cl	8.02
PD378	5-NH_2-pyrazin-2-yl	OH	Cl	8.04
PD379	5-OMe-pyrazin-2-yl	OH	Cl	8.00
PD381	5-Cyclopropyl-pyrazin-2-yl	OH	Cl	8.11
PD382	5-CN-pyrazin-2-yl	OH	Cl	9.13
PD386	5-CO_2H-pyrazin-2-yl	OH	Cl	8.29
PD387	5-$CONH_2$-pyrazin-2-yl	OH	Cl	8.01
PD393	3-OH-pyrazin-2-yl	OH	Cl	8.05
PD398	3-CN-pyrazin-2-yl	OH	Cl	8.80
PD399	3-NO_2-pyrazin-2-yl	OH	Cl	8.46
PD406	4-Br-Pyrimidin-2-yl	OH	Cl	8.16
PD408	4-Me-Pyrimidin-2-yl	OH	Cl	8.03
PD409	4-OH-Pyrimidin-2-yl	OH	Cl	9.16
PD412	4-Et-Pyrimidin-2-yl	OH	Cl	8.00
PD416	4-CF_3-Pyrimidin-2-yl	OH	Cl	8.21

Table 3.32 Highly predicted activity benzothiophene compounds (continued)

Cpd.	R_1	R_2	R_3	CoMSIA[a]
PD425	5-OH-Pyrimidin-2-yl	OH	Cl	8.21
PD426	5-NH_2-Pyrimidin-2-yl	OH	Cl	8.82
PD430	5-CN-Pyrimidin-2-yl	OH	Cl	8.66
PD438	4-Br-1,3,5-triazin-2-yl	OH	Cl	8.61
PD439	4-I-1,3,5-triazin-2-yl	OH	Cl	9.24
PD440	4-Me-1,3,5-triazin-2-yl	OH	Cl	8.31
PD444	4-Et-1,3,5-triazin-2-yl	OH	Cl	8.28
PD446	4-CN-1,3,5-triazin-2-yl	OH	Cl	8.05
PD454	1,3-Benzodioxol-5-yl	OH	Br	9.06
PD455	1,3-Benzodioxol-5-yl	OH	I	8.23
PD458	1,3-Benzodioxol-5-yl	OH	CH_2CH_3	8.48
PD460	1,3-Benzodioxol-5-yl	OH	CF_3	8.21
PD461	1,3-Benzodioxol-5-yl	F	Cl	9.18
PD466	3-OH-isoxazol-5-yl	OH	H	8.63
PD467	3-OH-isoxazol-5-yl	OH	F	8.42
PD469	3-OH-isoxazol-5-yl	OH	I	9.18
PD470	3-OH-isoxazol-5-yl	OH	CH_3	8.01
PD471	3-OH-isoxazol-5-yl	OH	OCH_3	8.32
PD472	3-OH-isoxazol-5-yl	OH	CH_2CH_3	8.40
PD474	3-OH-isoxazol-5-yl	OH	CF_3	9.84
PD475	3-OH-isoxazol-5-yl	F	Cl	9.46
PD476	3-OH-isoxazol-5-yl	Cl	Cl	9.44
PD477	3-OH-isoxazol-5-yl	CO_2H	Cl	8.93
PD478	3-OH-isoxazol-5-yl	CN	Cl	8.80
PD479	3-OH-isoxazol-5-yl	NO_2	Cl	9.28

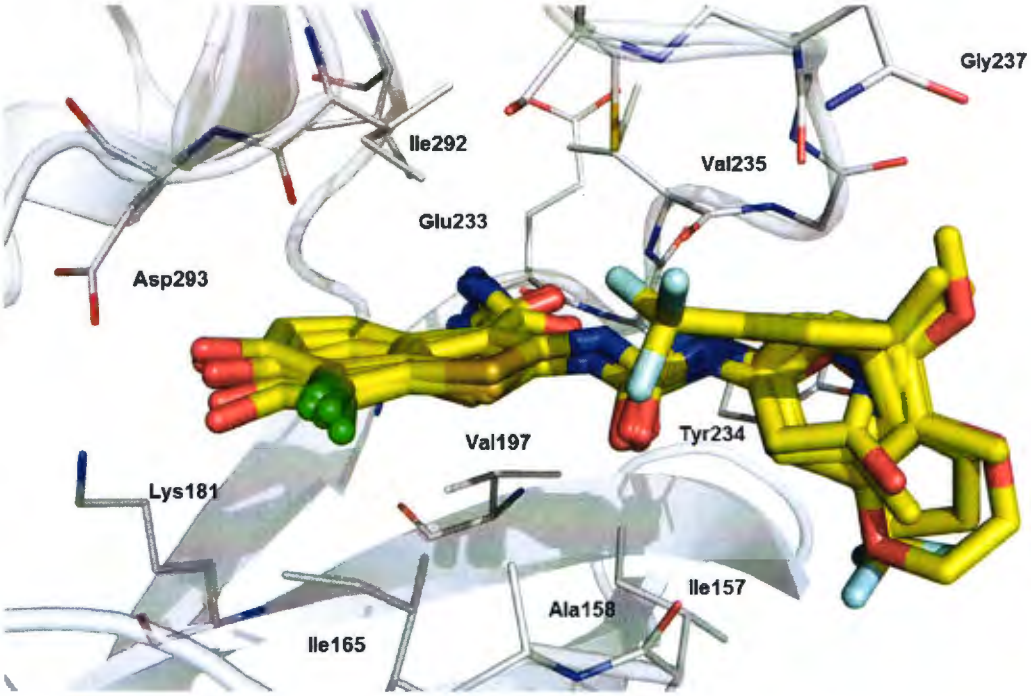


Figure 3.52 Superimposition binding mode of designed benzothiophene compounds

The binding mode of **PD118** and **PD240** were displayed in Figure 3.53 and Figure 3.54, respectively. Six hydrogen bond interactions were observed in both designed compounds. Primary amide of benzothiophene core structure formed two hydrogen bond interactions with oxygen of carbonyl backbone of Glu233 and NH backbone of Val235. Two hydrohen bond interactions of NH urea linker closed to R_1 position formed hydrogen bond interactions with carbonyl (C=O) backbone of Val235. Hydroxyl group of R_2 position formed two hydrogen bond interactions with Lys197 and Asp293 side chains. R_1 position of designed compounds interacted with amino acid residues surrounding the active site *via* hydrophobic interactions.

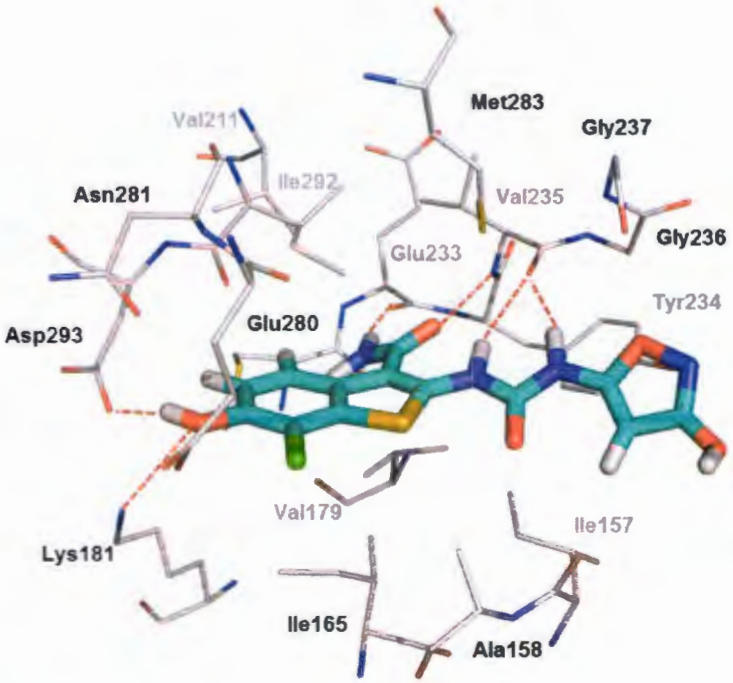


Figure 3.53 Binding mode of PD118 in PknG binding pocket derived from molecular docking calculation

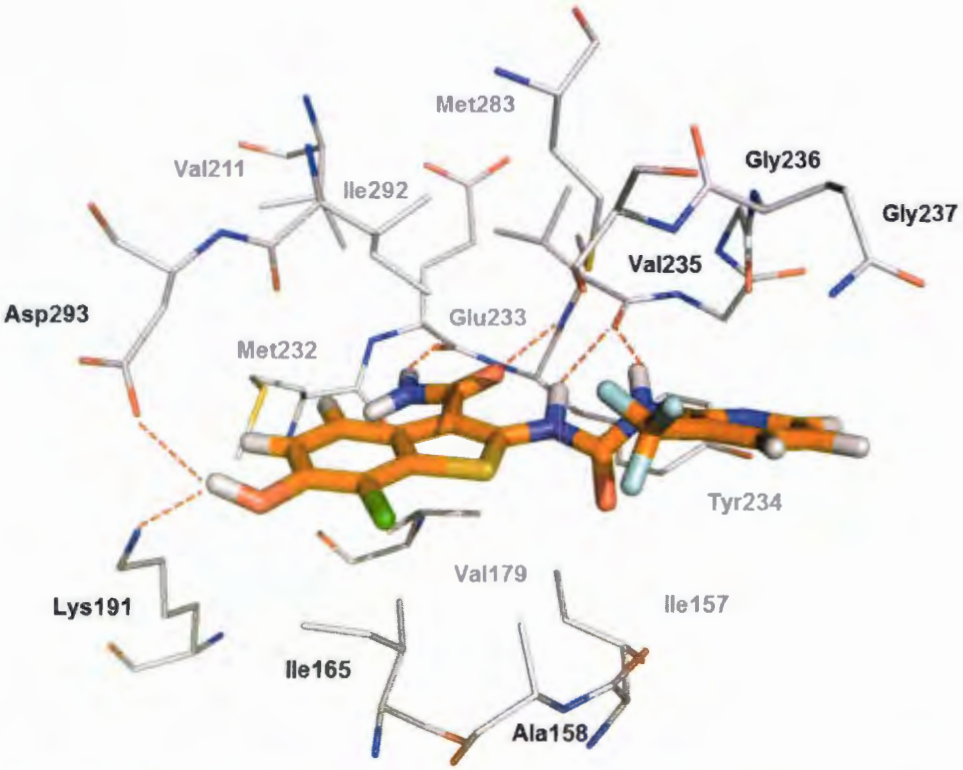


Figure 3.54 Binding mode of PD240 in PknG binding pocket derived from molecular docking calculation

3.1.2.7 Virtual screening of novel PknG inhibitors

1) Receptor site identifications

There are three available of the X-ray structures of PknG enzyme. Only one X-ray crystal structure of PknG complexed with thtrahydrobenzothiophene inhibitor (AX20017) was reported. Therefore, this structure (PDB code: 2PZI) was selected as receptor for virtual screening in this work. The x-ray structure of PknG contained 681 residues (73-750). This PknG structure contained three parts of proteins, a tetratricopeptide repeat (TPR) domain, kinase domain and rubredoxin domain, respectively. AX20017 compound was located at the ATP catalytic binding site of kinase domain. The binding site of AX20017 compound was defined as the receptor for identification of new PknG inhibitors in this study. This pocket contained two water molecules in the binding site. Therefore, two different types of receptor (contained water molecules in the binding site and no water molecule in the binding site) in this study were investigated.

2) Validation of docking program

AX20017 compound was docked into the kinase binding site of 2PZI using Glide XP docking programs to validate the performant of docking parameter. Glide XP docking was validated by root mean square deviations (RMSD) as shown in Figure 3.55. The RMSD values lower than 1 Å were obtained for both different receptors. The RMSD values of contained water molecules and no-water molecules are 0.43 and 0.38 Å, respectively. It can be seen that water molecules don't have important rules for binding of AX20017. Therefore, the AX20017 with no-water in the binding site was selected as the receptor for virtual screening.

Contained water molecules No-water molecules

Figure 3.55 Superimposition of X-ray crystal structure (Green stick) and docking conformations (Cyan stick) of AX20017 compound

3) Virtual screening of new PknG inhibitors from GSK database

To find novel PknG inhibitors, structure based virtual screening was performed. Two steps of virtual screening; molecular docking using Glide XP docking and binding free energy calculation were applied to identify new PknG inhibitors as shown in Figure 3.56. Firstly, 793 compounds from GSK database were docked into PknG binding pocket using the Glide XP scoring function with the same parameters with AX20017 derivatives. Then, top 100 compounds based on Glide XP scoring were rescored by MM-GBSA calculations with igb8. The highest binding free energy of know PknG inhibitors (-34.23 kcal/mol) was used as cut-off energy to select hit compounds. Based on this cut-off energy, 81 compounds were obtained. Then, the lipinski rule of 5 was applied to select 53 good drug property compounds. These hits were clustered to 18 classes and compounds with lowest binding free energy from each cluster were selected. Then, two hydrogen bond interactions with Glu233 and Val235 were considered to select potential hit compounds. Finally, 8 compounds were obtained as shown in Table 3.33 and Figure 3.57.

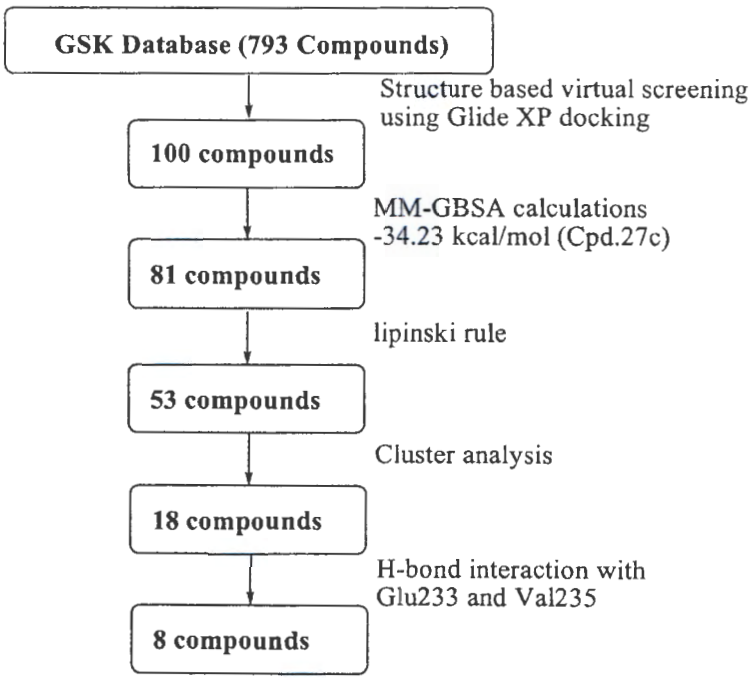
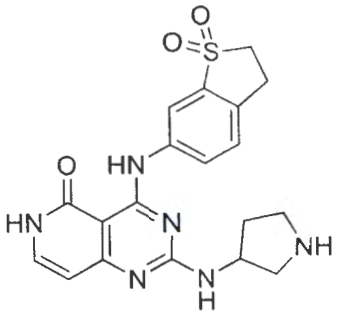


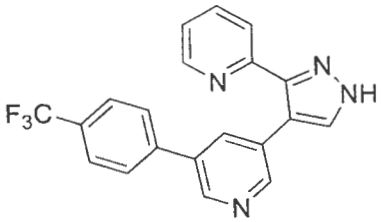
Figure 3.57 Schematic representation of the virtual screening workflow for PknG inhibitors

Table 3.33 Hit compounds of PknG inhibitors derived from virtual screening

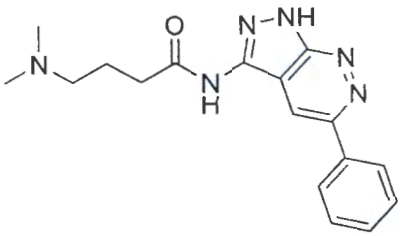
GSK ID	Chemical name	ΔH	Nrot	ΔG
GSK2181306A	4-[(1,1-Dioxo-2,3-dihydro-1-benzothiophen-6-yl)amino]-2-(pyrrolidin-3-ylamino)-6*H*-pyrido[4,3-d] pyrimidin-5-one	-51.60	4	-55.60
GW599550X	2-[4-(Trifluoromethyl)phenyl]-4-(3-pyridin-2-yl-1*H*-pyrazol-4-yl)pyridine	-45.37	4	-49.37
SB-627772-A	4-Dimethylamino-*N*-(5-phenyl-1*H*-pyrazolo[5,4-c]pyridazin-3-yl) butanamide	-42.35	6	-48.35
GSK2298859A	(3*S*,6*R*)-1-[6-(3-Amino-1*H*-indazol-6-yl)-2-(methylamino)-4-pyrimidinyl]-6-methyl-*N*-phenyl-3-piperidine carboxamide	-40.74	5	-45.74
SB-222516	4-[5-(4-Fluorophenyl)-3-(3-morpholinopropyl)imidazol-4-yl]-*N*-methyl-pyrimidin-2-amine	-38.67	7	-45.67
SB-390767	3-(3,5-Dichloro-4-hydroxyanilino)-4-(3-methoxyphenyl)pyrrole-2,5-dione	-39.18	4	-43.18
GW276655X	4-[1-(5-Oxazol-5-yl-2-oxo-1,2-dihydro-indol-3-ylidene)-hydrazino] benzenesulfonamide	-36.32	4	-40.32
GSK993273A	*N*-[4-(1*H*-Pyrrolo[2,3-b]pyridin-4-yl) phenyl]methanesulfonamide	-34.51	3	-37.51



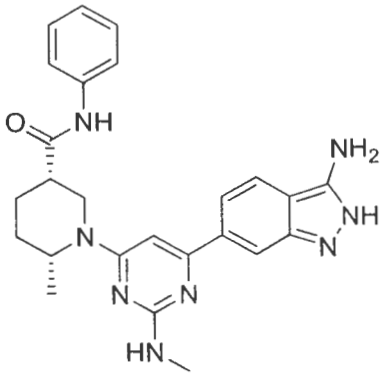
GSK2181306A



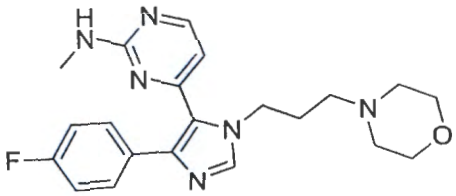
GW599550X



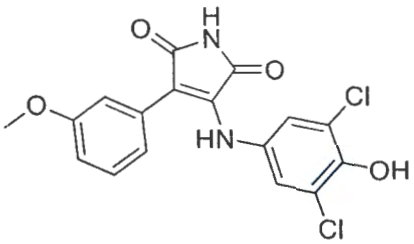
SB-627772-A



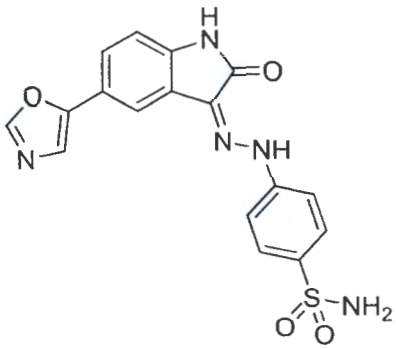
GSK2298859A



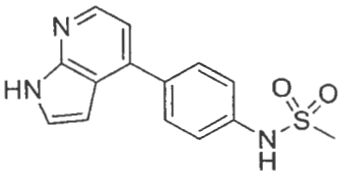
SB-222516



SB-390767



GW276655X



GSK993273A

Figure 3.57 Chemical structure of new hit PknG inhibitors derived from virtual screening

The binding interactions of hit compounds in PknG binding site derived from molecule docking were analyzed. Two hydrogen bond interactions of 4-[(1,1-Dioxo-2,3-dihydro-1-benzothiophen-6-yl) amino]-2-(pyrrolidin-3-ylamino)-6*H*-pyrido[4,3-d] pyrimidin-5-one with Glu233 and Val235 were found from the criteria to select potential hit compounds as shown in Figure 3.58. Moreover, three additional hydrogen bond interactions were observed. Pyrrolidine ring (NH) from acts as hydrogen bond donor to from hydrogen bond interactions with Glu280 and Asn281 side chain. Oxygen atom of sulfone functional acts as hydrogen bond acceptor to form hydrogen bond interaction with NH backbone of Ser239. Sigma-pi interaction between side chain of Val179 with pyrido[4,3-d]pyrimidin-5(6*H*)-one was observed. Moreover, hydrophobic interactions of this compound with amino acids in the binding site were observed.

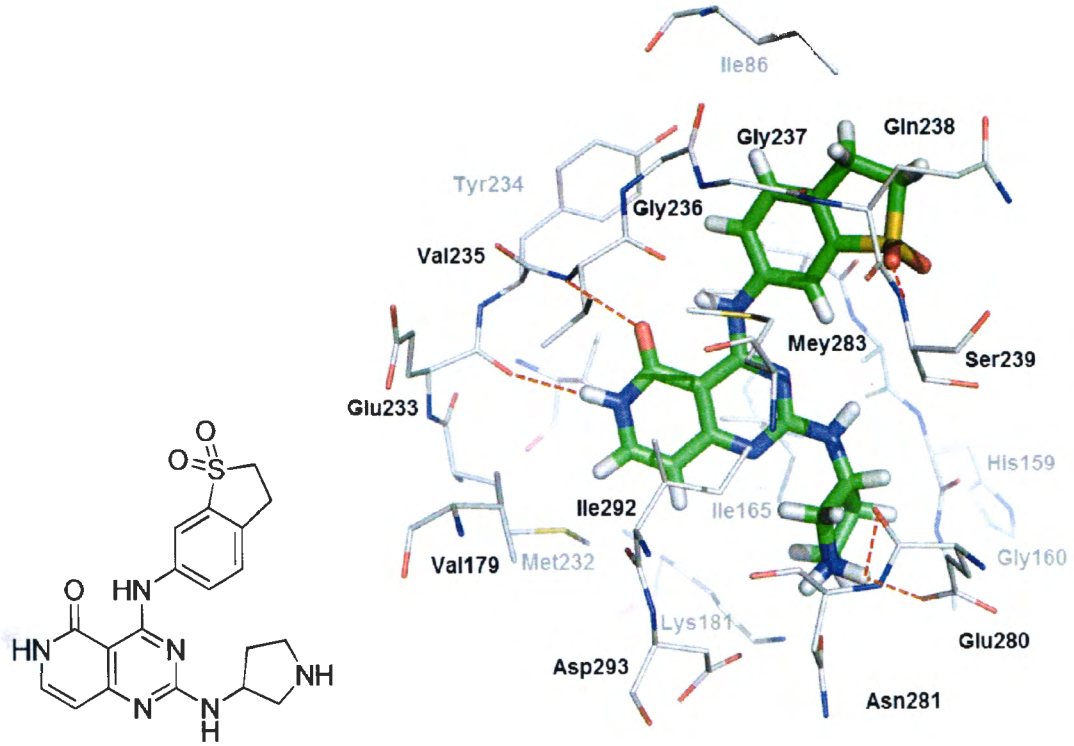


Figure 3.58 Binding mode of GSK2181306A

3.2 Rational design of anti-cancer agents

3.2.1 CoMSIA models

The training set chosen contains 24 compounds with activity range from 4.58 to 8.10 in log(1/IC_{50}) units. Only one compound, compound **12**, has a high activity value. If it is removed from the dataset, the activity range narrows. It is important to note that compound **12** shows great influence on the CoMSIA model. Removing this compound from the train set results in statistically unsatisfied model (unpublished data). Therefore, compound **12** was included as the training set in this study. Table 3.34 lists the statistical parameters of CoMSIA models obtained from the PLS analysis. With the highest q^2 of 0.65, the CoMSIA model including steric, electrostatic, hydrophobic and hydrogen acceptor fields was selected as the best CoMSIA model. The contribution of steric, electrostatic, hydrophobic and hydrogen acceptor fields is 12 %, 18 %, 42 % and 28 %, respectively, indicating that the hydrophobic field shows greatest influence on the activity of azanaphthoquinone annelated pyrrole derivatives. The selected CoMSIA model has high power to estimate the activities of training set with r^2 of 0.99 and q^2 of 0.65. In order to assess the predictive ability of this CoMSIA model, the activities of the test set compounds were predicted. Experimental and predicted activities (log(1/IC_{50}) for the training set and test set are reported, while distribution of experimental and predicted values for the training and the test sets according to the best CoMSIA model is represented in Figure 3.59. The calculated data of compounds in training set fit well with experimental results with error less than 0.1 for all compounds and the prediction error for all tested compounds are less than 1.0 Therefore, the best CoMSIA model could be utilized to predict the activities for new designed azanaphthoquinone annelated pyrrole derivatives.

Table 3.34 Statistical results of various CoMSIA models with different combined fields

Models	Statistical parameters						Fraction
	q^2	r^2	s	SEE	N	F	
S/E	0.31	0.89	0.73	0.30	4	38	44/56
S/H	0.44	0.96	0.66	0.19	4	101	21/79
S/D	0.02	0.38	0.81	0.65	1	14	25/75
S/A	0.44	0.99	0.70	0.08	6	386	37/63
S/E/H	0.61	0.99	0.58	0.08	6	394	14/25/61
S/E/D	0.12	0.45	0.77	0.61	1	18	18/27/55
S/E/A	0.61	0.99	0.58	0.04	6	1819	25/30/45
S/E/H/D	0.42	0.99	0.71	0.09	6	316	12/22/51/16
S/E/H/A[a]	**0.65**	**0.99**	**0.55**	**0.04**	**6**	**1335**	**12/18/42/28**
S/E/H/D/A	0.54	0.99	0.63	0.07	6	510	9/15/33/16/27

[a] The best CoMSIA model. q^2 (leave-one-out cross-validated correlation coefficient), r^2 (non-cross-validated correlation coefficient), N (optimum number of components), s (standard error of prediction), SEE (standard error of estimate), F (F-test value), S (steric field), E (electrostatic field), H (hydrophobic field), D (hydrogen donor field), A (hydrogen acceptor field)

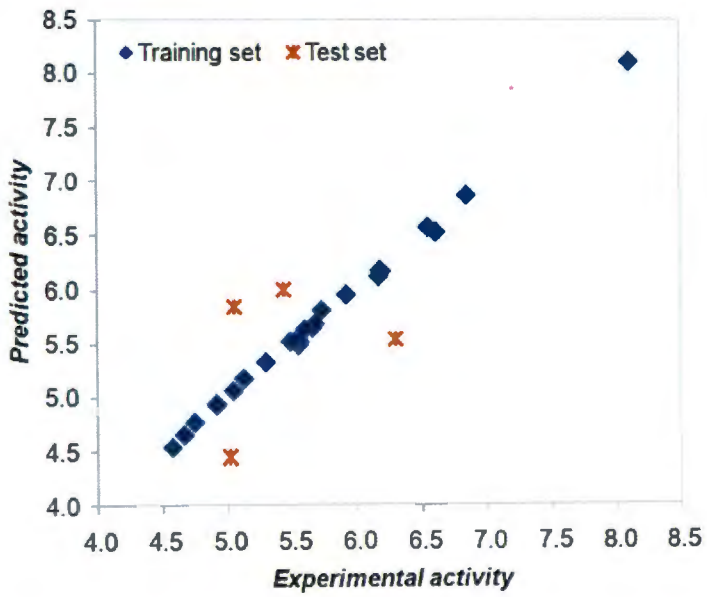


Figure 3.59 Plot between the experimental and predicted activities of the training and test sets derived from the best CoMSIA model

3.2.2 CoMSIA contour maps

Figures 3.60 and Figure 3.61 present the CoMSIA contour maps which reveal the influence of steric, electrostatic, hydrophobic and hydrogen acceptor fields to the activity of azanaphthoquinone annelated pyrrole derivatives. Favorable and unfavorable steric regions are represented in green and yellow contours, respectively, while blue and red contours indicate the regions which favor positive and negative charges, respectively. The magenta and white contours represent the favorable and unfavorable hydrophobic regions, respectively. The cyan and orange contours indicate regions that favor the hydrogen acceptor group and unfavor hydrogen acceptor group, respectively. The interpretation of CoMSIA contour maps reveals the structural requirement of each substituent position in azanaphthoquinone annelated pyrrole scaffold helpful for rational design of novel and potent azanaphthoquinone annelated pyrrole derivatives.

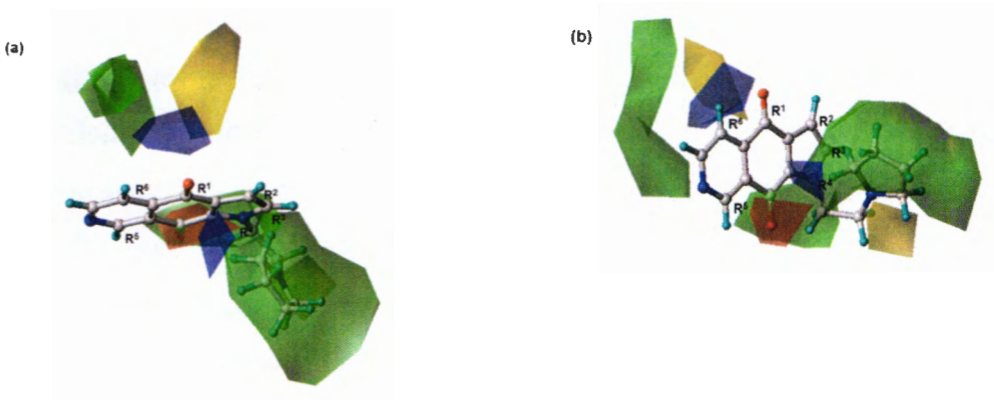


Figure 3.60 Steric and electrostatic contours obtained from the best CoMSIA model in combination with the most active compound (compound 12), (a) and (b) are the steric and electrostatic contours with the horizontal and vertical plane of azanaphthoquinone annelated pyrrole scaffold, respectively

Figure 3.61 Hydrophobic and hydrogen acceptor contours obtained from the best CoMSIA model in combination with the most active compound (compound 12), (a) and (b) are the hydrophobic and hydrogen acceptor contours with the horizontal and vertical plane of azanaphthoquinone annelated pyrrole scaffold, respectively

3.2.2.1 Structural requirement for the R_1 position

Among all selected CoMSIA descriptors, a large yellow, contour overlapping with a blue contour locate near the R_1 position shown in Figure 3.60. These contours suggest that this position prefers the small substituent which possesses low electron density. This suggestion is supported by all compounds presenting the R_1 substituent as the carbonyl oxygen showing the activities above five logarithmic units. In particular, the R_1 substituent of the most active compound, compound **12**, is the carbonyl oxygen. Therefore, the carbonyl oxygen should be optimal substituent for the R_1 position.

3.2.2.2 Structural requirement for the R_2 and R_3 positions

The R_3 position is buried in a large unfavorable hydrophobic region, white contour, indicating that the hydrophobic substituent should be not presented at this position (Figure 3.61). Accordingly, the presence of hydrophobic groups at the R_3 position of compounds **23–28** might be one factor responsible for lower activities of these compounds as compared to that of compound **12**. Additionally, a large favorable steric contour locates near the R_3 position (Figure 3.60a). Therefore, the introduction of bulky substituents possessing hydrophilic properties onto this position could enhance the activity of azanaphthoquinone annelated pyrrole derivatives. In case of the R_2 position, no structural requirement is

suggested from the best CoMSIA. Most of the compounds in the training set contain the same substituent at the R_2 position (the CH moiety). Only compound **28**, bearing (aziridine-1-yl)butyl at the R_2 position, is different from those of other compounds. That means the substituent at this position does not significantly contribute to the binding affinity of the compounds. As exemplified by the comparison of the inhibitory activities of compound **15** (R_2: H, log ($1/IC_{50}$)= 5.73), and that of compound **18** (R_2: -$CH_2CH_2NMe_2$, log ($1/IC_{50}$)= 5.85), the bulkier substituent attached to the R_2 position does not confer to the inhibitory activities of both compounds.

3.2.2.3 The Structural requirement for the R_4 position

As shown in Figure 3.60, a large green contour corresponds to the location of the group attached to the R_4 substituent. However, the tolerated steric requirement of this region is shown by a yellow contour located on the opposite side of the favorable steric region. It is indicated that steric occupancy with bulky groups would increase the binding affinity, but the size of the substituent should be optimum and not be too large. In addition, a predominant feature of hydrophobic contour, magenta area, in the proximity of the R_4 substituent (Figure 3.61) predicts favorable hydrophobic substituents. The reliability of the suggestions derived from the CoMSIA contour maps is verified by compound **12**, the highest active compound, with optimum bulky group and preferably hydrophobic property of -NCH_2CH_2-pyrrolidine substituent attached to the R_4 substituent. Compounds, such as compounds **23–27**, occupying the small R_4 substituent i.e. hydrogen atom, lose to fill the bulkier favorable region, showing lower the activities than that of compound **12**. On the other hand, compounds **13** and **15**, occupying the R_4 substituent with too large substituents, display significantly reduced biological activities compared to compound **12**. Besides, lipophilic substituents attached to the similar position of compounds **11**, **14** and **16** are the reason why these compounds exhibit weak inhibitory activities compared to that of compound **12**.

3.2.2.4 Structural sequirement for the R_5 position

The orange contour is placed near the R_5 position indicating that this position disfavors the hydrogen acceptor substituent (Figure 3.61). This finding is supported by the lower activity of compound **21** bearing a nitrogen atom at the R_5 position, as compared to that of compound **19** containing the CH group.

3.2.2.5 Structural requirement for the R_6 position

The orange contour locates near the R_6 position indicating that this position disfavors the hydrogen acceptor substituent (Figure 3.61). This finding is explaining why compound **20** bearing a nitrogen atom at the R_6 position exhibits lower potency than compound **13**.

3.2.3 MD simulations

3.2.3.1 Structural stability

In order to investigate the structural stability during MD simulations, the RMSDs as a function of the simulation time of each complex with respect to the starting structure were analyzed as shown in Figure 3.62. The RMSDs of the complex structures of the selected compounds, **7**, **11**, **12**, **15** and **22** bound to $d(CGTACG)_2$ reach the plateau characteristic at 8 ns, 2 ns, 8 ns, 4 ns and 6 ns, respectively. These results indicate that each complex structure reaches an equilibrium state after that time. Therefore, the information in terms of energy and structure of each system were analyzed over an equilibrium state.

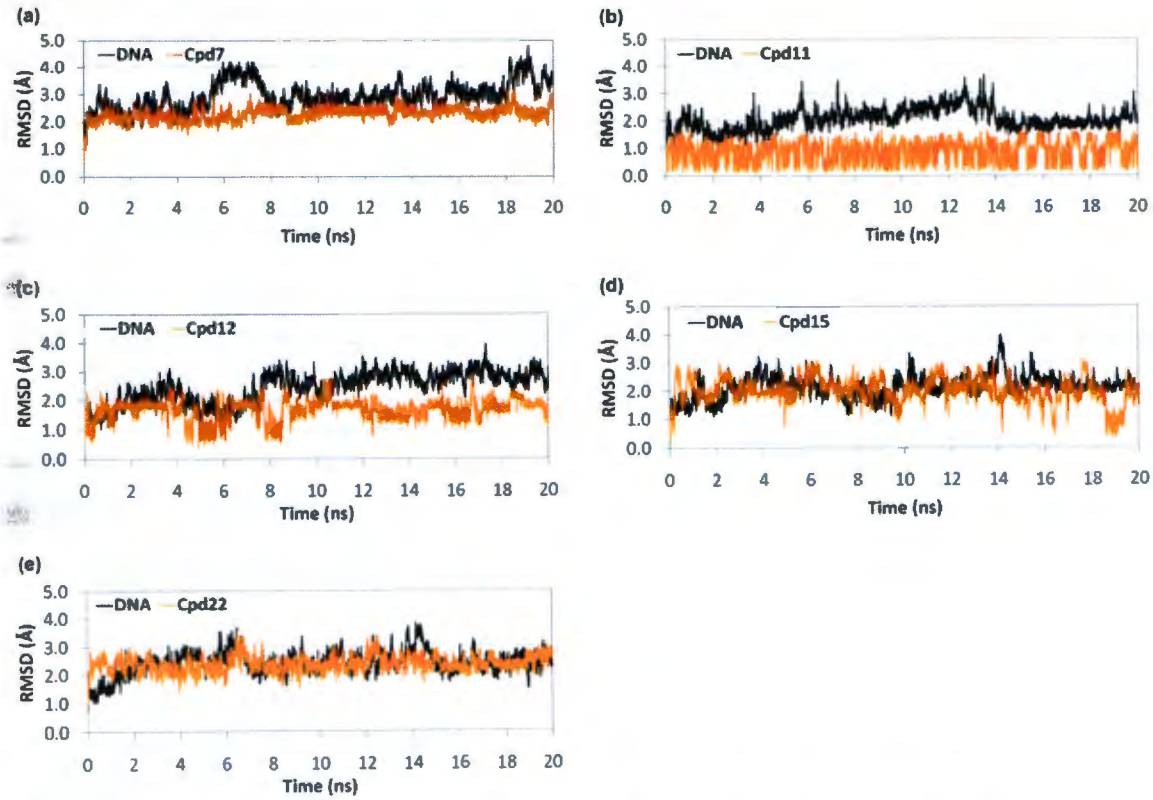


Figure 3.62 RMSD plots for the compound 7(a), 11(b), 12(c), 15(d) and 22(e)/$d(CGTACG)_2$

3.2.3.2 The binding free energy

To gain quantitative insights into the affinity for binding of the azanaphthoquinone derivatives in the intercalation binding site, the binding free energies of the selected compounds were calculated by the MM-PBSA method (Srinivasan *et.al*, 1998). The comparison between the experimental binding free energies (ΔG_{exp}) and the calculated binding free energies (ΔG_{cal}) of compounds **7**, **11**, **12**, **15** and **22** is shown in Table 3.35. The correlation of experimental IC_{50} and calculated free binding free energy is presented in Figure 3.63. It is notable that the calculated binding free energies of all selected compounds are in the correct order as compared with the IC_{50} values. The obtained results could be successfully used to validate the MD procedure in this study. This result shows the reliability of the MD simulations.

Table 3.35 ΔG_{exp} and ΔG_{cal} of the selected azanaphthoquinone derivatives bound to $d(CGTACG)_2$

Cpd.	IC_{50} (μM)	ΔH	-TΔS	ΔG_{cal}	ΔG^{a}_{exp}
7	35.481	-20.0	16.5	-3.5	-6.1
22	26.282	-25.8	20.4	-5.4	-6.3
15	1.862	-22.3	15.9	-6.4	-7.8
11	0.245	-22.9	15.5	-7.4	-9.0
12	0.008	-25.0	16.3	-8.7	-11.1

[a]derived from $\Delta G = RT \ln[\text{Activity}]$, where activity is the antiproliferative activity of compounds **7**, **22**, **15**, **11** and **12** on cervical carcinoma expressed in IC_{50}. R represents the gas constant (1.988 cal/mol K), T represents the temperature (300K).

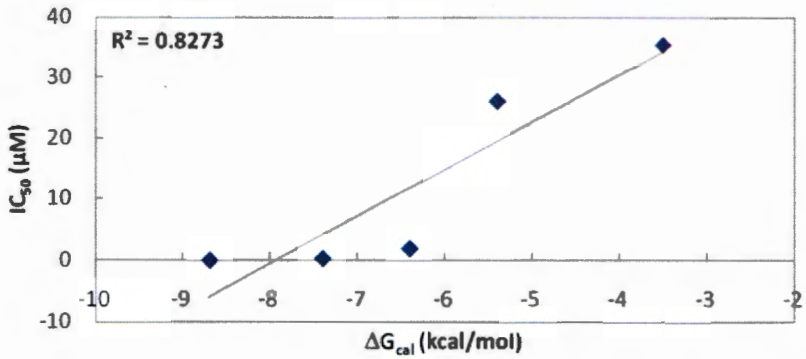


Figure 3.63 Correlation of experimental IC_{50} and calculated free binding free energy using MM-PBSA method

3.2.3.3 Structural analysis

The binding mode analysis of the selected azanaphthoquinone annelated pyrrole derivatives/d(CGTACG)$_2$ complexes started with compound **12** (IC_{50}=0.008 mM), which is the most active compound of the series studies. Compound **12** inserts at the C1G2(A)/C5G6(B) of d(CGTACG)$_2$ as shown in Figure 3.64. The horizontal plane of azanaphthoquinone annelated pyrrole scaffold is bound perpendicular to the horizontal plane of the CG base pairs. The interactions of azanaphthoquinone annelated pyrrole scaffold of this compound and the CG base pairs were observed by pi–pi stacking interactions between quinone and pyrrole moieties with the purine and pyrimidine ring of C1G2(A) and C5G6(B). Compound **12** display extensively hydrogen bond contacts between: (i) the oxygen carbonyl of quinone ring at the R_1 substituent with CH group of C1(A) and G2(A) deoxyribose (ii) the CH group at the alkyl group of the R_4 substituent and NH_2 group of C1(A). Notably, compound **12** is engaged in an additional hydrogen-pi interaction between CH of the pyrrolidini-1-yl-ethyl linker and pyrimidine ring of the C5(B). Moreover, hydrophobic interactions of pyrrolidine ring at the R_4 substituent with C5(B) and G6(B) of the DNA major groove were observed. The key structural features derived are in consistence with the CoMSIA interpretation. Numerous crucial interactions observed for compound **12** should be accounted for displaying the excellent binding free energy (-11.1 kcal/mol). Compounds **11** and **15** (IC_{50}=2.512 and 1.862 mM, respectively) are represented as moderate active compounds. Figure 3.65a shows the binding interactions of compound **11** bound to C1G2(A)/C5G6(B) of d(CGTACG)2.

Compound **11** is parallel to the long axes of the CG base pairs duplex to form the pi–pi stacking interaction with pyrimidine and purine rings of cytosine and guanine, respectively. The R_4 substituent protrude into the major groove of the DNA duplex with C1G2(A). Compound **11** could form hydrogen bond interactions between: (i) the CH group of methyl linker at the R_4 substituent with oxygen carbonyl of C1(A) (ii) the oxygen ether group of oxirane ring at the R_4 substituent with pyrimidine ring of C1(A) (iii) the CH group at chiral carbon of oxirane ring at the R_4 substituent and nitrogen atom of G2(A). However, any hydrogen-p interaction and hydrophobic interactions concerning the R_4 substituent with the base pairs of the DNA duplex are missing. This is the appropriate explanation why compound **11** exhibited lower inhibitory activity compared to compound **12**. The binding interactions of compound **15** bound to of $d(CGTACG)_2$ is presented in Figure 3.65b. The azanaphthoquinone annelated pyrrole scaffold of compound **15** interact with C1C2(A)/C5G6(B) in the intercalation binding site of $d(CGTACG)_2$ are similar to that of compound **12**. The azanaphthoquinone annelated pyrrole axes of these compounds are perpendicular to the long axes of the CG base pairs Therefore, the scaffold of compound **15** inserts in the CG steps of the DNA duplex to form the pi–pi stacking interactions between quinone and pyrrole moieties with the purine and pyrimidine ring. Only two hydrogen bond interactions between the CH of butyl linker at the R_4 substituent of compound **15** and the nitrogen atom of purine base G6(B) were observed. Similar to compound **11**, any hydrogen-pi interaction and hydrophobic interactions concerning the R_4 substituent with the base pairs of the DNA duplex were observed. Moreover, the conformation of the DNA helix strand of compound **15**/$d(CGTACG)_2$ complex was significantly changed. The number and the quality of interactions between the R_4 substituent and base pairs of the DNA duplex decreased could explain their moderate to weak inhibitory activities of compounds **11** and **15** compared to that of compound **12**. Compounds **22** and **7** (IC_{50}=26.282 mM and 35.481 mM, respectively), are representative compounds possessing weak inhibitory activities in this analogues. Interestingly, the chemical structure of compound **22** is highly similar to that of compound **12**, the most active compound, except the steric hindrance of the R_1 substituent. To investigate the effect of the R_1 substituent on the inhibitory activities of the compounds, the binding mode of compound **22**/ $d(CGTACG)_2$ was compared to that of compound **12**. In contrary,

the horizontal plane of azanaphthoquinone annelated pyrrole scaffold of compound **22** is parallel only to the horizontal plane of the C5(B)G2(A) base pair (Figure 3.66a). The CH_2CH_2-pyrrolidine side chain at the R_4 position of this compound protrudes into the minor groove of the DNA duplex, whereas bulky substituent, $-NNHCH_2CH_2NMe_2$ group, at the R_1 position protrudes into the major groove. Because of the insertion of compound **22**, bearing too bulky substituent at the R_1 substituent, the hydrogen bonds between the C1(A) base and the G6(B) base were broken leading to open up of the C1(A)G6(B) base pair of the DNA duplex. The broken base pair results in the loss of the pi–pi interactions between the azanaphthoquinone annelated pyrrole scaffold and the CG base pairs of compound **22**. In case of compound **7**, the least active compound in the dataset which containing bulky substituent at the R_1 substituent, the binding mode of compound **7**/d$(CGTACG)_2$ complex is similar to that of compound **22** (Figure 3.66b). The C1(A)G6(B) base pair is opened up because of the insertion of compound **7**. This finding explains why compounds **22** and **7** display poor binding free energy (-6.3 kcal/mol and -3.5 kcal/mol, respectively), as compared to that of compound **12** (-11.1 kcal/mol). Therefore, these obtained results imply that the presence of bulky substituents at both R_1 and R_4 positions such as compounds **7** and **22** diminishes the binding affinity of azanaphthoquinone annelated pyrrole derivatives in the CG step of DNA duplex. There analyses are in well consistence with the CoMSIA suggestion that the small substituent is preferred for the R_1 position.

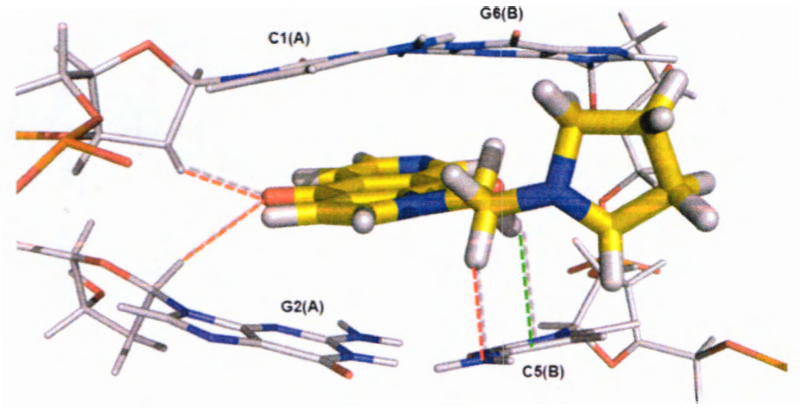


Figure 3.64 **Structure of compound 12/d(CGTACG)$_2$ complex averaged over last 5 ns of the simulation time. For clarity, only the structure in the intercalating part, C1G2(A)/C5G6(B) is shown. Carbon atoms of DNA are colored by gray. Carbon atoms of compound 12 are coloured by yellow. H-bond are colored in red. Hydrogen- π is colored in green.**

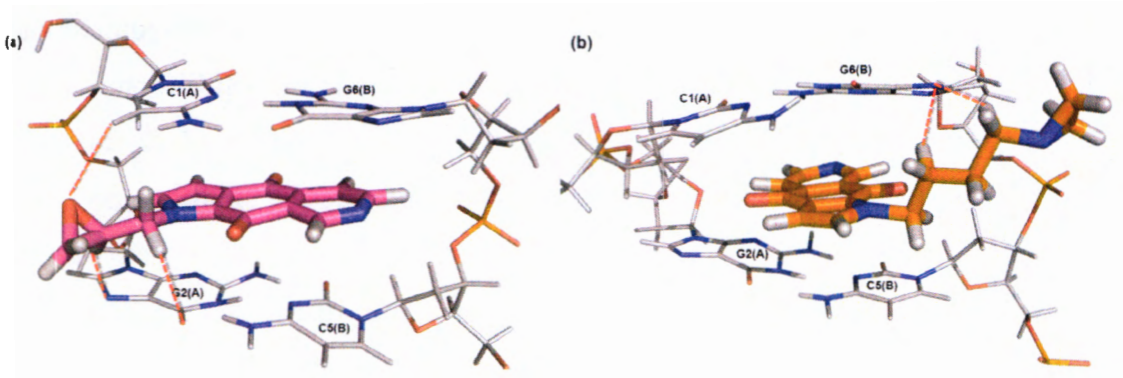


Figure 3.65 **Structure of compound 11/d(CGTACG)$_2$ complex (a) and compound 15/d(CGTACG)$_2$ complex (b) averaged over last 5 ns of the simulation time. For clarity, only the structure in the intercalating part, C1G2(A)/C5G6(B) is shown. Carbon atoms of DNA are colored by gray. Carbon atoms of compound 11 are colored by pink. Carbon atoms of compound 15 are colored by orange. H-bond are colored in red**

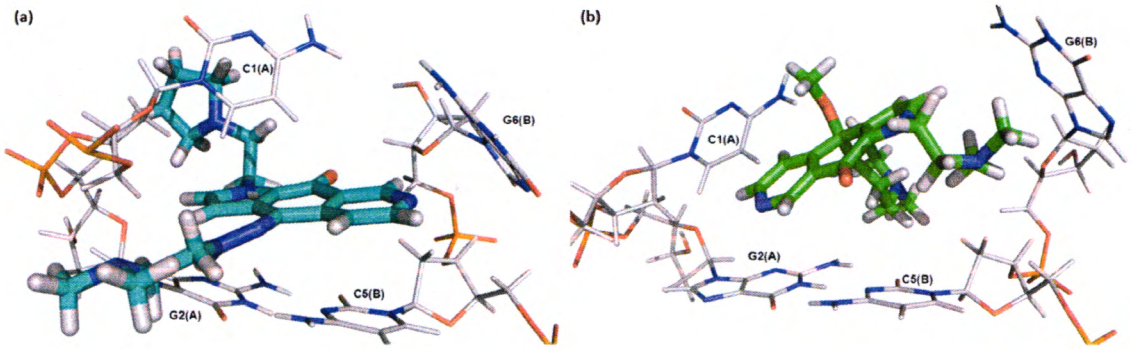


Figure 3.66 Structure of compound 7/d(CGTACG)$_2$ complex (a) and compound 22/d(CGTACG)$_2$ complex (b) averaged over last 5 ns of the simulation time. For clarity, only the structure in the intercalating part C1G2(A)/C5G6(B) is shown. Carbon atoms of DNA are colored by gray. Carbon atoms of compound 7 are colored by green. Carbon atoms of compound 22 are colored by cyan

3.2.3.4 A Comparison between the CoMSIA model and MD Analyses

In order to verify the correspondence of the structural requirements derived from the 3D-QSAR model with the MD analysis, the CoMSIA contour maps were superimposed to the equilibrium MD conformation of compound **12** shown in Figure 3.67. The CoMSIA and MD analyses clearly indicate the similar suggestions for the importance of the R_4 substituents to enhance the inhibitory activities of compounds in the dataset. As previously discussed, the orientation of bulky and hydrophobically favored substituents of the R_4 position bound into the base pairs of the DNA duplex is one of the key characteristics of the compound. The steric contour map highlights the importance of bulky substituent of the pyrrolidini-1-yl-ethyl group at the R_4 position which could be involved in hydrogen bond interaction and hydrogen-pi interaction with the key DNA base pairs, shown in Figure 3.67a. As shown in Figure 3.67b, the hydrophobic map points out the beneficial presence of the hydrophobic substituents at the R_4 position, enhancing hydrophobic contacts with C5(B) and G6(B) of the DNA major groove. Moreover, an unfavorable steric contour located in the vicinity of the R_1 position proves to match with the binding site topology by representing small substituent i.e. the carbonyl group as shown in

compound **12**. Satisfactory agreement obtained from the CoMSIA model and MD analyses may provide insight into crucial structural features effecting ligand receptor interactions and their binding affinities and thus can provide guideline for novel inhibitor design of azanaphthoquinone annelated pyrrole derivatives possessing better antiproliferative activity.

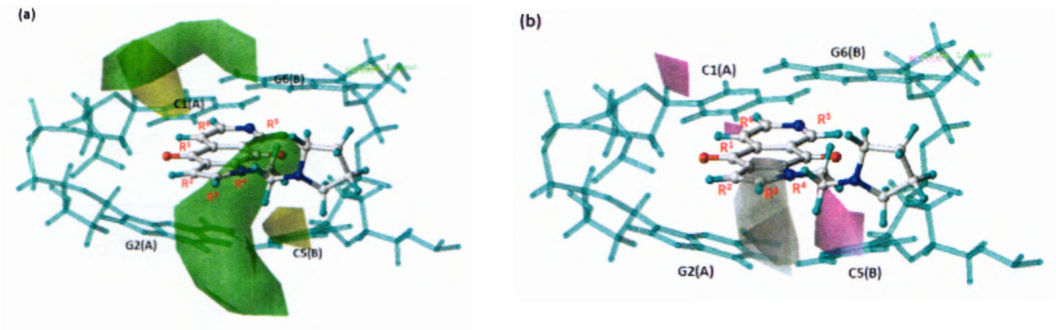
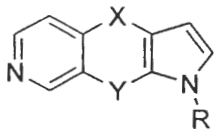


Figure 3.67 Superimposition of the average structure from the last 5 ns of the MD simulation (bond and stick) and the CoMSIA steric (a), and hydrophobic contour (b). For clarity, only the structure in the intercalating part C1G2(A)/C5G6(B) (stick) is shown. Carbon atoms of DNA are colored by light blue. Carbon atoms of compound 12 are colored by atom type. Sterically favored steric areas are represented by green and disfavored steric areas are represented by yellow. Hydrophobically favored areas are represented by magenta and disfavored hydrophobic areas are represented by white

3.2.4 Designed new azanaphthoquinone annelated pyrrole derivatives

The integrated results derived from 3D-QSAR CoMSIA and MD simulations on azanaphthoquinone annelated pyrrole derivatives were used to design new DNA intercalating agents as anti-cancer agents. To design new DNA intercalating agents, compound **12** was selected as the template structure. Structural concepts to modify two carbonyls of compound **12** were summarized. 127 compounds were designed as shown in in Table 3.36. Biological activity against cancer cell line of new designed compounds was predicted. The obtained results show that 16 compounds showed high predicted biological activity than template compound **12**.

Table 3.36 Chemical structure and predicted activity of new designed azanaphthoquinone annelated pyrrole derivatives



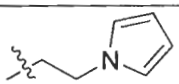
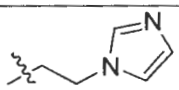
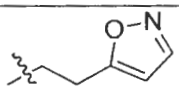
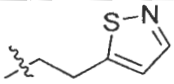
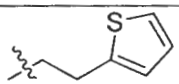
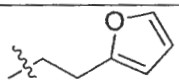
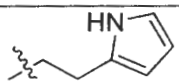
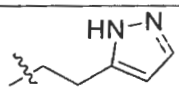
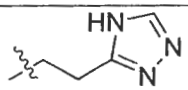
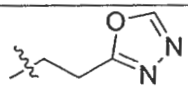
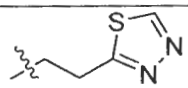
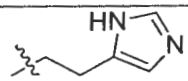
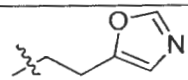
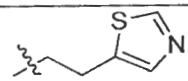
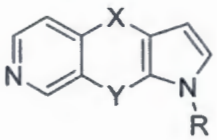
Cpd.	X	Y	R	log(1/IC$_{50}$)
AD1	C=O	C=O		7.40
AD2	C=O	C=O		7.38
AD3	C=O	C=O		7.24
AD4	C=O	C=O		6.83
AD5	C=O	C=O		7.18
AD6	C=O	C=O		7.38
AD7	C=O	C=O		7.45
AD8	C=O	C=O		7.11
AD9	C=O	C=O		6.73
AD10	C=O	C=O		6.66
AD11	C=O	C=O		6.46
AD12	C=O	C=O		7.14
AD13	C=O	C=O		7.06
AD14	C=O	C=O		6.86

Table 3.36 Chemical structure and predicted activity of new designed azanaphthoquinone annelated pyrrole derivatives (continued)



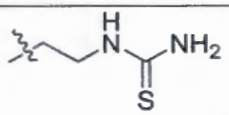
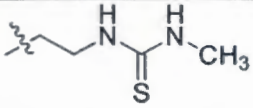
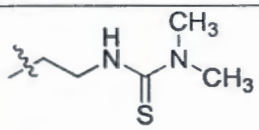
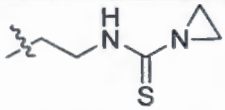
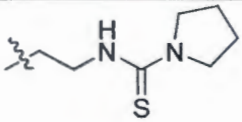
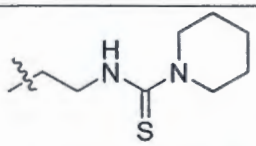
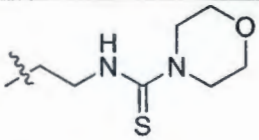
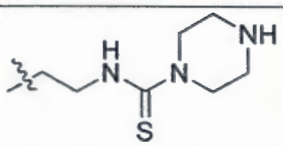
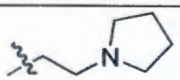
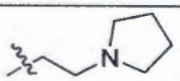
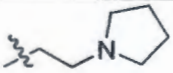
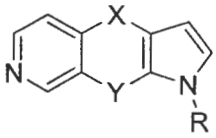
Cpd.	X	Y	R	log(1/IC$_{50}$)
AD15	C=O	C=O		6.86
AD16	C=O	C=O		7.10
AD17	C=O	C=O		7.41
AD18	C=O	C=O		7.08
AD19	C=O	C=O		7.20
AD20	C=O	C=O		7.27
AD21	C=O	C=O		6.89
AD22	C=O	C=O		6.88
AD23	C=O	O		7.97
AD24	C=O	NH		7.94
AD25	C=O	S		7.92

Table 3.36 Chemical structure and predicted activity of new designed azanaphthoquinone annelated pyrrole derivatives (continued)



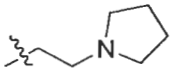
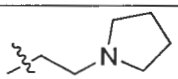
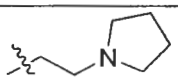
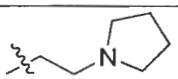
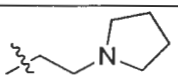
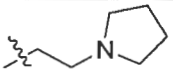
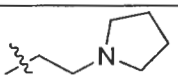
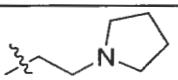
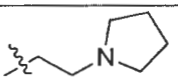
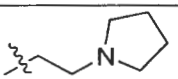
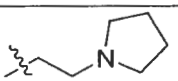
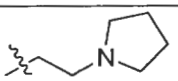
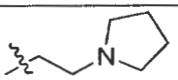
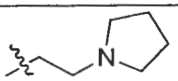
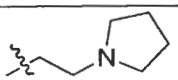
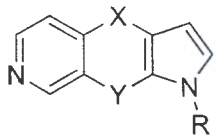
Cpd.	X	Y	R	log(1/IC$_{50}$)
AD26	C=O	S=O		7.79
AD27	C=O	C=CH$_2$		7.99
AD28	C=O	C=NH		8.04
AD29	C=O	C=S		7.98
AD30	C=O	CH$_2$		7.88
AD31	**C=S**	**C=S**		**8.23**
AD32	**C=S**	**C=NH**		**8.29**
AD33	**C=S**	C=CH$_2$		**8.24**
AD34	C=S	C=O		8.17
AD35	C=S	S=O		8.04
AD36	**C=S**	**S**		**8.17**
AD37	**C=S**	**O**		**8.23**
AD38	**C=S**	**NH**		**8.19**
AD39	**C=S**	**CH$_2$**		**8.13**
AD40	**C=NNH(C=S)NH$_2$**	**C=CH$_2$**		**8.24**

Table 3.36 Chemical structure and predicted activity of new designed azanaphthoquinone annelated pyrrole derivatives (continued)



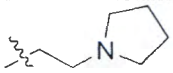
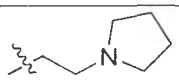
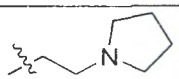
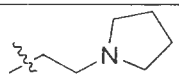
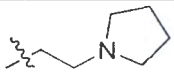
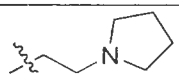
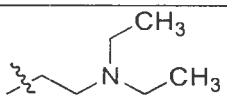
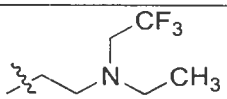
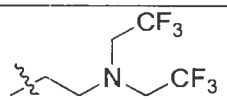
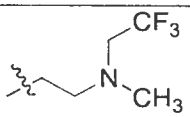
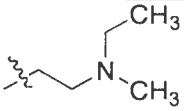
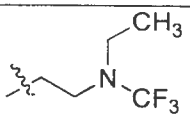
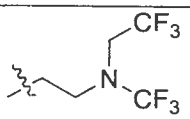
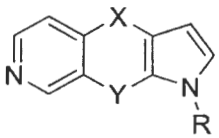
Cpd.	X	Y	R	log(1/IC$_{50}$)
AD41	**C=NNH(C=S)NH$_2$**	**C=NH**		**8.17**
AD42	**C=NNH(C=S)NH$_2$**	**C=S**		**8.23**
AD43	**C=NNH(C=S)NH$_2$**	**CH$_2$**		**8.13**
AD44	**C=NNH(C=S)NH$_2$**	**S**		**8.17**
AD45	**C=NNH(C=S)NH$_2$**	**O**		**8.23**
AD46	**C=NNH(C=S)NH$_2$**	**NH**		**8.19**
AD47	C=O	C=O		7.93
AD48	C=O	C=O		8.10
AD49	C=O	C=O		8.05
AD50	C=O	C=O		7.62
AD51	C=O	C=O		7.67
AD52	C=O	C=O		7.53
AD53	C=O	C=O		7.48

Table 3.36 Chemical structure and predicted activity of new designed azanaphthoquinone annelated pyrrole derivatives (continued)



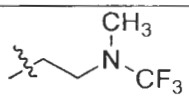
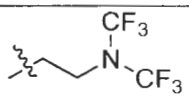
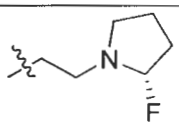
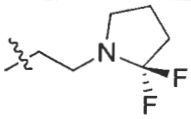
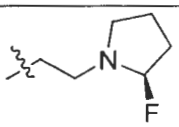
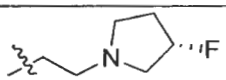
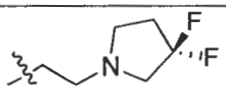
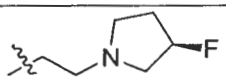
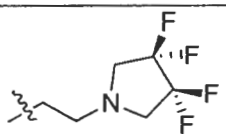
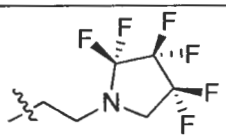
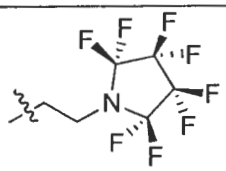
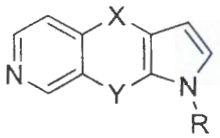
Cpd.	X	Y	R	log(1/IC$_{50}$)
AD54	C=O	C=O		7.25
AD55	C=O	C=O		6.64
AD56	C=O	C=O		7.85
AD57	C=O	C=O		7.95
AD58	C=O	C=O		7.81
AD59	C=O	C=O		7.91
AD60	C=O	C=O		7.81
AD61	C=O	C=O		7.89
AD62	C=O	C=O		7.61
AD63	C=O	C=O		7.66
AD64	C=O	C=O		7.23

Table 3.36 Chemical structure and predicted activity of new designed azanaphthoquinone annelated pyrrole derivatives (continued)



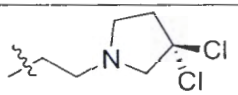
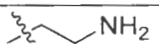
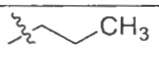
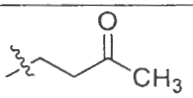
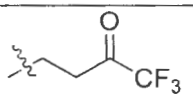
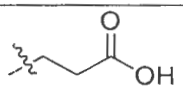
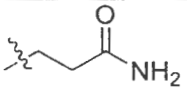
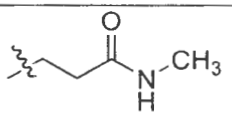
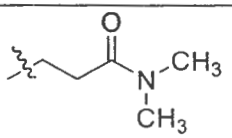
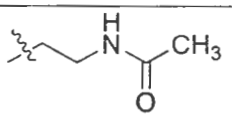
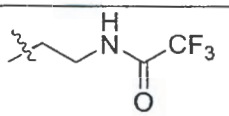
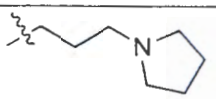
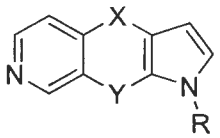
Cpd.	X	Y	R	log(1/IC$_{50}$)
AD65	C=O	C=O		8.10
AD66	C=O	C=O		7.05
AD67	C=O	C=O		6.56
AD68	C=O	C=O		6.38
AD69	C=O	C=O		6.62
AD70	C=O	C=O		6.57
AD71	C=O	C=O		6.76
AD72	C=O	C=O		6.82
AD73	C=O	C=O		7.07
AD74	C=O	C=O		7.15
AD75	C=O	C=O		6.77
AD76	C=O	C=O		6.77
AD77	C=O	C=O		7.51

Table 3.36 Chemical structure and predicted activity of new designed azanaphthoquinone annelated pyrrole derivatives (continued)



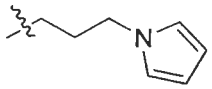
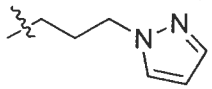
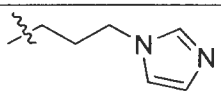
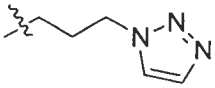
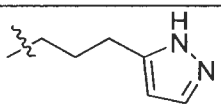
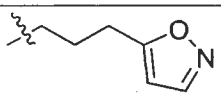
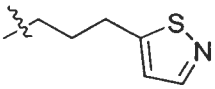
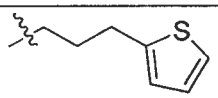
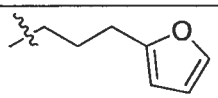
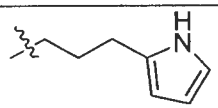
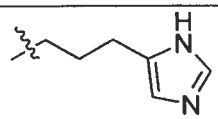
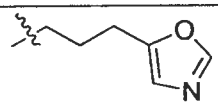
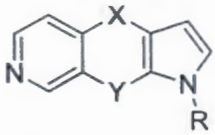
Cpd.	X	Y	R	log(1/IC$_{50}$)
AD78	C=O	C=O		7.20
AD79	C=O	C=O		7.04
AD80	C=O	C=O		7.06
AD81	C=O	C=O		7.20
AD82	C=O	C=O		6.93
AD83	C=O	C=O		6.94
AD84	C=O	C=O		6.87
AD85	C=O	C=O		7.05
AD86	C=O	C=O		7.08
AD87	C=O	C=O		7.11
AD88	C=O	C=O		6.91
AD89	C=O	C=O		5.87

Table 3.36 Chemical structure and predicted activity of new designed azanaphthoquinone annelated pyrrole derivatives (continued)



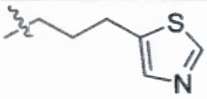
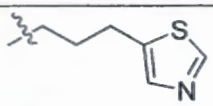
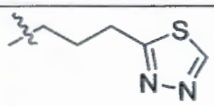
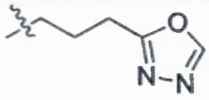
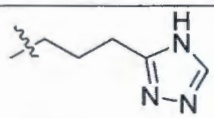
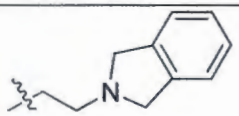
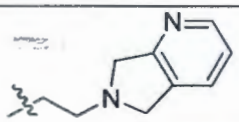
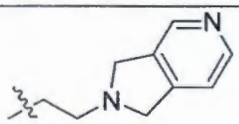
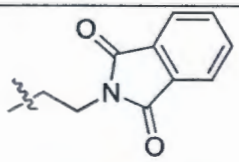
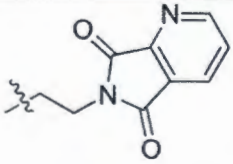
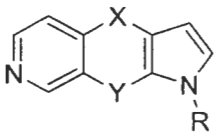
Cpd.	X	Y	R	log(1/IC$_{50}$)
AD90	C=O	C=O		6.85
AD90	C=O	C=O		6.85
AD91	C=O	C=O		6.70
AD92	C=O	C=O		6.73
AD93	C=O	C=O		6.76
AD94	C=O	C=O		8.03
AD95	C=O	C=O		7.25
AD96	C=O	C=O		7.70
AD97	C=O	C=O		7.36
AD98	C=O	C=O		7.28

Table 3.36 Chemical structure and predicted activity of new designed azanaphthoquinone annelated pyrrole derivatives (continued)



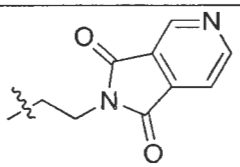
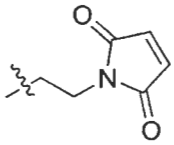
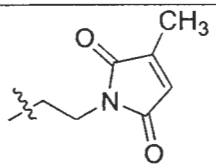
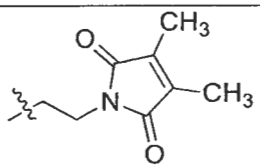
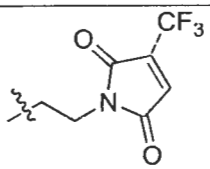
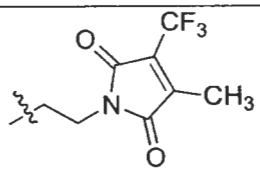
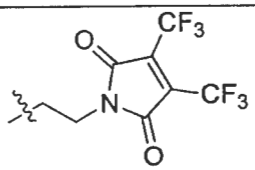
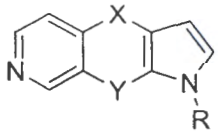
Cpd.	X	Y	R	log(1/IC$_{50}$)
AD99	C=O	C=O		7.04
AD100	C=O	C=O		7.04
AD101	C=O	C=O		7.37
AD102	C=O	C=O		7.60
AD103	C=O	C=O		7.44
AD104	C=O	C=O		7.67
AD105	C=O	C=O		7.80

Table 3.36 Chemical structure and predicted activity of new designed azanaphthoquinone annelated pyrrole derivatives (continued)



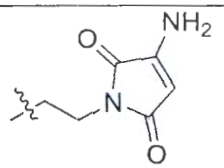
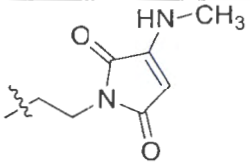
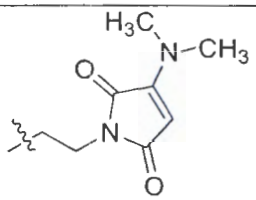
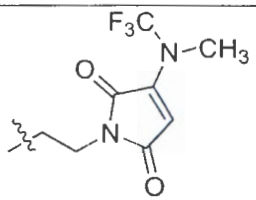
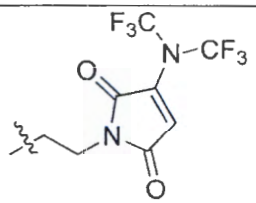
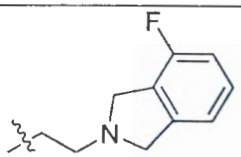
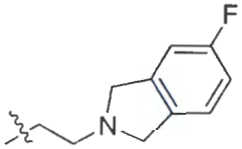
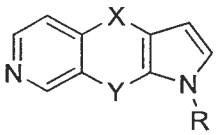
Cpd.	X	Y	R	log(1/IC$_{50}$)
AD106	C=O	C=O		6.95
AD107	C=O	C=O		7.08
AD108	C=O	C=O		7.20
AD109	C=O	C=O		7.49
AD110	C=O	C=O		7.98
AD111	C=O	C=O		7.82
AD112	C=O	C=O		7.66

Table 3.36 Chemical structure and predicted activity of new designed azanaphthoquinone annelated pyrrole derivatives (continued)



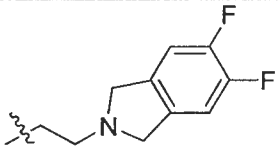
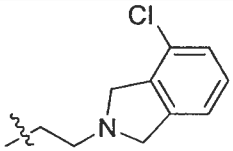
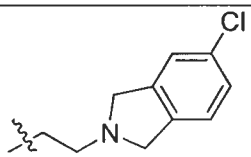
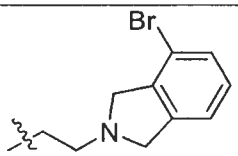
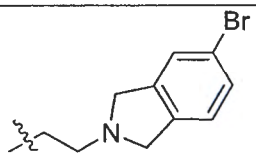
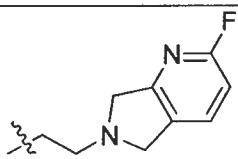
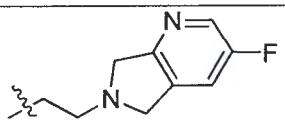
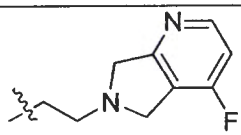
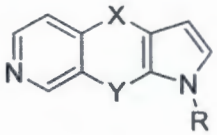
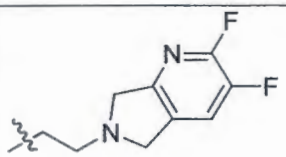
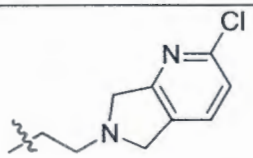
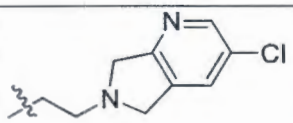
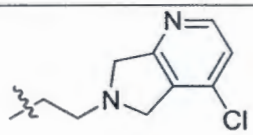
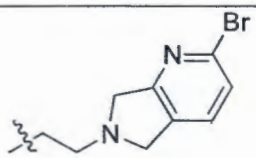
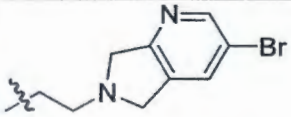
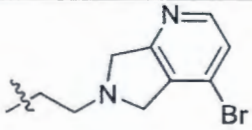
Cpd.	X	Y	R	log(1/IC$_{50}$)
AD113	C=O	C=O		7.28
AD114	C=O	C=O		7.90
AD115	C=O	C=O		7.07
AD116	C=O	C=O		7.94
AD117	C=O	C=O		7.72
AD118	**C=O**	**C=O**		**8.11**
AD119	C=O	C=O		7.55
AD120	C=O	C=O		7.61

Table 3.36 Chemical structure and predicted activity of new designed azanaphthoquinone annelated pyrrole derivatives (continued)



Cpd.	X	Y	R	log(1/IC$_{50}$)
AD121	C=O	C=O		7.74
AD122	C=O	C=O		7.15
AD123	C=O	C=O		7.59
AD124	C=O	C=O		7.70
AD125	**C=O**	**C=O**		**8.19**
AD126	C=O	C=O		7.62
AD127	C=O	C=O		7.75

3.2.5 Promising azanaphthoquinone annelated pyrrole derivatives based on rational design

127 compounds of azanaphthoquinone annelated pyrrole derivatives were designed based on the structural requirements and crucial interaction derived from 3D-QSAR CoMSIA and MD simulations. The obtained results showed that 16 compounds were obtained with higher predicted biological activity than compound **12**, the

template compounds as shown in Table 3.36. The binding mode of high predicted compounds was displayed in Figure 3.68.

Table 3.37 Highly predicted activity of azanaphthoquinone annelated pyrrole compounds

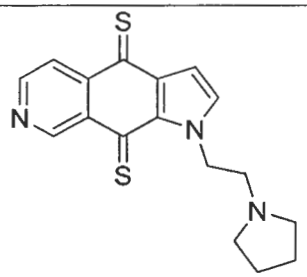
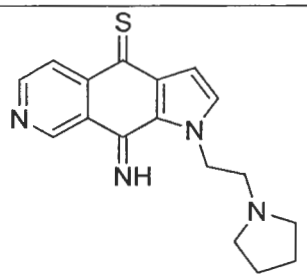
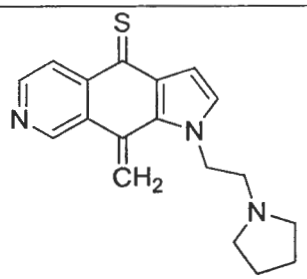
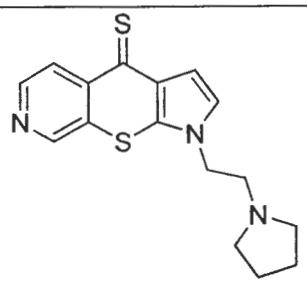
Cpd.	Structure	log(1/IC_{50})
AD31		8.23
AD32		8.29
AD33		8.24
AD36		8.17

Table 3.37 Highly predicted activity of azanaphthoquinone annelated pyrrole compounds (continued)

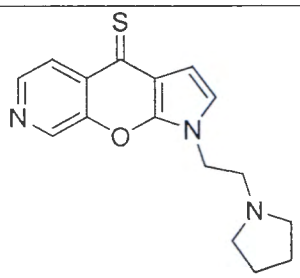
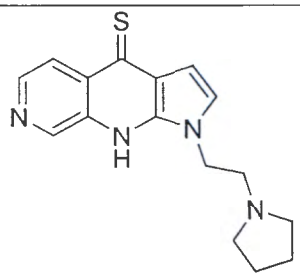
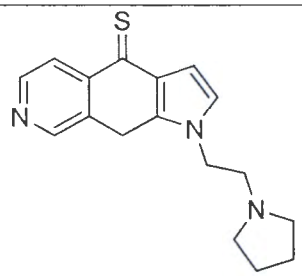
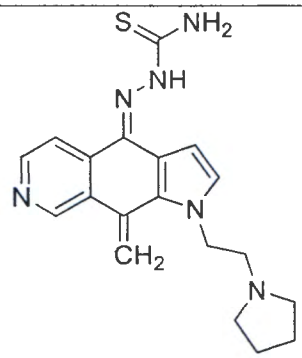
Cpd.	Structure	log(1/IC$_{50}$)
AD37		8.23
AD38		8.19
AD39		8.13
AD40		8.24

Table 3.37 The highly predicted activity of azanaphthoquinone annelated pyrrole compounds (continued)

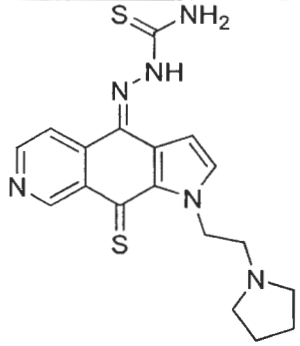
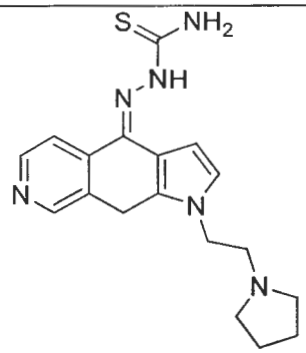
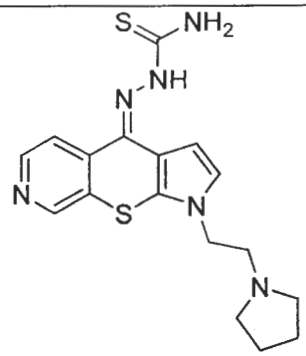
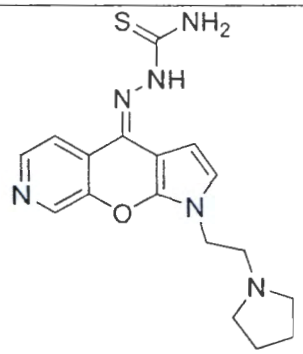
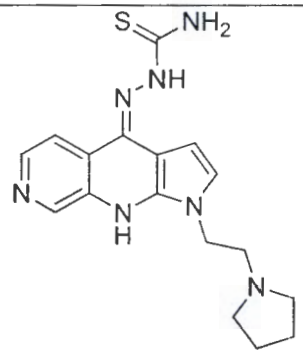
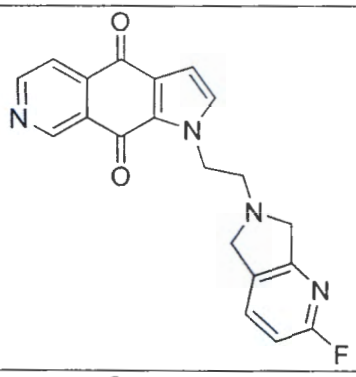
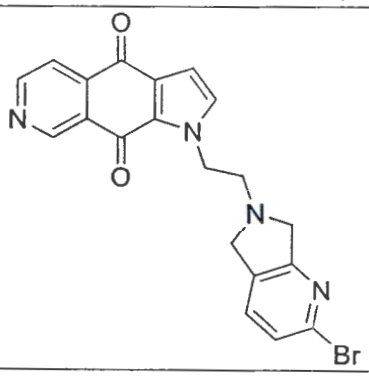
Cpd.	Structure	log(1/IC$_{50}$)
AD42		8.23
AD43		8.13
AD44		8.17

Table 3.37 The highly predicted activity of azanaphthoquinone annelated pyrrole compounds (continued)

Cpd.	Structure	log(1/IC$_{50}$)
AD45		8.23
AD46		8.19
AD118		8.11
AD125		8.19

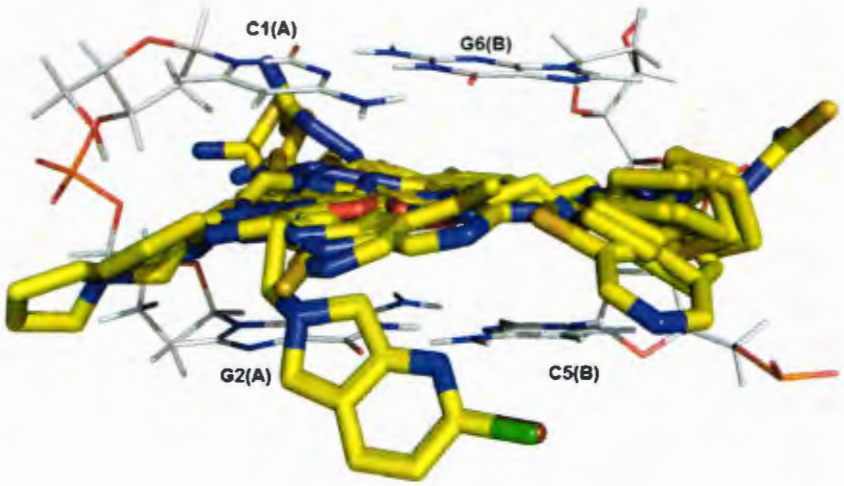


Figure 3.68 Superimposition binding mode of designed azanaphthoquinone annalated pyrrole compounds

The binding mode and binding interactions of **AD32**, the highest predicted biological activity of new designed azanaphthoquinone annelated pyrrole derivative was analyzed as shown in Figure 3.69. Core structure of **AD32** was inserted into the intercalating site and formed pi-pi interactions with C1(A)/G6(B) and G2(A)/C5(B) of DNA base pairs. Pyrrolidini-1-yl-ethyl linker located at major groove of DNA and formed hydrogen bond interactions with C5(B) and G(6) base pair.

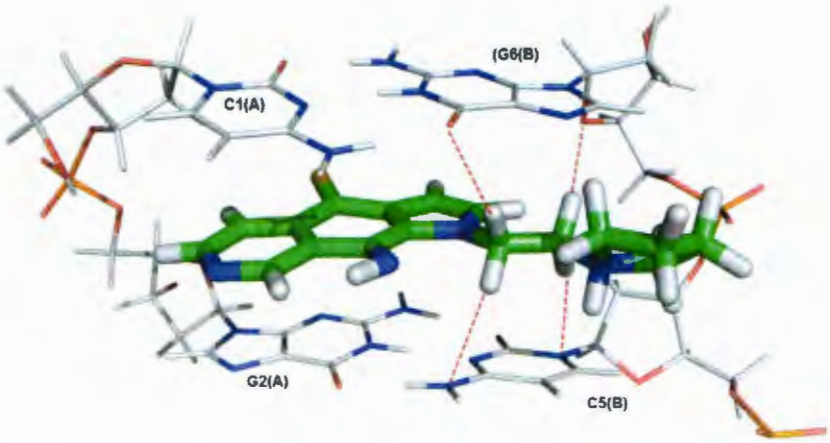


Figure 3.69 Structure of the highest predicted biological activity, compound AD32/d(CGTACG)$_2$ (complex from molecular docking

All of designed compounds in theosemicarbazole series showed high predicted biological activity. Therefore, the binding mode and binding interactions of compounds in this series were analyzed. The binding mode and binding interaction of the highest binding affinity of designed compound is displayed in Figure 3.70. **AD40** is the highest predicted biological activity in theosemicarbazole series. Core structure of **AD40** is inserted into the intercalating site of DNA with their long axis of new scaffold parallel to the long axes of the CG base pairs. The quinone and pyrrole moieties stack with the purine ring of C5(B) and G6(B), whereas the pyridine ring overlaps with two six membered rings of C1(A) and G2(A) as shown in Figure 3.70. The semithiocarbazole (X substituent) protrude into the minor groove of the DNA duplex with C1G2(A). The NH_2 the terminal of semithiocarbazole substituent could form hydrogen bond interaction with oxygen atom of ribose backbone of G2(A) and phosphate linker. Protonation hydrogen at pyrrolidini-1-yl formed hydrogen bond interaction with phosphate backbone of G6(B.

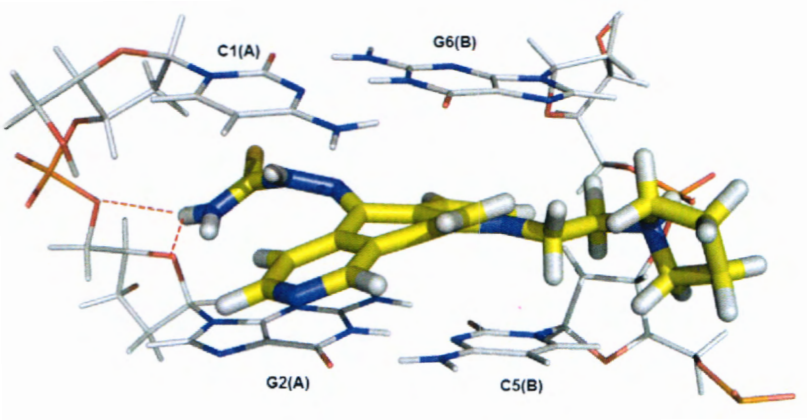


Figure 3.70 Structure of the highest predicted biological acitivity in a series of semithiocarbazole, compound AD40/d(CGTACG)$_2$ (complex from molecular docking

CHAPTER 4
CONCLUSIONS

Molecular modeling and computer aided molecular design approaches were successfully applied to rational design new and more potent anti-TB agents and anti-cancer agents. The first selected target was anti-TB agents. InhA and PknG enzymes were selected as two attractive targets. Diphenyl ether and benzofuran pyrrolidine pyrazole derivatives were chosen to rational design as new and highly effective InhA inhibitors in this study. For rational design of PknG inhibitors as anti-TB agents, benzothiophene derivatives were considered. The second selected target was anti-cancer agents. DNA intercalating agents in a series of azanaphthoquinone annelated pyrrole derivatives were elucidated.

3D-QSAR CoMISA method and MD simulations were performed to investigate the structural requirements and to rational design new and more potent diphenyl ether derivatives as InhA inhibitors of anti-TB agents. The graphic interpretation of the obtained CoMSIA model reveals the key structural elements of diphenyl ether derivatives necessary for good InhA inhibitory activities. MD simulations were successfully applied to reliably predict binding modes, inhibitor–enzyme interactions, and binding free energies of diphenyl ether derivatives in the InhA binding pocket. The structural requirements derived from the CoMSIA model correspond well with the binding interactions of diphenyl ether derivatives in the InhA pocket found in the MD simulations. The presented integrated results should be useful as guiding principles for the design of novel InhA inhibitors based on suitable modifications of the diphenyl ether scaffold. 22 compounds of new designed phenyl ether derivatives with higher predicted biological activity than template design were proposed.

From the data set collected from benzofuran pyrrolidine pyrazole derivatives based InhA inhibitors, 3D-QSAR CoMSIA models were set up with different inhibitory activity against InhA enzyme (IC_{50}) and intact *M. tuberculosis* cell (MIC_{90}).

To study the binding modes and binding interactions of inhibitors at the InhA active site, MD simulations of representative compounds was performed. The combination of graphical interpretation of IC_{50} and MIC_{90} CoMSIA models and MD simulations highlight the structural concept to correctly balance IC_{50} and MIC_{90} values of benzofuran pyrrolidine pyrazole derivatives. The core structure of template compound is crucial to attaining favorable IC_{50} values, whereas the R_2 substituent is a key group to enhance MIC_{90} values without negative effects on IC_{50} values. Based on these beneficial guidelines, new benzofuran pyrrolidine pyrazole derivatives were designed. Among of these designed compounds, 17 designed of novel benzofuran pyrrolidine pyrazole derivatives as InhA inhibitors with better potency against *M. tuberculosis* cells were proposed. Moreover, 2 designed compounds were found as high predicted biological activity in both InhA enzyme assay and intact *M. tuberculosis* cell.

To identify novel scaffolds as InhA inhibitors, two different InhA binding sites, Y158-in and Y158-out conformations were used as receptors for structure based virtual screening. Based on structure based virtual screening workflow in this work, 31 hit compounds with 12 different scaffolds were proposed for biological assay to obtain new scaffolds of InhA inhibitors.

3D-QSAR and MD simulation were used for understanding the structural requirements for the inhibitory activity of the benzothiophene derivatives and binding modes to the enzymatic receptors. Reliable 3D-QSAR CoMSIA model with high predictive ability of the best QSAR model was obtained. The structural requirements to improve biological activity against PknG of benzothiophene derivatives were suggested based on CoMSIA contour maps. Based on detailed CoMSIA contour maps analysis, improvement in PknG binding affinity can be achieved by substitutional modification at R_2 and X position on the benzothiophene ring. The key residues determined by MD simulations provided the key structural basis for binding in PknG binding pocket. The beneficial guideline for designing of new potent PknG inhibitors was obtained. 145 designed compounds based on structural basis derived from 3D-QSAR CoMSIA and MD simulations showed better predicted activity than template compound.

Binding free energy calculations using MM-GBSA method with two solvation models, igb7 and igb8 models were promising methods to estimate the binding affinity

of benzothiophene derivatives as PknG inhibitors. The high correlations of binding free energy calculations using these models with biological activity against PknG were obtained. The entropy contribution on binding free energies calculations was slightly improved the correlations coefficient between inhibitory activity against PknG and binding free energy calculations. These results were applied as filter tools for identification of novel PknG inhibitors via structure based virtual screening. Structure based virtual screening using docking combined with binding free energy calculations using MM-GBSA method was used to identify novel PknG inhibitors. 8 hit compounds were proposed as novel scaffolds for biological assay against PknG.

To develop highly effective anti-cancer agents of azanaphthoquinone annelated pyrrole derivatives as anti-cancer agents, 3D-QSAR CoMSIA study and MD simulations were applied to evaluate their key structural features, binding mode and binding interactions in the DNA duplex. Best CoMSIA model was satisfactory according to the statistical results as well as the contour maps analysis. MD simulations were successful to model the reliable binding modes of azanaphthoquinone annelated pyrrole derivatives in the CG step of $d(CGTACG)_2$. Complex structure of inhibitor/DNA provides the insight into the crucial ligand-DNA interaction and the key structural feature favorable for binding affinity of azanaphthoquinone annelated pyrrole derivatives in the intercalation site of DNA duplex. The finding obtained from MD simulations is supported to the CoMSIA guideline for designing new compounds with the improved biological activity. 17 new designed compounds with high predicted biological activity against cancer cell were proposed.

Accordingly, molecular modeling and computer aided molecular design approaches in this study provides an insight into the crucial structural requirements and the necessary chemical substitutions required to exhibit enhanced inhibitory activity against enzyme, DNA and whole cell of tuberculosis and cancer. Successfully, newly designed compounds with the higher predicted activities as compared with the parent compounds of anti-TB agents and anti-cancer agents were obtained. These proposed new designed compounds can be considered as potent anti-TB agents and anti-cancer agents. More importantly, the structure based virtual screening was applied

to identify novel scaffolds as anti-TB agents. Hit compounds were proposed for biological assay to obtain new anti-TB agents.

REFERENCES

REFERENCES

Agostinis, P. and et al. "Photodynamic therapy of cancer: an update", **CA: A Cancer Journal for Clinicians**. 61(4): 250-281; July-August, 2011.

Alber, T. "Signaling mechanisms of the Mycobacterium tuberculosis receptor Ser/Thr protein kinases", **Current Opinion in Structural Biology**. 19: 650-657; December, 2009.

Aldred, K.J. and et al. "Bacillus anthracis GrlAV96A topoisomerase IV, a quinolone resistance mutation that does not affect the water-metal ion bridge", **Antimicrobial Agents and Chemotherapy**. 58(12): 7182-7187; December, 2014.

Alimuddin, Z., Payam ,N. and Stewart T.C. "Advances in the development of new tuberculosis drugs and treatment regimens", **Nature Reviews Drug Discovery**. 12(5): 388-404; May, 2013.

Alonso, H., Bliznyuk, A. A. and Gready, J. E. "Combining docking and molecular dynamic simulations in drug design", **Medicinal Research Reviews**. 26(5): 531-568; September, 2006.

am Ende, C.W. and et al. "Synthesis and in vitro antimycobacterial activity of B-ring modified diaryl ether InhA inhibitors", **Bioorganic & Medicinal Chemistry Letters**. 18(10): 3029-3033; May, 2008.

Anand, N. and et al. "Synthesis and evaluation of small libraries of triazolylmethoxy chalcones, flavanones and 2-aminopyrimidines as inhibitors of mycobacterial FAS-II and PknG", **Bioorganic & Medicinal Chemistry**. 20(17): 5150-5163; September, 2012.

Andricopulo, A. D., Salum, L. B. and Abraham, D. J. "Structure-based drug design strategies in medicinal chemistry", **Current Topics in Medicinal Chemistry**. 9(9): 771-790, 2009.

Antonini, I. and et al. "Synthesis and antitumor evaluation of bis aza-anthracene-9,10-diones and bis aza-anthrapyrazole-6-ones", **Journal of Medicinal Chemistry**. 51(4): 997-1006; February, 2008.

REFERENCES (CONTINUED)

Aqvist, J., Medina, C. and Samuelsson, J. E. "A new method for predicting binding affinity in computer-aided drug design", **Protein Engineering, Design & Selection**. 7(3): 385-391; March, 1994.

Archontis, G. and et al. "Glycogen phosphorylase inhibitors: a free energy perturbation analysis of glucopyranose spirohydantoin analogues", **Proteins**. 61(4): 984–998; December, 2005.

Avendaño, C. and Menéndez, J. C. "Peptidomimetics in cancer chemotherapy", **Clinical and Translational Oncology**. 9(9): 563-570; September, 2007.

Avendano, C. and Menendez, J.C. **Anticancer Drugs**. Amsterdam: Elsevier, 2007(a).

_____. **Anticancer Drugs**. Amsterdam: Elsevier, 2015(b).

Av-Gay, Y. and Everett, M. "The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*", **Trends in Microbiology**. 8(5): 238-244; May, 2000.

Banhegyi, P.G. **Design and synthesis of benzothiophene and benzothienopyrimidine derivatives with kinase inhibitory activity**. Doctoral (Ph.D.) Thesis: Semmelweis University, 2008.

Barry, C.E. 3rd. and et al. "Mycolic acids: structure, biosynthesis and physiological functions", **Progress in Lipid Research**. 37(2-3): 143-179; July-August, 1998.

Baxter, C.A. and et al. "Flexible docking using Tabu search and an empirical estimate of binding affinity", **Proteins: Structure, Function, and Bioinformatics**. 33(3): 367-382; November, 1998.

Bayly, C.I. and et al. "A well-behaved electrostatic potential based method using charge-restraints for deriving charges: The RESP model", **Journal of Physical Chemistry**. 97(40): 10269–10280; October, 1993.

Benod, C. and et al. "Structure-based discovery of antagonists of nuclear receptor LRH-1", **Journal of Biological Chemistry**. 288(27): 19830–19844; July, 2013.

REFERENCES (CONTINUED)

Bielska E. and et al., "Virtual screening strategies in drug design-methods and applications", **Journal of Biotechnology, Computational Biology and Bionanotechnology**. 92(3): 249–264, October, 2011.

Bordás, B., Komíves, T. and Lopata, A. "Ligand-based computer-aided pesticide design. A review of applications of the CoMFA and CoMSIA methodologies", **Pest Management Science**. 59(4): 393–400; April, 2003.

Borhani, D. W. and Shaw, D. E. "The future of molecular dynamics simulations in drug discovery", **Journal of Computer-Aided Molecular Design**. 26(1): 15-26; January, 2012.

Boshoff, H.I. and Barry, C.E., "Is the mycobacterial cell wall a hopeless drug target for latent tuberculosis" **Drug Discovery Today: Disease Mechanisms**. 3(2): 237-245, Summer, 2006.

Brooijmans, N. and Kuntz, I. D. "Molecular recognition and docking algorithms", **Annual Review of Biophysics and Biomolecular Structure**. 32: 335-373; January, 2003.

Brown, S. B., Brown, E. A. and Walker, I. "The present and future role of photodynamic therapy in cancer treatment", **Lancet Oncology**. 15(8): 497-508; August, 2004.

Bruning, J.B. and et al. "Structure of the *Mycobacterium tuberculosis* d-alanine:d-alanine ligase, a target of the antituberculosis drug d-cycloserine", **Antimicrobial Agents and Chemotherapy**. 55(1): 291-301; January, 2011.

Bryskier, A., "Fluoroquinolones: mechanisms of action and resistance', **International Journal of Antimicrobial Agents**. 2(3): 151-184, 1993.

Burckhardt, G. and et al. "Binding of 2-azaanthraquinone derivatives to DNA and their interference with the activity of DNA topoisomerases *in vitro*", **Biochemistry**. 37(14): 4703-4711; April, 1998.

REFERENCES (CONTINUED)

Caminero, J.A. and et al. "Best drug treatment for multidrug-resistant and extensively drug-resistant tuberculosis", **Lancet Infectious Diseases**. 10(9): 621–629; September, 2010.

Carr, J. H. (2005) "Public Health Image Library (PHIL)", **Centers for Disease Control and Prevention, Office of the Associate Director for Communications, Division of Public Affairs**. http://phil.cdc.gov/phil/details.asp?pid=9997. March 10, 2016.

Calman, K. and Paul, J. **An introduction to cancer medicine**. London: The MacMillan Press. 1978.

Castano, A. P., Demidova, T. N. and Hamblin, M. R. "Mechanisms in photodynamic therapy: part one-photosensitizers, photochemistry and cellular localization", **Photodiagnosis and Photodynamic Therapy**. 1(4): 279-293; December, 2004.

_____. "Mechanisms in photodynamic therapy: part two-cellular signaling, cell metabolism and modes of cell death", **Photodiagnosis and Photodynamic Therapy**. 2(1): 1-23; March, 2005.

Cavalletti, E. and et al. "Pixantrone (BBR 2778) has reduced cardiotoxic potential in mice pretreated with doxorubicin: comparative studies against doxorubicin and mitoxantrone", **Investigational New Drugs**. 25(3): 187-95; June, 2007.

Chaires, J.B. "Drug--DNA interactions", **Current Opinion in Structural Biology**. 8(3): 314-320; June, 1998.

Chakraborti, P.K. and et al. "Signalling mechanisms in Mycobacteria", **Tuberculosis (Edinb)**. 91(5): 432-440; September, 2011.

Chakraborty, S. and et al. "Para-aminosalicylic acid acts as an alternative substrate of folate metabolism in *Mycobacterium tuberculosis*", **Science**. 339(6115): 88-91; January, 2013.

Chao, J. and et al. "Protein kinase and phosphatase signaling in *Mycobacterium tuberculosis* physiology and pathogenesis", **Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics**. 1804(3): 620-627; March, 2010.

REFERENCES (CONTINUED)

Chen Y.C. "Beware of docking!", **Trends in Pharmacological Sciences**. 36(2): 78-95; February, 2015.

Cheng, T. and et al. "Structure based virtual screening for drug discovery: a problem-centric review", **AAPS Journal**. 14(1): 133-141; March, 2012.

Cherkasov, A. and et al., "QSAR modeling: where have you been? Where are you going to?", **Journal of Medicinal Chemistry**. 57(12): 4977–5010; June, 2014.

Chollet, A. and et al. "Synthesis and evaluation of new GEQ derivatives as inhibitors of InhA enzyme and *Mycobacterium tuberculosis* growth", **European Journal of Medicinal Chemistry**. 101: 218-235; August, 2015.

Christian L., Andrew V. and Mario C. R. "New drugs and new regimens for the treatment of tuberculosis: review of the drug development pipeline and implications for national programmes", **Current Opinion in Pulmonary Medicine**. 16(3): 186–193; May, 2010.

Clark, M. and et al. "Comparative Molecular Field Analysis (CoMFA). 2. Toward its use with 3D-structural databases", **Tetrahedron Computer Methodology**. 3(1): 47-59; February, 1990.

Cole, J.C. and et al. "Comparing protein-ligand docking programs is difficult", **Proteins: Structure, Function, and Bioinformatics**. 60(3): 325-332; August, 2005.

Cole, S. T. and et al. "Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence", **Nature**. 393(6685): 537-544; June, 1998.

Collino, F. (2011) "Genes and cancer", **Gene signature profiles in cancer**. http://flipper.diff.org/apptagsaccount/items/3635. March 12, 2016.

Corbeil, C.R., Williams, C.I. and Labute, P. "Variability in docking success rates due to dataset preparation", **Journal of Computer-Aided Molecular Design**. 26(6):775-786; June, 2012.

REFERENCES (CONTINUED)

Cornell, W.D. and et al. "Application of RESP charges to calculation conformational energies, hydrogen bond energies, and free energies of salvation", **Journal of the American Chemical Society**. 115(21): 9620–9631, October, 1993.

Cosconati, S. and et al. "Virtual screening with AutoDock: theory and practice", **Expert Opinion on Drug Discovery**. 5(6): 597-607; June, 2010.

Cozzone, A. J. "Role of protein phosphorylation on serine/threonine and tyrosine in the virulence of bacterial pathogens", **Journal of Molecular Microbiology and Biotechnology**. 9(3-4): 198-213, 2005.

Cramer, R.D., Patterson, D.E. and Bunce, J.D. "Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins", **Journal of the American Chemical Society**. 10(18): 5959–5967; August, 1988.

Darden, T., York, D. and Pedersen, L. "Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems", **Journal of Chemical Physics**. 98(12): 10089–10092; April 1993.

DeBarber, A.E. and et al. "Ethionamide activation and sensitivity in multidrug-resistant *Mycobacterium tuberculosis*", **Proceedings of the National Academy of Sciences of the United States of America**. 97(17): 9677-9682; August, 2000.

Dessen, A. and et al. "Crystal structure and function of the isoniazid target of *Mycobacterium tuberculosis*", **Science**. 267(5204): 1638-1641; March, 1995.

De Vivo and et al. "Role of Molecular Dynamics and Related Methods in Drug Discovery", **Journal of Medicinal Chemistry**. 59(9): 4035-4061; May, 2016.

DiMasi, J.A., Hansen, R.W. and Grabowski, H.G. "The price of innovation: new estimates of drug development costs", **Journal of Health Economics**. 22(2): 151–185; March, 2003.

Dolmans, D.E., Fukumura, D. and Jain, R,K. "Photodynamic therapy for cancer", **Nature Reviews Cancer**. 3(5): 380-387; May, 2003.

REFERENCES (CONTINUED)

Dougherty, T. J. and et al. "Photodynamic therapy", **Journal of the National Cancer Institute**. 90(12): 889-905; June, 1998.

Drlica, K. "The mutant selection window and antimicrobial resistance", **Journal of Antimicrobial Chemotherapy**. 52(1): 11-17; June, 2003(a).

Drlica, K. and et al. "Quinolones: Action and Resistance Updated", **Current Topics in Medicinal Chemistry**. 9(11): 981-998; August, 2009(b).

Dror, R. O. and et al. "Biomolecular simulation: a computational microscope for molecular biology", **Annual Review of Biophysics**. 41: 429-452; June, 2012.

Duan, Y. and et al. "A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations", **Journal of Computational Chemistry**. 24(16): 1999–2012; December, 2003.

Durrant, J. D. and McCammon, J. A. "Molecular dynamics simulations and drug discovery", **BMC Biology**. 9: 71; October, 2011.

Edan, G., Morrissey, S., Le Page, E. "Rationale for the use of mitoxantrone in multiple sclerosis", **Journal of the Neurological Sciences**. 223(1): 35-39; August, 2004.

Eldridge, M.D. and et al. "Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes", **Journal of Computer-Aided Molecular Design**. 11(5): 425-445; September, 1997.

Eliav, E. **Selection of a basis set**. http://www.tau.ac.il/~ephraim/Appendix_1.pdf. March 20, 2016.

Encinas, L. and et al. "Encoded library technology as a source of hits for the discovery and lead optimization of a potent and selective class of bactericidal direct inhibitors of *Mycobacterium tuberculosis* InhA", **Journal of Medicinal Chemistry**. 57(4): 1276-1288; February, 2014.

REFERENCES (CONTINUED)

Engert, A. and et al. "Pegfilgrastim support for full delivery of BEACOPP-14 chemotherapy for patients with high-risk Hodgkin's lymphoma: results of a phase II study", **Haematologica**. 91(4): 546-549; April, 2006.

Ertl, P., Rohde, B. and Selzer, P. "Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties.", **Journal of Medicinal Chemistry**. 43(20): 3714-3717: October, 2000.

Fàbrega, A. and et al. "Mechanism of action of and resistance to quinolones", **Microbial Biotechnology**. 2(1): 40–61; January, 2009.

Ferreira, L.G. and et al. "Molecular docking and structure-based drug design Strategies", **Molecules**. 20(7): 13384-13421; June, 2015.

Fillion, A. et al. "Impact of fluoroquinolone resistance on bactericidal and sterilizing activity of a moxifloxacin-containing regimen in murine tuberculosis", **Antimicrobial Agents and Chemotherapy**. 57(9): 4496-4500; September, 2013.

Foloppe, N. and Hubbard, R. "Towards predictive ligand design with free-energy based computational methods?", **Current Medicinal Chemistry**. 13(29): 3583-3608, 2006.

Forli, S. and Olson, A. J. "A force field with discrete displaceable waters and desolvations entropy for hydrated ligand docking", **Journal of Medicinal Chemistry.** 55(2): 623-638; January, 2012.

Fox, M. E. and Smith, P. J. "Long-term inhibition of DNA synthesis and the persistence of trapped topoisomerase II complexes in determining the toxicity of the antitumor DNA intercalators mAMSA and mitoxantrone", **Cancer Research**. 50(18): 5813-5818; September, 1990.

Fox, W., Ellard, G.A. and Mitchison, D.A. "Studies on the treatment of tuberculosis undertaken by the British Medical Research Council tuberculosis units, 1946-1986, with relevant subsequent publications", **International Journal of Tuberculosis and Lung Disease**. 3(10 Suppl 2): S231-279; October, 1999 .

REFERENCES (CONTINUED)

Freundlich, J.S. and et al. "Triclosan derivatives: towards potent inhibitors of drug-sensitive and drug-resistant *Mycobacterium tuberculosis*", **ChemMedChem**. 4(2): 241-248; Febuary, 2009.

Friesner, R.A. and et al. "Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy", **Journal of Medicinal Chemistry**. 47(7): 1739-1749; March, 2004(a).

_____. "Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes", **Journal of Medicinal Chemistry**. 49(21): 6177-6196, October, 2006(b).

Frishman, W. H. and et al. "Cardiovascular toxicity with cancer chemotherapy", **Current Problems in Cancer**. 21(6): 301-360; November-December, 1997.

Gavernet, L. and et al. "Affinity of sulfamates and sulfamides to carbonic anhydrase II isoform: Experimental and molecular modeling approaches", **Journal of Chemical Information and Modeling**. 50 (6): 1113–1122; January, 2010.

Gengenbacher, M. and Kaufmann S.H., "Mycobacterium tuberculosis: success through dormancy", **FEMS Microbiology Reviews**. 36(3): 514-532; May, 2012.

Genheden, S. and Ryde, U. "The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities", **Expert Opinion on Drug Discovery**. 10(5): 449-461, May, 2015.

Gilson, M. K. and Zhou, H. X. "Calculation of protein-ligand binding affinities", **Annual Review of Biophysics and Biomolecular Structure**. 36, 21-42, 2007.

Ginsberg, A.M. and Spigelman, M., "Challenges in tuberculosis drug research and development", **Nature Medicine**. 13(3): 290-294; March, 2007.

Glide (2010) "A complete solution for ligand-receptor docking", **Schrodinger**. https://isp.ncifcrf.gov/files/isp/uploads/2010/07/glide.pdf. March 15, 2016.

REFERENCES (CONTINUED)

Global Alliance for TB Drug Development. "Handbook of Anti-Tuberculosis Agents", **Tuberculosis**. 88(2):1-169; March, 2008.

GLOBOCAN2012 (2012) Cancer Fact Sheets", **International agency for research on cancer**. http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx. March 1, 2016.

Gohlke, H. and Klebe, G. "Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors", **Angewandte Chemie International Edition in English**. 41(15): 2644–2676; August, 2002.

Golbraikh, A. and Tropsha, A. "Beware of q2!", **Journal of Molecular Graphics and Modelling.** 20(4): 269-276; January, 2002.

Gonsette, R.E. "Compared benefit of approved and experimental immunosuppressive therapeutic approaches in multiple sclerosis", **Expert Opinion on Pharmacotherapy**. 8(8): 1103-1116; June, 2007.

González, M.A. "Force fields and molecular dynamics simulations", **Collection SFN**. 12: 169–200; June, 2011.

Goodford, P. J. "A computational procedure for determining energetically favorable binding sites on biologically important macromolecules", **Journal of Medicinal Chemistry.** 28(7): 849-857; June, 1985.

Guardia, A. and et al. "*N*-Benzyl-4-((heteroaryl)methyl)benzamides: A new class of direct NADH-dependent 2-*trans* enoyl-acyl carrier protein reductase (InhA) inhibitors with antitubercular activity", **ChemMedChem**. 11(7): 687-701; April, 2016.

Halgren, T.A. and et al. "Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening", **Journal of Medicinal Chemistry**. 47(7): 1750-1759; March, 2004.

Hansson, T., Oostenbrink, C. and van Gunsteren, W. "Molecular dynamics simulations", **Current Opinion in Structural Biology**. 12(2): 190-196; April, 2002.

REFERENCES (CONTINUED)

Hartkoorn, R.C. and et al. "Towards a new tuberculosis drug: pyridomycin - nature's isoniazid", **EMBO Molecular Medicine**. 4(10): 1032-1042; October, 2012.

Hayes, J. M. and Leonidas, D. D. "Computation as a tool for glycogen phosphorylase inhibitor design", **Mini Reviews in Medicinal Chemistry**. 10(12): 1156-1174; October, 2010.

Hayik, S. A., Dunbrack, R. Jr. and Merz, K. M. Jr. "A mixed QM/MM Scoring function to predict protein-ligand binding affinity", **Journal of Chemical Theory and Computation**. 6(10): 3079-3091; September, 2010.

He, X. and et al. "Pyrrolidine carboxamides as a novel class of inhibitors of enoyl acyl carrier protein reductase from *Mycobacterium tuberculosis*", **Journal of Medicinal Chemistry**. 49(21): 6308-6323; October, 2006.

He, X., Alian, A. and Ortiz de Montellano, P.R. "Inhibition of the *Mycobacterium tuberculosis* enoyl acyl carrier protein reductase InhA by arylamides", **Bioorganic & Medicinal Chemistry**. 15(21): 6649-6658; November, 2007.

Hegymegi-Barakonyi, B. and et al. "Signalling inhibitors against *Mycobacterium tuberculosis*--early days of a new therapeutic concept in tuberculosis", **Current Medicinal Chemistry**. 15(26): 2760-2770, 2008.

Hehre, W.J. **A Guide to Molecular Mechanics and Quantum Chemical Calculations**. Irvine, California: Wavefunction, Inc, 2003.

Heifetz, A. and et al. "From receptors to ligands: fragment-assisted drug design for GPCRs applied to the discovery of H3 and H4 receptor antagonists", **Medicinal Chemistry**. 4(1): 313-321; December, 2013.

Homeyer N. and Gohlke, H. "Free energy calculations by the molecular mechanics Poisson–Boltzmann surface area method", **Molecular Informatics**. 31(2): 114–122; February, 2012.

Hongmao, S. **A Practical Guide to Rational Drug Design**. Amsterdam: Elsevier, 2016.

REFERENCES (CONTINUED)

Hopper, C. "Photodynamic therapy: a clinical reality in the treatment of cancer", **Lancet Oncology**. 1: 212-219; December, 2000.

Hou, T. and et al. "Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations", **Journal of Chemical Information and Modeling**. 51(1): 69–82; January, 2011.

Huang, S.Y. and Zou, X. "Advances and challenges in protein-ligand docking", **International Journal of Molecular Sciences**. 11(8): 3016-3034; August, 2010.

Huey, R. and et al. "Grid-based hydrogen bond potentials with improved directionality", **Letters in Drug Design and Discovery**. 1: 178-183; April, 2004 (a).

_____. "A semiempirical free energy force field with charge-based desolvation", **Journal of Computational Chemistry**. 28: 1145-1152; April, 2007(b).

Hughes, J.P. and et al. "Principles of early drug discovery", **British Journal of Pharmacology**. 162: 1239–1249; March, 2011.

Ihmels, H. and Thomas, L. **In Materials Science of DNA Chemistry**. Boca Raton: CRC Press, 2011.

Ioerger, T.R. and et al. "Identification of new drug targets and resistance mechanisms in *Mycobacterium tuberculosis*", **PLoS One**. 8(9): e75245; September, 2013.

ISN Insider (2013) **New Medicines Have a Long Road to the Medicine Cabinet**. https://www.isnetworld.com/~isn/blog/2013/05/06/new-medicines-have-a-long-road-to-the-medicine-cabinet/. March 7, 2016.

Jassal, M. and Bishai, W.R. "Extensively drug-resistant tuberculosis", **Lancet Infectious Diseases**. 9(1): 19–30; January, 2009.

Johansen, S.K. and et al. "Capreomycin binds across the ribosomal subunit interface using tlyA-encoded 2'-O-methylations in 16S and 23S rRNAs", **Molecular Cell**. 23(2):173-182, July, 2006.

REFERENCES (CONTINUED)

Jones, G., Willett, P. and Glen, R.C. "Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation", **Journal of Molecular Biology**. 245(1): 43-53; January, 1995.

Jones, G. and et al. "Development and validation of a genetic algorithm for flexible docking", **Journal of Molecular Biology**. 267(3): 727-748; April, 1997.

Jorgensen W.L. "The many roles of computation in drug discovery", **Science**. 303(5665):1813-1818; March 19, 2004.

Kaledin, M. and et al. "Normal mode analysis using the driven molecular dynamics method. II. An application to biological macromolecules", **Journal of Chemical Physics**. 121(12): 5646–5653; September, 2004.

Kalyaanamoorthy, S. and Chen, Y. P. "Structure-based drug design to augment hit discovery", **Drug Discovery Today**. 16(17-18): 831-839; September, 2011.

Kana, B.D. and et al. "Future target-based drug discovery for tuberculosis", **Tuberculosis**. 94(6): 551-556; December, 2014.

Kanetaka, H. and et al. "Discovery of InhA inhibitors with anti-mycobacterial activity through a matched molecular pair approach", **European Journal of Medicinal Chemistry**. 94: 378-385; April, 2015.

Kapetanovic, I.M. "Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach", **Chemico-Biological Interactions**. 171(2): 165–176; January, 2008.

Kar, S. and Roy, K. "How far can virtual screening take us in drug discovery?", **Expert Opinion on Drug Discovery**. 8(3): 245-261; March, 2013.

Karaman, B. and Sippl, W. "Docking and binding free energy calculations of sirtuin inhibitors", **European Journal of Medicinal Chemistry**. 93: 584-598; March, 2015.

Karplus, M. and McCammon, J. A. "Molecular dynamics simulations of biomolecules", **Nature Structural & Molecular Biology**. 9(9): 646-652; September, 2002.

REFERENCES (CONTINUED)

Kenneth, C. C. and John, P. **An Introduction to Cancer Medicine**. London: The macmillan press ltd, 1978

King, R.J.B. and Robins, M.W. **Cancer Biology**. London: Pearson Education Limited, 2006.

Kinjo, T. and et al. "Identification of compounds with potential antibacterial activity against Mycobacterium through structure-based drug screening", **Journal of Chemical Information and Modeling**. 53(5): 1200-1212; May, 2013.

Kirkwood, J.G. "Statistical mechanics of fluid mixtures", **Journal of Chemical Physics.**, 3(5): 300–313; January, 1935.

Kitchen, D. B. and et al. "Docking and scoring in virtual screening for drug discovery: methods and applications", **Nature Reviews Drug Discovery**. 3(11): 935-949; November, 2004.

Kitchen, D. B., Stahura, F. L. and Bajorath, J. "Computational techniques for diversity analysis and compound classification", **Mini Reviews in Medicinal Chemistry**. 4(10): 1029-1039; December, 2004.

Klebe, G. "Virtual ligand screening: strategies, perspectives and limitations", **Drug Discovery Today**. 11(13–14): 580–594; July, 2006.

Klebe, G., Abraham, U. and Mietzner, T. "Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity", **Journal of Medicinal Chemistry**. 37(24):4130-4146; November, 1994.

Kolb, P. and et al., "Structure-based discovery of β2-adrenergic receptor ligands", **Proceedings of the National Academy of Sciences of the United States of America**. 106: 6843–6848; 2009.

Koul, A. and et al. "The challenge of new drug discovery for tuberculosis", **Nature**. 469(7331): 483-490; January, 2011.

Krapcho, A. P. "6,9-Bis[(aminoalkyl)amino]benzo[g]isoquinoline-5,10-diones. A novel class of chromophore-modified antitumor anthracene-9,10-diones: synthesis and antitumor evaluations", **Journal of Medicinal Chemistry**. 37(6): 828-837; March, 1994.

REFERENCES (CONTINUED)

Krapcho, A. P. and et al. "Aza and diaza bioisosteric anthracene-9,10-diones as antitumor agents", **Acta Biochimica Polonica**. 42(4): 427-432; 1995 (a).

_____. "Synthesis and antitumor evaluation of 2,5-disubstituted-indazolo[4, 3-gh]isoquinolin-6(2*H*)-ones (9-aza-anthrapyrazoles)", **Journal of Medicinal Chemistry**. 41(27): 5429-5444; December, 1998(b).

Krumrine, J. and et al. "Principles and methods of docking and ligand design", **Methods of Biochemical Analysis**. 44: 443-76; 2003.

Kubelka, J. "Quantum chemical calculations: the methods.", **Molecular Modeling - CHEM 4560/5560**. http://www.uwyo.edu/kubelka-chem/molecular_modeling_notes_9.pdf. March 25, 2016.

Kubinyi, H. **QSAR: Hansch Analysis and Related Approaches**. Weinheim, New York, Basil, Cambridge, Tokyo: VCH; 1993(a).

_____. (1998) "Comparative molecular field analysis (CoMFA)", **Encyclopedia of computational chemistry**. http://www.wiley.com/legacy/wileychi/ecc/samples/sample05.pdf. March 15, 2016(b).

_____. "Success stories of computer-aided design", **Computer Applications in Pharmaceutical Research and Development**. Hoboken: John Wiley & Sons, Inc., 2006(c).

Kumar, U.C. and et al. "Discovery of novel InhA reductase inhibitors: application of pharmacophore- and shape-based screening approach", **Future Medicinal Chemistry**. 5(3): 249–259; March, 2013.

Kumar, U.C. and Shaik, M. "3-D QSAR CoMSIA Models of arylamides for prediction of enoyl acyl carrier protein reductase inhibitory activity", **Journal of Pharmaceutical Science and Technology**. 3: 536–542; 2011.

Kuo, M.R. and et al. "Targeting tuberculosis and malaria through inhibition of Enoyl reductase: compound activity and structural data", **Journal of Biological Chemistry**. 278(23): 20851-20859; January, 2003.

REFERENCES (CONTINUED)

Lavecchia, A. and Di Giovanni, C. "Virtual screening strategies in drug discovery: a critical review", **Current Medicinal Chemistry**. 20(23): 2839-2860, 2013.

Lei, H., Wang, X. and Wu, C. "Early stage intercalation of doxorubicin to DNA fragments observed in molecular dynamics binding simulations", **Journal of Molecular Graphics and Modelling**. 38: 279-289; September, 2012.

Li, J. and et al. "Association of gyrA/B mutations and resistance levels to fluoroquinolones in clinical isolates of *Mycobacterium tuberculosis*", **Emerging Microbes & Infections**. 3(3): e19; March, 2014.

Li, W. and et al. "Unbinding pathways of GW4064 from Human Farnesoid X Receptor as revealed by molecular dynamics simulations", **Journal of Chemical Information and Modeling**. 52(11): 3043–3052; November, 2012.

Li, X. L. and et al. "Molecular spectroscopy evidence of berberine binding to DNA: comparative binding and thermodynamic profile of intercalation", **Biomacromolecules**. 13(3): 873-880; March, 2012.

Liang, J., Edelsbrunner, H. and Woodward, C. "Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design", **Protein Science**. 7(9): 1884-1897; September, 1998.

Lienhardt, C., Vernon, A. and Raviglione, M.C. "New drugs and new regimens for the treatment of tuberculosis: review of the drug development pipeline and implications for national programmes", **Current Opinion in Pulmonary Medicine**. 16(3): 186-193, May, 2010.

Lionta, E. and et al. "Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances", **Current Topics in Medicinal Chemistry**. 14(16): 1923–1938, 2014.

Liu, J. and Wang, R. "Classification of current scoring functions", **Journal of Chemical Information and Modeling**. 55(3): 475-482; March, 2015.

REFERENCES (CONTINUED)

Lu, C. and et al. "Novel biomarkers distinguishing active tuberculosis from latent infection identified by gene expression profile of peripheral blood mononuclear cells.", **PLoS ONE**. 6: e24290; August, 2011.

Lu. X, Huang, K. and You, Q. "Enoyl acyl carrier protein reductase inhibitors: a patent review (2006 - 2010)", **Expert Opinion on Therapeutic Patents**. 21(7): 1007-1022; July, 2011.

Luckner, S.R. and et al. "A slow, tight binding inhibitor of InhA, the enoyl-acyl carrier protein reductase from *Mycobacterium tuberculosis*", Journal of **Biological Chemistry**. 285(19): 14330-14337; May, 2010.

Ma, Z. and et al., "Global tuberculosis drug development pipeline: the need and the reality", **Lancet**. 375(9731): 2100-2109; June, 2010.

Mahoney M.W. and Jorgensen, W.L. "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions", **Journal of Chemical Physics**. 112(20): 8910–8922, 2000.

Mandal S., Moudgil, M. and Mandal, S.K. "Rational drug design", **European Journal of Pharmacology**. 625(1-3): 90–100; December, 2009.

Manjunatha, U. H. and et al. "Direct inhibitors of InhA are active against *Mycobacterium tuberculosis*", **Science Translational Medicine**. 7(269): 269ra3; January, 2015.

Manjunatha, U. H. and Smith, P. W. "Perspective: challenges and opportunities in TB drug discovery from phenotypic screening", **Bioorganic & Medicinal Chemistry**. 23(16): 5087-5097; August, 2015.

Martis, E.A. and Somani, R.R. "Drug designing, discovery and development techniques", **Promising Pharmaceuticals**. 19-36, 2012.

Mehler, E. L. and Solmajer, T. "Electrostatic effects in proteins: comparison of dielectric and charge models", **Protein Engineering, Design & Selection**. 4(8): 903-910; December, 1991.

Mellman, I., Coukos, G. and Dranoff, G. "Cancer immunotherapy comes of age", **Nature**. 480(7378): 480-489; December, 2011.

REFERENCES (CONTINUED)

Meng, X. Y. and et al. "Molecular docking: a powerful approach for structure-based drug discovery", **Current Computer - Aided Drug Design**. 7(2): 146-157; January, 2011.

Miller, B.R. 3rd. and et al. "MMPBSA.py: An efficient program for end-state free energy calculations", **Journal of Chemical Theory and Computation**. 8(9): 3314-3321; September, 2012.

Mongan, J. and et al. "Generalized born model with a simple, robust molecular volume correction", **Journal of Chemical Theory and Computation**. 3(1): 156–169; January, 2007.

Mooij, W.T. and Verdonk, M.L. "General and targeted statistical potentials for protein-ligand interactions", **Proteins: Structure, Function, and Bioinformatics**. 61(2): 272-287, November, 2005.

Morlock, G.P. and et al. "ethA, inhA, and katG loci of ethionamide-resistant clinical Mycobacterium tuberculosis isolates", **Antimicrobial Agents and Chemotherapy**. 47(12): 3799-3805; December, 2003.

Morris, G. M. and et al. "Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function", **Journal of Computational Chemistry**. 19(14): 1639-1662; June, 1998.

Morris, G. M. and et al. (2001) "Automated docking of flexible ligands to receptors", **AutoDock User's Guide**. http://autodock.scripps.edu/faqs-help/manual/autodock-3-user-s-guide/AutoDock3.0.5_UserGuide.pdf. March 15, 2016.

Muegge, I. and Oloff, S. "Advances in virtual screening", **Drug Discovery Today Technologies**. 3: 405–411; December, 2006.

Mukherji, D. and Pettengell, R. "Pixantrone maleate for non-Hodgkin's lymphoma", **Drugs Today (Barc)**. 45(11): 797-805; November, 2009.

Nair, P. C. and Miners, J. O. "Molecular dynamics simulations: from structure function relationships to drug discovery", **In Silico Pharmacology**. 21: 24; November, 2014.

REFERENCES (CONTINUED)

Nakamoto, K., Tsuboi, M. and Strahan, G.D. "Drug-DNA interactions: structures and spectra", **Methods of Biochemical Analysis**. 51: 1-366; 2008.

Narayan, A. and et al. "Serine threonine protein kinases of mycobacterial genus: phylogeny to function", **Physiol Genomics**. 29(1): 66-75; March, 2007.

Ng P. S. and et al. "Structure activity relationships of 4-hydroxy-2-pyridones: A novel class of antituberculosis agents", **European Journal of Medicinal Chemistry**. 106: 144-156; December, 2015.

Nguyen, L. and Pieters, J. "The Trojan horse: survival tactics of pathogenic mycobacteria in macrophages", **Trends in Cell Biology**. (5): 269-276; May, 2005.

Nguyen, L. and et al. "Role of protein kinase G in growth and glutamine metabolism of *Mycobacterium bovis* BCG", **Journal of Bacteriology**. 187(16): 5852-6; 2005.

Nissink, J.W. and et al. "A new test set for validating predictions of protein-ligand interaction", **Proteins: Structure, Function, and Bioinformatics**. 49(4): 457-471; December, 2002.

Norberg, J. and Nilsson, L. "Advances in biomolecular simulations: methodology and recent applications", **Quarterly Reviews of Biophysics**. 36(3): 257-306; January, 2003.

Norbury, C. and Nurse, P. "Animal cell cycles and their control", **Annual Review of Biochemistry**. 61: 441-470, 1992.

Nunes-Alves C. and et al. "In search of a new paradigm for protective immunity to TB", **Nature Reviews Microbiology**. 12(4): 289-299; April, 2014.

Olson, A.J. and et al. (2012) "Automated docking of flexible ligands to flexible receptors", **AutoDock Version 4.2 User Guide**. http://autodock.scripps.edu/faqs-help/manual/autodock-4-2-user-guide/AutoDock4.2_UserGuide.pdf. March 15, 2016.

REFERENCES (CONTINUED)

Onufriev, A., Bashford, D. and Case D.A. "Exploring protein native states and large-scale conformational changes with a modified generalized born model", **Proteins: Structure, Function, and Bioinformatics**. 55(2): 383–394; May, 2004.

Ooms, F. "Molecular modeling and computer aided drug desig. Examples of their applications in medicinal chemistry", **Current Medicinal Chemistry**. 7(2): 141-58; February, 2000.

Pajk, S. and et al. "New direct inhibitors of InhA with antimycobacterial activity based on a tetrahydropyran scaffold", **European Journal of Medicinal Chemistry**. 112: 252-257; April, 2016.

Palchaudhuri, R. and Hergenrother, P. J. "DNA as a target for anticancer compounds: methods to determine the mode of binding and the mechanism of action", **Current Opinion in Biotechnology**. 18(6): 497-503; December, 2007.

Pan, P. and et al. "Time-dependent diaryl ether inhibitors of InhA: structure-activity relationship studies of enzyme inhibition, antibacterial activity, and in vivo efficacy", **ChemMedChem**. 9(4): 776-791; April, 2014.

Paquet, E. and Viktor, H. L. "Molecular dynamics, monte carlo simulations, and langevin dynamics: a computational review", **BioMed Research International**. 183918; February, 2015

Parikh, S. and et al. "Roles of tyrosine 158 and lysine 165 in the catalytic mechanism of InhA, the enoyl-ACP reductase from *Mycobacterium tuberculosis*", **Biochemistry**. 38(41): 13623-13634; October, 1999.

Parikh, S.L., Xiao, G. and Tonge, P.J. "Inhibition of InhA, the enoyl reductase from *Mycobacterium tuberculosis*, by triclosan and isoniazid", **Biochemistry**. 39(26): 7645-7650; July, 2000.

Pass, H. I. "Photodynamic therapy in oncology: mechanisms and clinical use", **Journal of the National Cancer Institute**. 85(6): 443-456; March, 1993.

REFERENCES (CONTINUED)

Pauli, I. and et al. "Discovery of new inhibitors of *Mycobacterium tuberculosis* InhA enzyme using virtual screening and a 3D-pharmacophore-based approach", **Journal of Chemical Information and Modeling**. 53(9): 2390-2401; September, 2013.

Pearlman, D.A. "Evaluating the molecular mechanics Poisson–Boltzmann surface area free energy method using a congeneric series of ligands to p38 MAP kinase", **Journal of Medicinal Chemistry**. 48 (24): 7796–7807; December, 2005.

Pedgaonkar, G.S. and et al. "Development of benzo[d]oxazol-2(3*H*)-ones derivatives as novel inhibitors of *Mycobacterium tuberculosis* InhA", **Bioorganic & Medicinal Chemistry**. 22(21): 6134–6145; November, 2014(a).

_____. "Development of 2-(4-oxoquinazolin-3(4*H*)-yl)acetamide derivatives as novel enoyl-acyl carrier protein reductase (InhA) inhibitors for the treatment of tuberculosis", **European Journal of Medicinal Chemistry**. 86: 613-627; October, 2014(b).

Perilla, J. R. and et al. "Molecular dynamics simulations of large macromolecular complexes", **Current Opinion in Structural Biology**. 31: 64-74; April, 2015.

Pharmaceutical Research and Manufacturers of America (PhRMA). **Biopharmaceutical Research & Development: The Process Behind New Medicines**. http://www.phrma.org/sites/default/files/pdf/rd_brochure_022307.pdf. March 3, 2016.

Philip, R., "A unified classification of cancers: An oncotaxonomy with symbols", **Cancer**. 31(4): 963-982; April, 1972.

Prosser, G.A. and de Carvalho, L.P. "Reinterpreting the mechanism of inhibition of *Mycobacterium tuberculosis* D-alanine:D-alanine ligase by D-cycloserine", **Biochemistry**. 52(40): 7145-7149; October, 2013.

REFERENCES (CONTINUED)

Raha, K. and Merz, K. M. Jr. "A quantum mechanics-based scoring function: study of zinc ion-mediated ligand binding", **Journal of the American Chemical Society**. 126(4): 1020-1021; February, 2004.

Raviglione, M.C. and Smith, I.M., "XDR tuberculosis--implications for global public health", **The New England Journal of Medicine**. 356(7): 656-659, 2007.

Rescifina, A. and et al. "Recent advances in small organic molecules as DNA intercalating agents: synthesis, activity, and modeling", **European Journal of Medicinal Chemistry**. 74: 95-115; March, 2014.

Ripphausen, P., Nisius, B. and Bajorath, J. "State-of-the-art in ligand-based virtual screening", **Drug Discovery Today**. 16(9–10): 372–376; May, 2011.

Rizzo, V., Sacchi, N. and Menozzi, M. "Kinetic studies of anthracycline-DNA interaction by fluorescence stopped flow confirm a complex association mechanism", **Biochemistry**. 28(1): 274-282; January, 1989.

Roman, S. and et al. "Design, synthesis, and evaluation of new thiadiazole-based direct Inhibitors of enoyl acyl carrier protein reductase (InhA) for the treatment of tuberculosis", **Journal of Medicinal Chemistry**. 58(2): 613-624; January, 2015.

Pongprom, N. and et al. "Synthesis of anticancer compounds carbinol derivatives of azanaphthoquinone annelated pyrroles", **Monatshefte für Chemie**. 140: 309–313; September, 2009.

_____. "Synthesis of new Benzo[f]isoindole-4,9-diones as anticancer compounds", **Monatshefte für Chemie**. 141(1): 53–62; January, 2010.

Pongprom, N. and Pungpo, P. "Synthesis and cytotoxic activities of azanaphthoquinone annelated pyrrolo hydrazone derivatives", **KKU Research Journal**. 17(2): 195-202; March, 2012.

Rosenberg, S. A., Yang, J. C. and Restifo, N. P. "Cancer immunotherapy: moving beyond current vaccines", **Nature Medicine**. 10(9): 909-915; September, 2004.

REFERENCES (CONTINUED)

Rozwarski, D.A. and et al. "Modification of the NADH of the isoniazid target (InhA) from *Mycobacterium tuberculosis*", **Science**. 279(5347): 98-102; January, 1998(a).

_____. "Crystal structure of the *Mycobacterium tuberculosis* enoyl-ACP reductase, InhA, in complex with NAD^+ and a C16 fatty acyl substrate", **Journal of Biological Chemistry**. 274(22): 15582-15589; May, 1999(b).

Rubin, P. "A unified classification of cancers: an oncotaxonomy with symbols", **Cancer**. 31(4): 963-982; April, 1973.

Ruddon, R.W. **Cancer Biology**. Oxford: Oxford University Press, Inc, 2007.

Ryckaert, J.P. Ciccotti, G. and Berendsen, H.J.C. "Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-Alkanes", **Journal of Computational Physics**. 23(3): 327-341; March, 1977.

Sagui, C. and Darden, T.A. "Molecular dynamics simulations of biomolecules: long-range electrostatic effects", **Annual review of biophysics and biomolecular structure**. 28: 155-179; June, 1999.

Sakamoto, K., "The Pathology of *Mycobacterium tuberculosis* Infection", **Veterinary Pathology**. 49(3): 423-439; May, 2012.

Salian, S. and et al. "Structure-activity relationships among the kanamycin aminoglycosides: Role of ring I hydroxyl and amino groups", **Antimicrobial Agents and Chemotherapy**. 56(12): 6104-6108; May, 2012.

Schatz, A., Bugie, E. and Waksman, S.A. "Streptomycin, a substance exhibiting antibiotic activity against gram positive and gram-negative bacteria", **Experimental Biology and Medicine**. 55(1): 66-69; January, 1944.

Scherr, N. and et al. "Structural basis for the specific inhibition of protein kinase G, a virulence factor of *Mycobacterium tuberculosis*", **Proceedings of the National Academy of Sciences of the United States of America**. 104(29): 12151-12156; July, 2007.

REFERENCES (CONTINUED)

Schröder, J. and et al., "Docking-based virtual screening of covalently binding ligands: an orthogonal lead discovery approach", **Journal of Medicinal Chemistry**. 56(4): 1478–1490; February, 2013.

Schuitmaker, J. J. and et al. "Photodynamic therapy: a promising new modality for the treatment of cancer", **Journal of Photochemistry and Photobiology B: Biology**. 34(1): 3-12; June, 1996.

Shanab, K. and et al. "Synthesis and biological evaluation of novel cytotoxic azanaphthoquinone annelated pyrrolo oximes", **Bioorganic & Medicinal Chemistry Letters**. 17(22): 6091-6095; November, 2007(a).

_____. "Synthesis and biological evaluation of new cytotoxic azanaphthoquinone pyrrolo-annelated derivatives", **Bioorganic & Medicinal Chemistry Letters**. 20(13): 3950-3952; June, 2010(b).

_____. "Synthesis and antiproliferative activity of new cytotoxic azanaphthoquinone pyrrolo-annelated derivatives: Part II", **Bioorganic & Medicinal Chemistry Letters**. 21(10): 3117-3121; May, 2011(c).

Shchekotikhin, A. E. and et al. "Synthesis and cytotoxic properties of 4,11-bis[(aminoethyl)amino]anthra[2,3-b]thiophene-5,10-diones, novel analogues of antitumor anthracene-9,10-diones", **Bioorganic & Medicinal Chemistry**. 17(5): 1861-1869; March, 2009.

Shi, W. and et al. "Pyrazinamide Inhibits trans-translation in *Mycobacterium tuberculosis*", **Science**. 333 (6049): 1630-1632; September, 2011.

Shirude, P. S. and et al. "Methyl-thiazoles: a novel mode of inhibition with the potential to develop novel inhibitors targeting InhA in *Mycobacterium tuberculosis*", **Journal of Medicinal Chemistry**. 56(21): 8533-8542; November, 2013.

Sink, R. and et al. "Design, synthesis, and evaluation of new thiadiazole-based direct inhibitors of enoyl acyl carrier protein reductase (InhA) for the treatment of tuberculosis", **Journal of Medicinal Chemistry**. 58(2): 613-624; January, 2015.

REFERENCES (CONTINUED)

Singh, N. and et al. "Identification of novel inhibitors of *Mycobacterium tuberculosis* PknG using pharmacophore based virtual screening, docking, molecular dynamics simulation, and their biological evaluation", **Journal of Chemical Information and Modeling**. 55(6): 1120-1129; June, 2015.

Sipos, A. and et al. "Lead selection and characterization of antitubercular compounds using the Nested Chemical Library", **Tuberculosis (Edinb)**. 95: S200-206; June, 2015.

Sirgel, F.A. and et al. "Mutations in the rrs A1401G gene and phenotypic resistance to amikacin and capreomycin in Mycobacterium tuberculosis", **Microbial Drug Resistance**. 18(2): 193-197; July, 2011.

Sissi, C. and Palumbo, M. "Antitumor potential of aza-bioisosterism in anthracenedione-based drugs", **Current Topics in Medicinal Chemistry**. 4(2): 219-230, 2004.

Skylaris, C.K. "Gaussian basis sets", **CHEM6085: Density Functional Theory**. http://www.southampton.ac.uk/assets/centresresearch/documents/compchem/DFT_L8.pdf. March 20, 2016.

Slepikas, L. and et al. (2016) "In silico driven design and synthesis of rhodanine derivatives as novel antibacterials targeting the enoyl reductase InhA", **Journal of Medicinal Chemistry**. http://pubs.acs.org/doi/pdf/10.1021/acs.jmedchem.5b01620 , March 6, 2016.

Sloan, D.J., Davies, G.R. and Khoo, S.H. "Recent advances in uberculosis: New drugs and treatment regimens", **Current Respiratory Medicine Reviews**. 9(3): 200–210; January, 2013.

Snow. C. D. and et al. "How well can simulation predict protein folding kinetics and thermodynamics?", **Annual Review of Biophysics and Biomolecular Structure**. 34: 43-69, 2005.

REFERENCES (CONTINUED)

Slynko, I. and et al. "Virtual screening of PRK1 inhibitors: ensemble docking, rescoring using binding free energy calculation and QSAR model development", **Journal of Chemical Information and Modeling**. 54(1):138-150; January, 2014.

Sousa, S. F., Fernandes, P. A. and Ramos, M. J. "Protein-ligand docking: current status and future challenges". **Proteins: Structure, Function, and Bioinformatics**. 65(1): 15-26; October, 2006.

Sousa, S.F. and et al. "Protein-Ligand Docking in the New Millennium – A Retrospective of 10 Years in the Field", **Current Medicinal Chemistry**. 20: 2296-2314, 2013.

Spreitzer, H. and et al. "Synthesis of azanaphthoquinone annelated pyrroles", **Heterocycles**. 54(1): 111-121; 2001.

Srinivasan, J. and et al. "Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate–DNA helices", **Journal of the American Chemical Society**. 120 (37): 9401–9409; August, 1998.

Stec, J. and et al. "Biological evaluation of potent triclosan-derived inhibitors of the enoyl-acyl carrier protein reductase InhA in drug-sensitive and drug-resistant strains of *Mycobacterium tuberculosis*", **ChemMedChem**. 9(11): 2528-2537; November, 2014.

Straatsma, T.P. and McCammon J.A. "Multiconfiguration thermodynamic integration", **Journal of Chemical Physics**. 95 (2): 1175–1188, 1991.

Sullivan, T.J. and et al. "High affinity InhA inhibitors with activity against drug-resistant strains of *Mycobacterium tuberculosis*", **ACS Chemical Biology**. 1(1): 43-53; February, 2006.

Sumit, C. and et al. "Para-aminosalicylic acid acts as an alternative substrate of folate metabolism in *Mycobacterium tuberculosis*", **Science**. 339: 88-91; January, 2013.

Sun, H. "Pharmacophore-Based Virtual Screening", **Current Medicinal Chemistry**. 15(10): 1018-1024, 2008.

REFERENCES (CONTINUED)

Szabo, A. and Ostlund, N.S. **Modern Quantum Chemictry-Introduction to Advance Electronic Structure Theory**. New York: McGraw-Hill, Inc, 1989.

Székely, R. and et al. "A novel drug discovery concept for tuberculosis: inhibition of bacterial and host cell signaling", **Immunology Letters**. 116(2): 225-231; March, 2008.

Takayama, K., Wang, C. and Besra, G.S. "Pathway to synthesis and processing of mycolic acids in Mycobacterium tuberculosis.", **Clinical Microbiology Reviews**. 18(1): 81-101; January, 2005.

Talele, T.T., Khedkar, S.A. and Rigby, A.C. "Successful applications of computer aided drug discovery: Moving drugs from concept to the clinic", **Current Topics in Medicinal Chemistry**. 10: 127-141, 2010.

Tang, J., Yam, W.C. and Chen, Z. "*Mycobacterium tuberculosis* infection and vaccine development", **Tuberculosis**. 98: 30-41; May, 2016.

Tang, Y. and et al. "New technologies in computer-aided drug design: Toward target identification and new chemical entity discovery", **Drug Discovery Today: Technologies**. 3: 307-313, 2006.

Thuston, D. E. and Lobo, S. G. M. J. "The Chemotherapy of cancer", **Introduction to the Principles of Drug Design and Action**. Amsterdam. 331, 1998.

Tomillero, A. and Moral M.A. "Gateways to clinical trials", **Methods and Findings in Experimental and Clinical Pharmacology**. 31: 661-700; December, 2009.

Tonge, P.J., Kisker, C. and Slayden, R.A. "Development of modern InhA inhibitors to combat drug resistant strains of *Mycobacterium tuberculosis*", **Current Topics in Medicinal Chemistry**. 7(5): 489-498, 2007.

Tsui, V. and Case, D.A. "Theory and applications of the generalized born solvation model in macromolecular simulations", **Biopolymers**. 56(4): 275–291, 2001.

REFERENCES (CONTINUED)

Vannelli, T.A., Dykman, A. and Ortiz de Montellano, .PR. "The antituberculosis drug ethionamide is activated by a flavoprotein monooxygenase", **Journal of Biological Chemistry**. 277:12824–12829; April, 2002.

Verdonk, M.L. and et al. "Improved protein-ligand docking using GOLD", **Proteins: Structure, Function, and Bioinformatics**. 52(4): 609-623; September, 2003(a).

_____. "Modeling water molecules in protein-ligand docking using GOLD", **Journal of Medicinal Chemistry**. 48(20): 6504-6515; October, 2005(b).

Verma, J., Khedkar, V.M. and Coutinho, E.C. "3D-QSAR in drug design--a review", **Current Topics in Medicinal Chemistry**. 10(1): 95-115, 2010.

Vermeulen, K., Van Bockstaele, D. R. and Berneman, Z. N. "The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer", **Cell Proliferation**. 36(3): 131-149; June, 2003.

Vilchèze, C. and et al. "Altered NADH/NAD^+ ratio mediates coresistance to isoniazid and ethionamide in mycobacteria", **Antimicrobial Agents and Chemotherapy**. 49(2): 708-720; February, 2005.

Vuimo, T. A. and et al. **New Research on Biotechnology and Medicine**. Nova Science Publishers. 87, 2006.

Walburger, A. and et al. "Protein kinase G from pathogenic mycobacteria promotes survival within macrophages", **Science**. 304(5678): 1800-1804; June, 2004.

Wang, J. and et al. "Use of MM-PBSA in reproducing the binding free energies to HIV1 RT of TIBO derivatives and predicting the binding mode to HIV1 RT of efavirenz by docking and MM-PBSA", **Journal of Chemical Physics**. 123(22): 5221–5230; June, 2001(a).

Wang, J. and et al. "Development and testing of a general AMBER force field", **Journal of Computational Chemistry**. 25(9): 1157-1174; July, 2004(b).

_____. "Automatic atom type and bond type perception in molecular mechanical calculations", **Journal of Molecular Graphics and Modelling**. 25(2): 247–260; October, 2006(c).

REFERENCES (CONTINUED)

Wang, J. Cieplak, P. and Kollman, P.A. "How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?", **Journal of Computational Chemistry**. 21(12): 1049–1074; July, 2000.

Wang, J., Hao, T. and Xu, X. "Recent advances in free energy calculations with a combination of molecular mechanics and continuum models", **Current Computer-Aided Drug Design**. 2(3): 95-103; September, 2006.

Wang, R.X., Lai, L.H. and Wang, S.M. "Further development and validation of empirical scoring functions for structure-based binding affinity prediction", **Journal of Computer-Aided Molecular Design**. 16(1): 11–26; January, 2002.

Wargel, R.J., Hadur, C.A. and Neuhaus, F.C. "Mechanism of d-cycloserine action: transport mutants for d-alanine, d-cycloserine, and glycine", **Journal of Bacteriology**. 105(3): 1028-1035; March, 1971.

Warner, D. F. and Mizrahi, V. "The survival kit of *Mycobacterium tuberculosis*", **Nature Medicine**. 13(3): 282-284; March, 2007.

Weiner, S. J. and et al. "A new force field for molecular mechanical simulation of nucleic acids and proteins", **Journal of the American Chemical Society**. 106: 765-784; February, 1984.

Wendy, D. and et al. "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules", **Journal of the American Chemical Society**. 117: 5179-5197; 1995.

WHO. (1991) "WHO model prescribing information: drugs used in mycobacterial diseases", **WHO**. http://apps.who.int/medicinedocs/pdf/s5511e/s5511e.pdf. March 1, 2016(a).

_____. (2010) "Treatment of tuberculosis: guidelines", **Stop TB department**. http://apps.who.int/iris/bitstream/10665/44165/1/9789241547833_eng.pdf?ua=1. March 10, 2016(b).

REFERENCES (CONTINUED)

WHO. (2015) "Global tuberculosis report 2015" **WHO**. http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf. March 1, 2016(c).

Wichapong, K. and et al. "Postprocessing of protein-ligand docking poses using linear response MM-PB/SA: application to Wee1 kinase inhibitors", **Journal of Chemical Information and Modeling**. 50(9): 1574-1588; September, 2010(a).

_____. "Application of docking and QM/MM-GBSA rescoring to screen for novel Myt1 kinase inhibitors", **Journal of Chemical Information and Modeling**. 54(3): 881-893; March, 2014(b).

Woods, C.J. and et al. "A water-swap reaction coordinate for the calculation of absolute protein-ligand binding free energies", **The Journal of Chemical Physics**. 134(5): 054114; February, 2011(a).

_____. "Rapid decomposition and visualisation of protein-ligand binding free energies by residue and by water", **Faraday Discussions**. 169: 477-499; January, 2014(b).

Xiaoyun, L., Kun, H. and Qidong, Y. "Enoyl acyl carrier protein reductase inhibitors: a patent review (2006 -2010)", **Expert Opinion on Therapeutic Patents**. 21(7): 1007-1022; July, 2011.

Xiong, X. and et al. "Key targets and relevant inhibitors for the drug discovery of tuberculosis", **Current Drug Targets**. 14(6): 676-699, June, 2013.

Xue, W. and et al. "Understanding the effect of drug-resistant mutations of HIV-1 intasome on raltegravir action through molecular modeling study", **Molecular BioSystems**. 8: 2135-2144; August, 2012.

Ying, Z. and et al. "The catalase—peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*", **Nature**. 358: 591 – 593; August, 1992.

Zumla, A., Nahid, P. and Cole, S.T. "Advances in the development of new tuberculosis drugs and treatment regimens", **Nature Reviews Drug Discovery**. 12(5): 388-404; May, 2013.

REFERENCES (CONTINUED)

Zhang, Y. and et al. "The catalase-peroxidase gene and isoniazid resistance of Mycobacterium tuberculosis", **Nature**. 358(6387): 591-593; August 1992.

Zhao, H. and Caflisch, A. "Molecular dynamics in drug design", **European Journal of Medicinal Chemistry**. 91: 4-14; February, 2015.

Zhenkun, M. and et al. "Global tuberculosis drug development pipeline: the need and the reality", **Lancet**. 375: 2100–2109; June, 2010.

APPENDIX

PUBLICATIONS

1. **P. Kamsri**, A. Punkvang, S. Hannongbua, P. Saparpakorn and P. Pungpo, Elucidating structural basis of benzofuran pyrrolidine pyrazole derivatives for enhancing potency against both the InhA enzyme and intact *M. tuberculosis* cells: A combined MD simulations and 3D-QSAR study, RSC Advance, 2015, 5, 52926-52937. DOI: 10.1039/C5RA08103C.

2. **P. Kamsri**, A. Punkvang, P. Saparpakorn, S. Hannongbua, S. Irle and P. Pungpo, Elucidating the structural basis of diphenyl ether derivatives as highly potent enoyl-ACP reductase inhibitors through molecular dynamics simulations and 3D-QSAR study, Journal of Molecular Modeling, 2014, Online publications, DOI: 10.1007/s00894-014-2319-0.

3. **P. Kamsri**, N. Koohatammakun, A. Srisupan, P. Meewong, A. Punkvang, P. Saparpakorn, S. Hannongbua, P. Wolschann, S. Prueksaaroon, U. Leartsakulpanich and P. Pungpo, Rational design of InhA inhibitors in the class of diphenyl ether derivatives as highly potential anti-tubercular agents using molecular dynamics simulations, SAR and QSAR in Environmental Research, 2014, 25(6), 473-488, DOI: 10.1080/1062936X.2014.898690.

4. **P. Kamsri**, A. Punkvang, N. Pongprom, A. Srisupan, P. Saparpakorn, S. Hannongbua, P. Wolschann and P. Pungpo, Key structural features of azanaphthoquinone annelated pyrrole derivative as anticancer agents based on the rational drug design approaches, Molecular Informatics, 2013, 32, 541 – 554, DOI: 10.1002/minf.201200132.

RSC Advances

PAPER

Cite this: RSC Adv., 2015, 5, 52926

Elucidating structural basis of benzofuran pyrrolidine pyrazole derivatives for enhancing potency against both the InhA enzyme and intact *M. tuberculosis* cells: a combined MD simulations and 3D-QSAR study†

Pharit Kamsri,[a] Auradee Punkvang,[b] Supa Hannongbua,[c] Patchreenart Saparpakorn[c] and Pornpan Pungpo[a]

A 2-*trans* enoyl-acyl carrier protein (ACP) reductase or InhA of *M. tuberculosis* is a drug target of isoniazid (INH), the first-line drug for tuberculosis treatment. Many series of compounds have been developed as novel inhibitors of this enzyme. However, they lack good potency against purified InhA and activity against intact *M. tuberculosis* cells. Benzofuran pyrrolidin pyrazole derivatives are potent direct InhA inhibitors. These compounds show high potency for InhA inhibition with IC_{50} values at nanomolar levels. However, their activities against *M. tuberculosis* cells in terms of MIC_{90} were about one-thousand fold than IC_{50}. Accordingly, in this work, IC_{50} and MIC_{90} values of benzofuran pyrrolidin pyrazole derivatives were subjected to CoMFA and CoMSIA studies in order to investigate the structural basis required for good activity against both purified InhA and *M. tuberculosis* cells. Moreover, MD simulations were employed to evaluate key interactions for binding benzofuran pyrrolidin pyrazole derivatives in InhA. Based on MD results, the core structure of these compounds is the key portion for binding in the InhA pocket. Alternatively, R substituents showed weak interactions with the InhA pockets. Interpretation of IC_{50} and MIC_{90} CoMSIA contour maps revealed the structural requirements in terms of steric, electrostatic, hydrophobic and hydrogen donor and acceptor for IC_{50} and MIC_{90} values of InhA inhibitors. Finally, the integrated results obtained from MD simulations and graphic interpretation of CoMSIA models provided a structural concept for rational design of novel InhA inhibitors with better potency against both the InhA enzyme and intact *M. tuberculosis* cells.

Received 2nd May 2015
Accepted 10th June 2015

DOI: 10.1039/c5ra08103c

www.rsc.org/advances

1. Introduction

Tuberculosis (TB) is an infectious disease caused by *Mycobacterium tuberculosis* (*M. tuberculosis*) and remains one of the world's deadliest infectious diseases. The World Health Organization (WHO) reported that an estimated 9.0 million people developed new TB cases and 1.5 million people died from this disease in 2013. Moreover, the incidence of new TB cases and deaths in 2013 was higher than those reported previously.[1] The high mortality rate of TB is caused by multi drug-resistant tuberculosis (MDR-TB),[2-7] extensively drug-resistant tuberculosis (XDR-TB),[8,9] totally drug-resistant tuberculosis (TDR-TB)[10,11] and human immunodeficiency virus (HIV) co-infection.[1] A NADH-dependent 2-*trans* enoyl-acyl carrier protein (ACP) reductase or InhA has been identified as potential drug target for tuberculosis treatment.[12] This enzyme catalyzes the reduction of α,β-unsaturated fatty acids, the last step in fatty acids biosynthesis in *M. tuberculosis*.[13-14] InhA was reported as the drug target of isoniazid (INH), the first-line drug against tuberculosis.[20-23] Since INH is a prodrug, it requires the activation process of catalase-peroxidase (KatG) to generate the acyl radical active form. This radical is then covalently bound to nicotinamide adenine dinucleotide (NAD^+) to produce an active INH-NAD adduct acting as a potent InhA inhibitor.[18-22] The high potency of INH against InhA was lost by mutations in KatG. Therefore, many researchers aimed to discover novel inhibitors that can directly inhibit InhA without the KatG activation process. Inhibitors that can act like this are called direct InhA inhibitors. A class of *N*-((3*R*,5*S*)-1-(benzofuran-3-carbonyl)-5-carbamoylpyrrolidin-3-yl)-1*H*-pyrazole-5-carboxamide derivatives (benzofuran pyrrolidin pyrazole derivatives) have been identified as potent direct InhA inhibitors.[24] The majority of

[a]Department of Chemistry, Faculty of Science, Ubon Ratchathani University, Ubonratchathani, Thailand. E-mail: pornpan_ubu@yahoo.com; Fax: +664 535 8379; Tel: +664 535 3400 ext. 4124

[b]Faculty of Science, Nakhon Phanom University, Nakhon Phanom, Thailand

[c]Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok, Thailand

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c5ra08103c

benzofuran pyrrolidine pyrazole derivatives show high potency against purified InhA with inhibitory concentration of compound required to inhibit InhA at 50% (IC_{50}) values at the nanomolar level. However, these compounds show weak cellular activity against *M. tuberculosis*, with the minimum inhibitory concentration of compound that resulted in complete inhibition in growth of *M. tuberculosis* 90% (MIC_{90}) at the micromolar level. These results show poor correlation between IC_{50} and MIC_{90} values of benzofuran pyrrolidine pyrazole derivatives. In this work, IC_{50} and MIC_{90} values of benzofuran pyrrolidine pyrazole derivatives were used for comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) studies in order to investigate the structural basis of these compounds for good activity against both InhA and *M. tuberculosis*. Moreover, molecular dynamics (MD) simulations were employed to evaluate the key interactions for binding of benzofuran pyrrolidin pyrazole derivatives in InhA. Therefore, the integrated results obtained from MD simulations and graphic interpretation of quantitative structure activity relationship (QSAR) models should provide crucial structural concepts for improving the correlation between IC_{50} and MIC_{90} values of benzofuran pyrrolidin pyrazole derivatives.

2. Material and methods

2.1 Data sets and biological activities

Thirty-four benzofuran pyrrolidin pyrazole derivatives used for CoMFA and CoMSIA studies were identified from the published literature.[24] Chemical structures and experimental biological activities in terms of MIC_{90} and IC_{50} values of these compounds are shown in Table 1. MIC_{90} and IC_{50} values were nominally converted into $\log(1/MIC_{90})$ and $\log(1/IC_{50})$ values for CoMFA and CoMSIA studies. Based on the diversity of structures and wide range of activities, the data set of compounds was divided into 30 training set compounds for final model development and 4 test set compounds for model validation. All chemical structures of benzofuran pyrrolidin pyrazole derivatives were constructed using the standard tools available in the GaussView 3.07 program and were then fully optimized using the HF/6-31G method implemented in the Gaussian 09 program.[25] The harmonic vibrational frequencies of the optimized geometries have also been calculated. All elements in the calculated Hessian matrix are positive, which indicate that the structures are true minima on the potential energy surface.

2.2 Molecular docking calculations

In this study, molecular docking calculations using the GOLD Program[26–28] were employed with the aims of generating the initial structure for MD simulations and performing molecular alignment to set up CoMFA and CoMSIA models. The available X-ray structure of InhA in a complex with compound 1 (PDB code 4COD) was used as an initial structure for molecular docking calculations. All atoms of the protein were kept rigid, whereas ligand was flexible during the molecular docking calculations. The number of Genetic Algorithm (GA) runs was set to 15 runs with the default search algorithm parameters. The docking calculations were validated using the root-mean-square deviation (RMSD) value between the docked and observed X-ray conformations of compound 1 in its pocket. A RMSD value lower than 1 Å was acceptable. Then, molecular docking calculations with validated parameters were used to dock all remaining compounds into the InhA binding pocket. The binding mode that showed the lowest binding energy was selected for each compound and was used to set up CoMFA and CoMSIA models. It was then used as the initial structure for MD simulations of compounds 2, 22, 23 and 28.

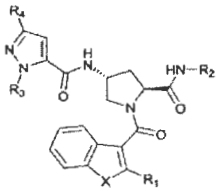
2.3 Molecular dynamics simulations

Compound 28, with the best IC_{50} value, was selected to investigate its binding mode in InhA. Moreover, the binding modes of compounds 2, 22 and 23 were modelled by MD simulations in order to investigate the effect of R_1 and R_2 substituents on the IC_{50} value. The AMBER12 program[29] was employed to perform molecular dynamics simulations. The complex structures of compounds 2, 22, 23 and 28 in InhA obtained from molecular docking calculations were used as the initial structure in MD simulations. The Amber ff03 force field was used for the physical description of InhA.[22] The general Amber force field (GAFF)[23,24] and restrained electrostatic potential (RESP) partial charges[25–28] of ligands and NAD^+ were generated by the *antechamber* module implemented in the AMBER12 package. To generate the system for MD simulations, the initial complex structure was solvated by TIP3P water[29] in a truncated octahedral box extending up to 10 Å from the solute species. Five Na^+ ions were added to neutralize the system charge. Initially, the energy of system was minimized using a steepest decent method followed by the conjugate gradient method. Then, the system was gradually warmed from 0 K to 300 K in 30 ps by restraining all atoms of the complex with a restraint weight of 2 kcal mol^{-1} $Å^{-2}$. This was followed by 70 ps of the position-restrained dynamics simulations with a restraining weight of 2 kcal mol^{-1} $Å^{-2}$ at 300 K under an isobaric condition. Finally, 10 ns MD simulations without any restraints were performed using the same conditions. Long-range electrostatic interactions were applied using the Particle Mesh Ewald method (PME)[30] during the simulations. The cut-off distance for the long-range van der Waals interaction was set to 8 Å. The SHAKE method[41] was applied to constrain the bond lengths of hydrogen atoms attached to heteroatoms. Coordinates and energy outputs during MD simulations were recorded at 2 ps intervals.

2.4 Binding free energy calculations

The Molecular Mechanics/Poisson-Boltzmann Surface Area (MM-PBSA) method[42–45] was employed for calculating the binding free energy of compounds 2, 22 and 23 in InhA. In this calculation, 250 snapshots of the complex, receptor and ligand were extracted every 8 ps from the last nanosecond of the MD trajectory, which represents the equilibrium state. The binding free energy (ΔG_{bind}) of compounds 2, 22 and 23 complexed with InhA were estimated from eqn (1), where ΔG_{vacuum} and ΔG_{solv}

Table 1 The chemical structures and activities against InhA and *M. tuberculosis* of thirty-four benzofuran pyrrolidin pyrazole derivatives



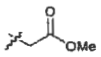
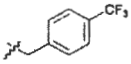
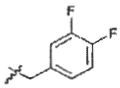
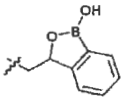
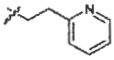
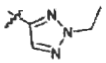
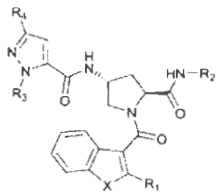
Cpd.	X	R_1	R_2	R_3	R_4	IC_{50} (μM)	MIC_{90} (μM)	$\log(1/IC_{50})$	$\log(1/MIC_{90})$
1	O	H	Et	Me	Et	0.034	8.00	7.47	5.10
2[a]	O	H		Me	Et	0.005	0.50	8.30	6.30
3	O	H	H	Me	Et	0.012	3.00	7.92	5.52
4	O	H	CH_2CF_3	Me	Et	0.046	4.00	7.34	5.40
5	O	H	$CH_2CH_2CH_3$	Me	Et	0.021	15.60	7.68	4.81
6	O	H	CH_2CH_2OMe	Me	Et	0.014	4.00	7.85	5.40
7[a]	O	H	CH_2CH_2COOEt	Me	Et	0.022	4.00	7.66	5.40
8	O	H		Me	Et	0.045	4.00	7.35	5.40
9	O	H		Me	Et	0.040	4.00	7.40	5.40
10	O	H		Me	Et	0.042	16.00	7.38	4.80
11[a]	O	H		Me	Et	0.009	2.00	8.05	5.70
12	O	H		Me	Et	0.035	3.00	7.46	5.52
13	O	H		Me	Et	0.112	1.00	6.95	6.00
14	O	H		Me	Et	0.025	1.00	7.60	6.00
15	O	H		Me	Et	0.018	16.00	7.74	4.80
16	O	H		Me	Et	0.009	8.00	8.05	5.10
17	O	H		Me	Et	0.003	4.00	8.52	5.40

Table 1 (Contd.)



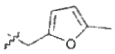
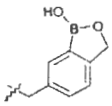
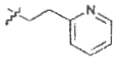
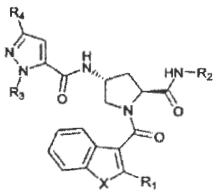
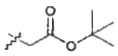
Cpd.	X	R_1	R_2	R_3	R_4	IC_{50} (μM)	MIC_{90} (μM)	$\log(1/IC_{50})$	$\log(1/MIC_{90})$
18	O	H		Me	Et	0.032	4.00	7.49	5.40
19	O	H		Me	Et	0.005	1.00	8.30	6.00
20	O	H		Me	Et	0.021	1.50	7.68	5.82
21	O	H		Me	Cyclopropyl	0.015	1.00	7.82	6.00
22	O	H		Et	Et	0.003	0.05	8.52	7.30
23	O	H	H	Et	Et	0.004	0.50	8.40	6.30
24	O	H		Et	Et	0.002	0.20	8.70	6.70
25	O	H		Et	Et	0.004	0.50	8.40	6.30
26	O	H		Et	Et	0.004	0.50	8.40	6.30
27	O	H	CH_2CH_2OH	Et	Et	0.003	1.00	8.52	6.00
28	O	H		Et	Et	0.002	0.70	8.70	6.15
29	O	Et		Me	Et	0.005	0.70	8.30	6.15
30	S	H		Me	Et	0.029	1.00	7.54	6.00
31[a]	O	Ph		Me	Et	0.003	1.50	8.52	5.82
32	O	H		Me	Et	0.018	2.00	7.74	5.70

Table 1 (Contd.)



Cpd.	X	R_1	R_2	R_3	R_4	IC_{50} (μM)	MIC_{90} (μM)	log(1/IC_{50})	log(1/MIC_{90})
33	O	H		Me	Et	0.007	2.00	8.15	5.70
34	O	H		Me	Et	0.008	2.00	8.10	5.70

[a] Test set.

were the binding free energy of the complex in vacuum and the solvation free energy, respectively. In the MM-PBSA approach, the solvation free energy was calculated by solving a linearized Poisson–Boltzman equation. ΔG_{vacuum} was obtained by calculating the interaction energy between InhA and compounds **2**, **22** and **23** (ΔE_{MM}) and taking the entropy change ($T\Delta S$) as shown in eqn (2). ΔE_{MM} is divided into three components, non-covalent van der Waals energy (ΔG_{vdW}), electrostatic energy (ΔG_{ele}) and internal energy (ΔG_{int}), as shown in eqn (3). ΔE_{MM} and ΔG_{solv} were calculated using the SANDER module and a PBSA program of the AMBER suite, respectively. The entropy contribution was estimated using normal mode analysis with the NMODE module.[46] The entropy contribution was estimated using 250 snapshots for the binding free energy calculation.

$$\Delta G_{bind} = \Delta G_{vacuum} + \Delta G_{solv} \quad (1)$$

$$\Delta G_{vacuum} = \Delta E_{MM} - T\Delta S \quad (2)$$

$$\Delta E_{MM} = \Delta G_{vdW} + \Delta G_{ele} + \Delta G_{int} \quad (3)$$

2.5 CoMFA and CoMSIA methods

IC_{50} and MIC_{90} values of compounds were used to set up CoMFA[47] and CoMSIA[48] models in order to evaluate the key structural features relating to the activity against both InhA and *M. tuberculosis*. The predicted binding modes of training set compounds obtained from molecular docking calculations were used for molecular alignment to set up CoMFA and CoMSIA models. SYBYL 8.0 molecular modelling software was used to run CoMFA and CoMSIA models. Partial least square (PLS) analysis was employed to derive a linear relationship between CoMFA and CoMSIA descriptor fields and activities. The PLS analysis, using the leave-one-out (LOO) cross-validation method, was performed to determine the optimal number of components. Sequentially, a final analysis with the optimal number of components was performed to construct CoMFA and CoMSIA models that were not cross-validated. The non-cross-validated correlation coefficient (r^2) and the leave-one-out cross-validated correlation coefficient (q^2) were used to evaluate the predictive ability of CoMFA and CoMSIA models. Selected CoMFA and CoMSIA models were employed to predict IC_{50} and MIC_{90} values of test set compounds that were not used to construct models. This was done to evaluate the external predictive ability of these models.

3. Results

3.1 Stability of the complex models

To reveal the structural stability of simulation system, the RMSD values for the position of all solute species were separately analyzed. The RMSD plots for the four simulation systems over 10 ns are shown in Fig. 1. Convergent RMSD plots indicated that the equilibrium state was reached for each system during this simulation period. As shown, the RMSDs for compounds **2**, **22**, **23** and **28** in InhA converged after approximately 2 ns.

3.2 Reliability of the calculation methods

MD simulations were employed to model the binding modes of compounds **2**, **22**, **23** and **28** in the InhA pocket. The experimental binding free energy (ΔG_{exp}) lying within the experimental error of the calculated values (ΔG_{bind}) considered as the correlation between the experimental binding free energy and

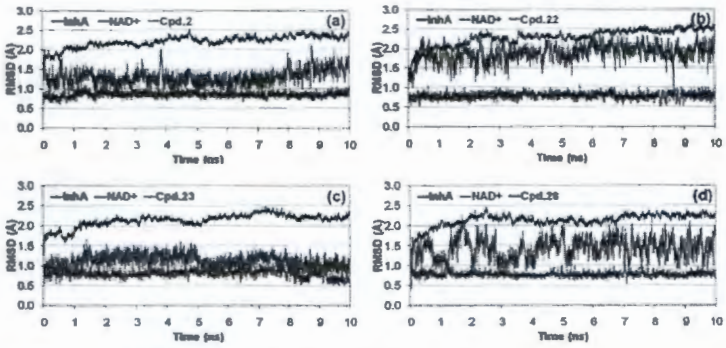


Fig. 1 RMSD plots of compounds 2 (a), 22 (b), 23 (c), and 28 (d) complexed with InhA.

Table 2 ΔG_{bind} and ΔG_{exp} of compounds 2, 22, 23 and 28 in InhA (kcal mol^{-1})

Cpd.	ΔH	$-T\Delta S$	ΔG_{bind}	ΔG_{exp}
2	−46.91 ± 5.08	−31.03 ± 6.06	−15.88 ± 5.14	−15.52
22	−49.69 ± 3.87	−33.15 ± 6.41	−16.54 ± 4.80	−15.82
23	−49.61 ± 3.71	−32.79 ± 5.57	−16.82 ± 4.79	−15.65
28	−49.26 ± 4.45	−32.52 ± 6.58	−16.74 ± 5.34	−16.07

the calculated values was used to indicate the reliability of the modelled binding modes of these compounds. ΔG_{bind} values of compounds 2, 22, 23 and 28 were close to their ΔG_{exp} values (Table 2). Therefore, we concluded that MD simulations reliably modelled binding modes of compounds 2, 22, 23 and 28 in the InhA pocket.

3.3 Binding mode of compound 28

The binding mode of compound 28 complexed with InhA obtained from MD simulations is shown in Fig. 2. Residues located near each substituent and the core structure are listed in Fig. 3. A hydrogen atom (the R_1 substituent) is near the carbonyl backbone of Met103. 2-pyridinyl methyl (the R_2 substituent) protrudes from the InhA pocket and interacts with the solvent (Fig. 2). The ethyl moiety (the R_3 substituent) is located near backbones of Gly96, Phe97 and pyrophosphate and ribose groups of NAD$^+$. The ethyl group (the R_4 substituent) was located in the hydrophobic side chains of Phe149, Tyr158, Met199 and nicotinamide of NAD$^+$. With regard to the core structure, the pyrazole ring in the core structure was sandwiched between two hydrophobic side chains of Met161 and Ala198. CO and NH of pyrazole amide formed hydrogen bonds with the backbones of Met98 and Ala198, respectively. The benzofuran core was buried in the hydrophobic side chains of Ile215, Ala157, Ile202 and Ala201, and was sandwiched between the hydrophobic side chains of Leu207 and Met103. The carbonyl of benzofuran core formed a hydrogen bond with the NH backbone of Ala201. NH of pyrrolidine amide formed a hydrogen bond with the CO backbone of Leu197.

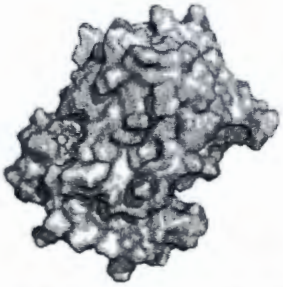


Fig. 2 Compound 28 (cyan) in its complex with whole InhA (grey) obtained from MD simulations.

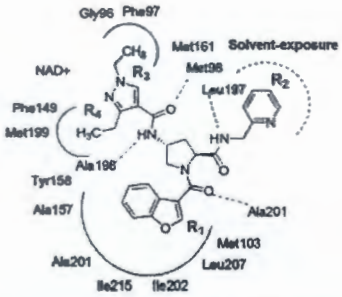


Fig. 3 List of residues surrounding within 4 Å from compound 28.

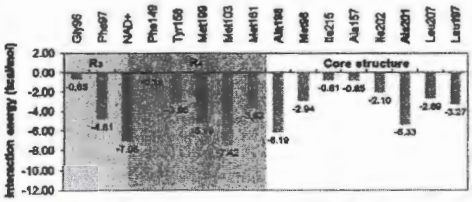


Fig. 4 Interaction energy profile of compound 28 and surrounding residues within 4 Å.

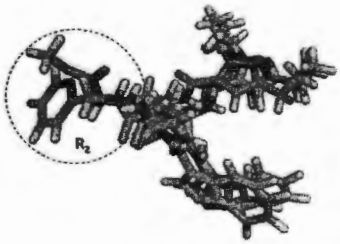


Fig. 5 The superimposition of binding modes of compounds 22 (pink), 23 (cyan) and 28 (green).

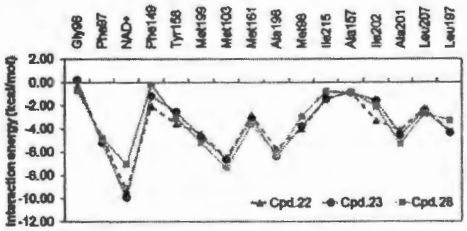


Fig. 6 Comparison of the interaction energy profiles of compounds 22 (green), 23 (blue) and 28 (yellow) with surrounding pocket within 4 Å.

3.4 Interaction energy

Free-energy decomposition calculations were used to investigate the interaction energies between compound 28 and each residue in the InhA pocket. Fig. 4 shows these interaction energies obtained from free-energy decomposition calculations. The lowest interaction energy (−7.42 kcal mol^{-1}) was observed for Met103, indicating that this residue had the largest contribution to binding of compound 28 in the InhA pocket. As previously mentioned, Met103 and Leu207 were sandwiched in the benzofuran core. Another remarkable interaction energy (−7.06 kcal mol^{-1}) was found for NAD$^+$. This was responsible for van der Waal and electrostatic interactions with the R_3 and R_4 substituents of compound 28 (Fig. 3). Ala198 showed an interaction energy (−6.16 kcal mol^{-1}), comparable with those of Met103 and NAD$^+$. This residue formed hydrogen bonds with the NH of pyrazole amide and sandwiched the pyrazole ring (Fig. 3). Met98, Leu197 and Ala201 formed other hydrogen bonds with the core structure with interaction energies of −2.94, −3.27 and −5.33 kcal mol^{-1}, respectively. Based on interaction energy profile of compound 28, the core structure formed more attractive interactive energies with surrounding residues than R substituents (Fig. 4). This result indicates that the core structure is the key fragment for binding of this compound in the InhA pocket.

3.5 The effect of the R_2 substituent on IC_{50} and MIC_{90} values

As compared with the positions of other R substituents, the R_2 position had the most varied substituents (Table 1). Compound 28 exposing the 2-pyridylmethyl at the R_2 position showed the best activity for InhA inhibition with an IC_{50} of 0.002 μM. When the R_2 substituent of this compound was replaced by CH_2-COOMe (compound 22), the IC_{50} value was slightly changed to 0.003 μM. In contrast, the MIC_{90} value against whole *M. tuberculosis* cell was greatly changed from 0.7 μM to 0.05 μM (Table 1). To reveal the effect of the R_2 substituent on the IC_{50} value, the binding modes of compounds 28 and 22 were compared (Fig. 5). The binding modes of these compounds in the InhA pocket were similar, and the R_2 substituents occupied in the same positions. Moreover, the interaction energy profiles of compounds 28 and 22 with residues in InhA pocket were similar (Fig. 6). As discussed above, the R_2 substituent of compound 28 protruded from the InhA pocket leading to weak interaction of this substituent with the pocket. Therefore, the IC_{50} value against InhA was not significantly changed when the R_2 substituent was varied. When the R_2 substituent was replaced by a hydrogen atom (compound 23), the binding mode and interaction energy profile of this compound were similar to those of compounds 22 and 28 (Fig. 5 and 6). With regard to IC_{50} values, compound 23 showed a comparable IC_{50} value with those of compounds 22 and 28. However, the MIC_{90} value of this compound (0.5 μM) was largely increased over that of compound 22 (0.05 μM). These results indicate that the R_2 substituent had a small effect on the IC_{50} value against InhA due to its weak interaction with the InhA pocket. Alternatively, this substituent is crucial to controlling the MIC_{90} against intact *M. tuberculosis* cells.

3.6 The effect of the R_3 substituent on IC_{50} and MIC_{90} values

The R_3 substituent of compounds in the data set was varied as ethyl (Et) or methyl (Me) groups (Table 1). Compounds 2 and 22 with structural differences at the R_3 substituent were selected to show the effect of the R_3 substituent on IC_{50} and MIC_{90} values. IC_{50} values of these compounds (0.005 and 0.003 μM, respectively) were not significant different, but their MIC_{90} values were tenfold different (0.5 and 0.05 μM, respectively). Fig. 7 shows the binding modes of compounds 2 and 22 in InhA obtained from MD simulations. The R_3 substituents of these compounds were located in the same position and surrounded by backbones of Gly96, Phe97 as well as pyrophosphate and ribose groups of NAD$^+$. The ethyl group (The R_3 substituent) of

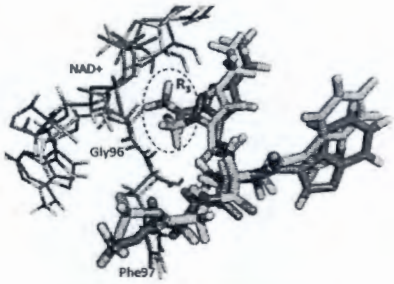


Fig. 7 The superimposition of binding modes of compounds 2 (yellow) and 22 (pink).

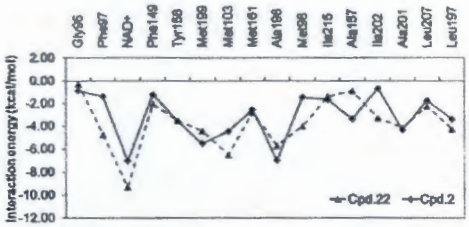


Fig. 8 Comparison of the interaction energy profiles of compounds 2 (gray) and 22 (green) with surrounding pocket within 4 Å.

compound 22 is close to Phe97 and pyrophosphate and ribose groups of NAD^+ more than the methyl group of compound 2. Therefore, interaction energies of compound 22 with Phe97 and NAD^+ had greater attraction than those of compound 2 (Fig. 8). Moreover, the presence of a methyl group at the R_3 position of compound 2 shifted the position of benzofuran core surrounded by Met103 and Ile202, and disrupted hydrogen bond interaction with Met98. Accordingly, interaction energies of compound 2 with Met98, Met103 and Ile202 showed less attraction than those of compound 22 (Fig. 8). These results indicate that compound 22 should have a better IC_{50} against InhA compared to compound 2. However, other than the interaction energies of Met98, Met103, Ile202, Phe97 and NAD^+, compounds 2 and 22 are comparable. The IC_{50} value for InhA inhibition by compound 22 was slightly better than that of compound 2. However, its MIC_{90} value was tenfold better than that of compound 2. The results indicated that the ethyl group at the R_3 position is more conducive to favorable IC_{50} and MIC_{90} values than the methyl group.

3.7 CoMFA and CoMSIA models

In this study, CoMFA and CoMSIA models were constructed from IC_{50} and MIC_{90} where prefixed with IC_{50} and MIC_{90}, respectively. IC_{50} and MIC_{90} CoMSIA models were constructed based on various combinations of molecular descriptor fields, in order to develop a highly predictive CoMSIA model (Tables 3 and 4). An IC_{50} CoMSIA model constructed from the combination of steric (*S*), electrostatic (*E*), hydrophobic (*H*) and hydrogen acceptor (*A*) fields[48] gave the highest q^2 (0.646), whereas an MIC_{90} CoMSIA model including steric, electrostatic, hydrophobic and hydrogen donor (*D*) fields[48] showed the highest q^2 (0.639). Therefore, these models were selected for graphical interpretation of IC_{50} and MIC_{90} CoMSIA contour maps. In order to assess the predictive abilities of IC_{50} and MIC_{90} CoMSIA models, IC_{50} and MIC_{90} values of the test set were predicted. Both IC_{50} and MIC_{90} CoMSIA models showed good ability to predict IC_{50} and MIC_{90} values of the test set data as shown in Fig. 9. In case of IC_{50} and MIC_{90} CoMFA models, they had poor predictive ability with q^2 values of 0.464 and 0.432, respectively. Accordingly, these CoMFA models were not used further in this work.

3.8 CoMSIA contour maps

To reveal the importance of molecular descriptor fields in both IC_{50} and MIC_{90} values of InhA inhibitors, CoMSIA contour maps were established. Compound 22 presented the best MIC value. Graphical interpretation of its IC_{50} and MIC_{90} CoMSIA contour maps was done. Interpretation of its IC_{50} and MIC_{90} CoMSIA contour maps revealed structural requirements in terms of steric, electrostatic, hydrophobic and hydrogen donor and acceptor fields for IC_{50} and MIC_{90} values of InhA inhibitors.

3.9 Steric requirements for IC_{50} and MIC_{90} values

Fig. 10 shows the CoMSIA steric contour maps obtained from selected IC_{50} and MIC_{90} CoMSIA models. These contours highlight the steric requirements for IC_{50} and MIC_{99} values of benzofuran pyrrolidine pyrazole derivatives. Both IC_{50} and MIC_{90} CoMSIA models show a green contour at the R_3 substituent. These results indicated that a bulky R_3 substituent is favourable for both IC_{50} and MIC_{90} values. Accordingly, an ethyl group is more preferred for the steric requirement of the R_3 substituent than a methyl group. This is consistent with the MD simulations since an ethyl group can form more interactions with InhA. At the R_2 position, IC_{50} and MIC_{90} CoMSIA models present a large yellow contour. However, IC_{50} CoMSIA model shows a favorable green steric contour at the terminal of the R_2 substituent (Fig. 10a). Based on MD simulations results, the R_2 substituent had weak interaction with the InhA pocket leading to less influence on the IC_{50} value. Therefore, the steric requirement of R_2 substituent should be based on the MIC_{90} CoMSIA steric contour that presented only a yellow contour near this substituent (Fig. 10b).

3.10 Electrostatic requirements for IC_{50} and MIC_{90} values

Electrostatic requirements for IC_{50} and MIC_{90} values of benzofuran pyrrolidine pyrazole derivatives are visualized in Fig. 11. Both IC_{50} and MIC_{90} CoMSIA contours show only an electrostatic requirement at the R_2 substituent. The IC_{50} CoMSIA shows a red contour at the ester moiety of R_2 substituent, whereas MIC_{90} CoMSIA presents a blue contour at this position. These results show different electrostatic requirements for IC_{50}

Table 3 Statistical results of IC_{50} CoMFA and CoMSIA models[a]

Models	Statistical parameters						Fraction
	q^2	r^2	s	SEE	N	F	
CoMFA							
S/E	0.464	0.996	0.392	0.035	6	909.618	60.3/39.7
CoMSIA							
S/E	0.084	0.977	0.512	0.081	6	162.845	32.1/67.9
S/H	0.465	0.950	0.383	0.118	5	90.431	29.1/70.9
S/D	0.624	0.923	0.321	0.145	5	57.579	54.3/45.7
S/A	0.146	0.970	0.495	0.093	6	123.724	39.7/60.3
S/E/H	0.260	0.981	0.460	0.074	6	194.704	16.6/44.5/38.9
S/E/D	0.592	0.980	0.342	0.076	6	185.576	21.0/53.7/25.3
S/E/A	0.281	0.975	0.454	0.085	6	149.701	22.5/42.8/34.7
S/E/H/D	0.646	0.990	0.318	0.055	6	363.962	13.1/35.8/28.5/22.6
S/E/H/A	0.336	0.983	0.436	0.070	6	222.520	12.3/31.5/29.4/26.8
S/E/H/D/A	0.610	0.991	0.334	0.050	6	437.341	10.0/25.4/22.6/20.7/21.4

[a] Bold values indicate the best CoMSIA model. *N* optimum number of components; *s* standard error of prediction; *SEE* standard error of estimate; *F* F-test value; *S* steric field; *E* electrostatic field; *H* hydrophobic field; *D* hydrogen donor field; *A* hydrogen acceptor field.

Table 4 Statistical results of MIC_{90} CoMFA and CoMSIA models[a]

Models	Statistical parameters						Fraction
	q^2	r^2	s	SEE	N	F	
CoMFA							
S/E	0.432	0.853	0.442	0.225	2	78.451	53.2/46.8
CoMSIA							
S/E	0.456	0.949	0.469	0.143	6	71.455	25.1/74.9
S/H	0.459	0.780	0.432	0.275	2	47.970	34.4/65.6
S/D	0.261	0.732	0.514	0.310	3	23.642	52.7/47.3
S/A	0.602	0.978	0.401	0.093	6	174.060	46.3/53.7
S/E/H	0.477	0.961	0.460	0.126	6	93.558	13.8/52.8/33.4
S/E/D	0.210	0.912	0.553	0.184	5	49.990	17.7/64.4/18.0
S/E/A	0.550	0.955	0.426	0.134	6	82.091	19.9/48.1/32.0
S/E/H/D	0.415	0.938	0.476	0.155	5	72.712	10.9/45.8/29.3/13.9
S/E/H/A	0.639	0.973	0.382	0.105	6	136.014	12.5/35.6/42.2/27.7
S/E/H/D/A	0.494	0.961	0.442	0.123	5	118.951	9.3/33.4/22.8/10.4/24.2

[a] Bold values indicate the best CoMSIA model. *N* optimum number of components; *s* standard error of prediction; *SEE* standard error of estimate; *F* F-test value; *S* steric field; *E* electrostatic field; *H* hydrophobic field; *D* hydrogen donor field; *A* hydrogen acceptor field.

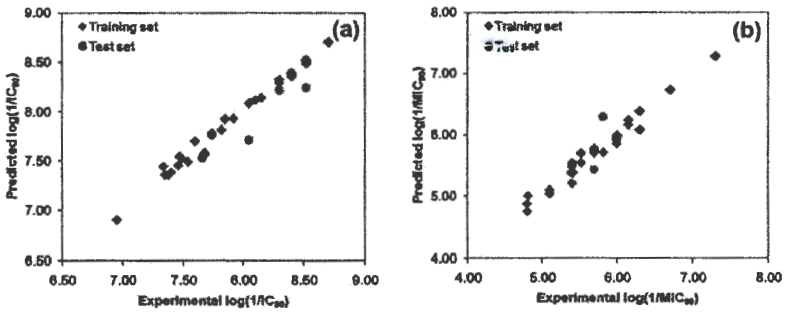


Fig. 9 The plot of experimental and predicted activities of the training and test data sets derived from IC_{50} (a) and MIC_{90} (b) CoMSIA models.

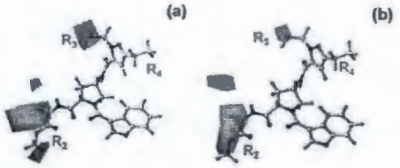


Fig. 10 Steric contour maps of IC_{50} (a) and MIC_{90} (b) CoMSIA models in combination with compound 22.

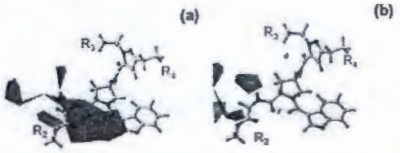


Fig. 11 Electrostatic contour maps of IC_{50} (a) and MIC_{90} (b) CoMSIA models in combination with compound 22.

and MIC_{90} values of benzofuran pyrrolidin pyrazole derivatives. However, the R_2 substituent has weak influence on the IC_{50} value. Therefore, the electrostatic requirement of R_2 substituent for MIC_{90} values should take more priority.

3.11 Hydrophobic requirements for IC_{50} and MIC_{90} values

Both IC_{50} and MIC_{90} CoMSIA contours show a purple contour at the R_3 substituent of compound 22 (Fig. 12). This shows that the hydrophobic requirements of the R_3 substituent for both IC_{50} and MIC values were similar. The R_3 substituent was either a methyl or ethyl group. As seen in Fig. 12, the terminal of ethyl group was buried in a purple R_3 contour. Therefore, the ethyl group was preferable for the hydrophobic requirement of the substituent. IC_{50} and MIC_{90} values of compound 2 with the methyl group at the R_3 substituent were weaker than those of compound 22 containing an ethyl group. At the R_4 substituent, both IC_{50} and MIC_{90} CoMSIA contours display a purple contour at this position (Fig. 12). Therefore, the presence of a hydrophobic substituent at this purple region should enhance both IC_{50} and MIC_{90} values. The grey contour located at the carbonyl moiety of the R_3 substituent in both IC_{50} and MIC_{90} CoMSIA contours indicated that this moiety is important for both IC_{50} and MIC_{90} values. Another important hydrophobic contour is located at the R_4 substituent. The MIC_{90} CoMSIA shows a purple

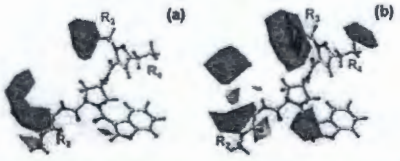


Fig. 12 Hydrophobic contour maps of IC_{50} (a) and MIC_{90} (b) CoMSIA models in combination with compound 22.

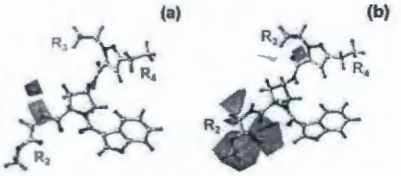


Fig. 13 Hydrogen donor contour of IC_{50} CoMSIA model (a) and hydrogen acceptor contour MIC_{90} CoMSIA model (b) in combination with compound 22.

region near the R_4 substituent (Fig. 12b), but this contour disappeared in the IC_{50} CoMSIA contour (Fig. 12a). Therefore, a hydrophobic moiety could be presented at purple region to enhance the MIC_{90} value without a negative contribution to the IC_{50} value.

3.12 Hydrogen donor and acceptor requirements for IC_{50} and MIC_{90} values

The hydrogen donor field was included in the selected IC_{50} CoMSIA model, but this molecular descriptor was instead changed to a hydrogen acceptor field in the selected MIC_{90} CoMSIA model (Fig. 13). The IC_{50} CoMSIA model did not show any hydrogen donor contour near any R substituents. However, this model showed a favourable hydrogen donor contour at the amide moiety of the core structure. The amide moiety appears to impact the IC_{50} value. Consistent with the MD simulations results, this moiety can form hydrogen bonds with Leu197. The MIC_{90} CoMSIA model shows a favourable hydrogen acceptor contour at the carbonyl moiety of R_2 substituent, indicating that this moiety is essential to a good MIC_{90} value.

3.13 The structural concept for good IC_{50} and MIC_{90} correlation

Based on the MD simulations results, the core structure of benzofuran pyrrolidine pyrazole derivatives is of key

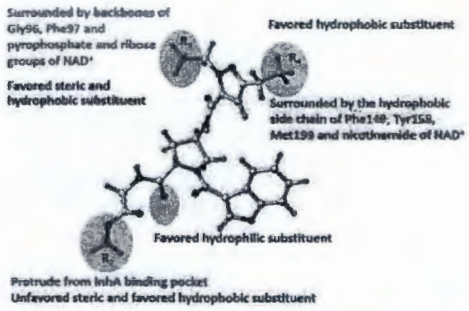


Fig. 14 The structural concept for good IC_{50} and MIC_{90} correlation summarized from MD simulations and CoMSIA results. Red and black letters indicate the results obtained from MD simulations and CoMSIA results, respectively.

importance for binding in the InhA pocket. Therefore, this fragment is crucial for favorable IC_{50} values. Among all R substituents, the R_2 substituent has the least interaction with the InhA pocket because it protrudes from the pocket. Modifications of the R_2 substituent did not significantly change IC_{50} values, but rather produced a tenfold increase in MIC_{90} values (compounds 22 and 23). Accordingly, the R_2 substituent is a key group that can be used to adjust the MIC_{90} value without negative contribution to the IC_{50} value. Based on the results obtained from our MD simulations and CoMSIA studies, the structural concept to correctly balance IC_{50} and MIC_{90} values of benzofuran pyrrolidin pyrazole derivatives is summarized in Fig. 14. New compounds designed based on this concept should show better IC_{50} and MIC_{90} values.

4. Conclusion

The combination of MD simulations and graphical interpretation of IC_{50} and MIC_{90} CoMSIA models highlight the structural concept to correctly balance IC_{50} and MIC_{90} values of benzofuran pyrrolidin pyrazole derivatives. The core structure of template compound is crucial to attaining favorable IC_{50} values, whereas the R_2 substituent is a key group to enhance MIC_{90} values without negative effects on IC_{50} values. Modifications of R substituents following the structural concept suggested here should allow design of novel InhA inhibitors with better potency against both the InhA enzyme and intact *M. tuberculosis* cells.

Acknowledgements

This research was supported by the Thailand Research Fund (DBG5680003, MRG5680169) the National Research Council of Thailand and Higher Education Research Promotion. The financial support from Royal Golden Jubilee Ph.D. Program (PHD/0004/2554) to P. Kamsri is gratefully acknowledged. Faculty of Science, Ubon Ratchathani University, Kasetsart University and NECTEC are gratefully acknowledged for supporting this research.

References

1 *Global Tuberculosis Report*, World Health Organization, Geneva, 2014.
2 M. C. Becerra, J. Bayona, J. Freeman, P. E. Farmer and J. Y. Kim, *International Journal of Tuberculosis and Lung Disease*, 2000, 4, 387–394.
3 C. Dye, M. A. Espinal, C. J. Watt, C. Mbiaga and B. G. Williams, *J. Infect. Dis.*, 2002, 185, 1197–1202.
4 M. A. Espinal, *Tuberculosis*, 2003, 83, 44–51.
5 A. Wright, M. Zignol, A. Van Deun, D. Falzon, S. R. Gerdes, K. Feldman, S. Hoffner, F. Drobniewski, L. Barrera, D. van Soolingen, F. Boulabhal, C. N. Paramasivan, K. M. Kam, S. Mitarai, P. Nunn and M. Raviglione, *Lancet*, 2009, 373, 1861–1873.
6 C. Y. Chiang, R. Centis and G. B. Migliori, *Respirology*, 2010, 15, 413–432.
7 *Multidrug and Extensively Drug-Resistant TB (M/XDR-TB):2010 Global Report on Surveillance and Response (WHO/HTM/TB/2010.3)*, World Health Organization, Geneva, 2010.
8 N. S. Shah, A. Wright, G. H. Bai, L. Barrera, F. Boulahbal, N. Martin-Casabona, F. Drobniewski, C. Gilpin, M. Havelkova, R. Lepe, R. Lumb, B. Metchock, F. Portaels, M. F. Rodrigues, S. RüschGerdes, A. Van Deun, V. Vincent, K. Laserson, C. Wells and J. P. Cegielski, *Emerging Infect. Dis.*, 2007, 13, 380–387.
9 M. Berry and O. M. Kon, *Eur. Respir. Rev.*, 2009, 18, 195–197.
10 A. A. Velayati, P. Farnia, M. R. Masjedi, T. A. Ibrahim, P. Tabarsi, R. Z. Haroun, H. O. Kuan, J. Ghanavi and M. Varahram, *Eur. Respir. J.*, 2009, 34, 1202–1203.
11 Z. F. Udwadia, R. A. Amale, K. K. Ajbani and C. Rodrigues, *Clin. Infect. Dis.*, 2012, 54, 579–581.
12 A. Quemard, J. C. Sacchettini, A. Dessen, C. Vilcheze, R. Bittman, W. R. Jacobs and J. S. Blanchard, *Biochemistry*, 1995, 34, 8235–8241.
13 C. Vilcheze, H. R. Morbidoni, T. R. Weisbrod, H. Iwamoto, M. Kuo, J. C. Sacchettini and W. R. Jacobs Jr, *J. Bacteriol.*, 2000, 182, 4059–4067.
14 D. A. Rozwarski, C. Vilchèze, M. Sugantino, R. Bittman and J. C. Sacchettini, *J. Biol. Chem.*, 1999, 274, 15582–15589.
15 D. A. Rozwarski, G. A. Grant, D. H. Barton, W. R. Jacobs Jr and J. C. Sacchettini, *Science*, 1998, 279, 98–102.
16 C. Vilcheze, F. Wang, M. Arai, M. H. Hazbon, R. Colangeli, L. Kremer, T. R. Weisbrod, D. Alland, J. C. Sacchettini and W. R. Jacobs Jr, *Nat. Med.*, 2006, 12, 1027–1029.
17 A. Dessen, A. Quemard, J. S. Blanchard, W. R. Jacobs Jr and J. C. Sacchettini, *Science*, 1995, 267, 1638–1641.
18 K. Johnsson and P. G. Schultz, *J. Am. Chem. Soc.*, 1994, 116, 7425–7426.
19 B. Lei, C. J. Wei and S. C. Tu, *J. Biol. Chem.*, 2000, 275, 2520–2526.
20 K. Johnsson, D. S. King and P. G. Schultz, *J. Am. Chem. Soc.*, 1995, 117, 5009–5010.
21 Y. Zhang, B. Heym, B. Allen, D. Young and S. Cole, *Nature*, 1992, 358, 591–593.
22 A. Quemard, A. Dessen, M. Sugantino, W. R. Jacobs Jr, J. C. Sacchetini and J. S. Blanchard, *J. Am. Chem. Soc.*, 1996, 118, 1561–1562.
23 A. Banerjee, E. Dubnau, A. Quemard, V. Balasubramanian, K. S. Um, T. Wilson, D. Collins, G. de Lisle and W. R. Jacobs Jr, *Science*, 1994, 263, 227–230.
24 L. Encinas, H. O'Keefe, M. Neu, M. J. Remuiñán, A. M. Patel, A. Guardia, C. P. Davie, N. Pérez-Macías, H. Yang, M. A. Convery, J. A. Messer, E. Pérez-Herrán, P. A. Centrella, D. Alvarez-Gómez, M. A. Clark, S. Huss, G. K. O'Donovan, F. Ortega-Muro, W. McDowell, P. Castañeda, C. C. Arico-Muendel, S. Pajk, J. Rullás, I. Angulo-Barturen, E. Alvarez-Ruiz, A. Mendoza-Losan, L. Ballell Pages, J. Castro-Pichel and G. Evindar, *J. Med. Chem.*, 2014, 57, 1276–1288.
25 M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. Hratchian, A. Izmaylov, J. Bloino, G. Zheng,

J. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. Montgomery, J. Peralta, F. Ogliaro, M. Bearpark, J. Heyd, E. Brothers, K. Kudin, V. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. Burant, S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. Millam, M. Klene, J. Knox, J. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. Stratmann, O. Yazyev, A. Austin, R. Cammi, C. Pomelli, J. Ochterski, R. Martin, K. Morokuma, V. Zakrzewski, G. Voth, P. Salvador, J. Dannenberg, S. Dapprich, A. Daniels, O. Farkas, J. Foresman, J. Ortiz, J. Cioslowski and D. Fox, *Gaussian 09*, Gaussian, Inc., Wallingford CT, 2009.

26 G. Jones, P. Willett and R. C. Glen, *J. Mol. Biol.*, 1995, 245, 43–53.

27 G. Jones, P. Willett, R. C. Glen, A. R. Leach and R. Taylor, *J. Mol. Biol.*, 1997, 267, 727–748.

28 J. W. Nissink, C. Murray, M. Hartshorn, M. L. Verdonk, J. C. Cole and R. Taylor, *Proteins*, 2002, 49, 457–471.

29 M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray and R. D. Taylor, *Proteins*, 2003, 52, 609–623.

30 M. L. Verdonk, G. Chessari, J. C. Cole, M. J. Hartshorn, C. W. Murray, J. W. Nissink, R. D. Taylor and R. Taylor, *J. Med. Chem.*, 2005, 48, 6504–6515.

31 D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Götz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. A. Kollman, *AMBER 12*, University of California, San Francisco.

32 Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. A. Kollman, *J. Comput. Chem.*, 2003, 24, 1999–2012.

33 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, 25, 1157–1174.

34 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graphics Modell.*, 2006, 25, 247–260.

35 P. Cieplak, W. D. Cornell, C. Bayly and P. A. Kollman, *J. Comput. Chem.*, 1995, 16, 1357–1377.

36 W. D. Cornell, P. Cieplak, C. I. Bayly and P. A. Kollmann, *J. Am. Chem. Soc.*, 1993, 115, 9620–9631.

37 C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, 97, 10269–10280.

38 J. Wang, P. Cieplak and P. A. Kollman, *J. Comput. Chem.*, 2000, 21, 1049–1074.

39 M. W. Mahoney and W. L. Jorgensen, *J. Chem. Phys.*, 2000, 112, 8910–8922.

40 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, 98, 10089–10092.

41 J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, 23, 327–341.

42 N. Homeyer and H. Gohlke, *Mol. Inf.*, 2012, 31, 114–122.

43 J. Wang, T. Hou and X. Xu, *Curr. Comput.-Aided Drug Des.*, 2006, 2, 95–103.

44 J. Wang, P. Morin, W. Wang and P. A. Kollman, *J. Am. Chem. Soc.*, 2001, 123, 5221–5230.

45 T. Hou, J. Wang, Y. Li and W. Wang, *J. Chem. Inf. Model.*, 2011, 51, 69–82.

46 M. Kaledin, A. Brown, A. L. Kaledin and J. M. Bowman, *J. Chem. Phys.*, 2004, 121, 5646–5653.

47 R. D. Cramer III, D. E. Patterson and J. D. Bunce, *J. Am. Chem. Soc.*, 1988, 110, 5959–5967.

48 G. Klebe, U. Abraham and T. Mietzner, *J. Med. Chem.*, 1994, 37, 4130–4146.

J Mol Model (2014) 20:2319
DOI 10.1007/s00894-014-2319-0

ORIGINAL PAPER

Elucidating the structural basis of diphenyl ether derivatives as highly potent enoyl-ACP reductase inhibitors through molecular dynamics simulations and 3D-QSAR study

Pharit Kamsri · Auradee Punkvang · Patchareenart Saparpakorn · Supa Hannongbua · Stephan Irle · Pornpan Pungpo

Received: 30 December 2013 / Accepted: 26 May 2014 / Published online: 17 June 2014

Abstract Diphenyl ether derivatives are good candidates for anti-tuberculosis agents that display a promising potency for inhibition of InhA, an essential enoyl-acyl carrier protein (ACP) reductase involved in fatty acid biosynthesis pathways in *Mycobacterium tuberculosis*. In this work, key structural features for the inhibition were identified by 3D-QSAR CoMSIA models, constructed based on available experimental binding properties of diphenyl ether inhibitors, and a set of four representative compounds was subjected to MD simulations of inhibitor-InhA complexes for the calculation of binding free energies. The results show that bulky groups are required for the R_1 substituent on the phenyl A ring of the inhibitors to favor a hydrophobic pocket formed by residues Phe149, Met155, Pro156, Ala157, Tyr158, Pro193, Met199, Val203, Leu207, Ile215, and Leu218. Small substituents with a hydrophilic property are required at the R_3 and R_4 positions of the inhibitor phenyl B rings to form hydrogen bonds with the backbones of Gly96 and Met98, respectively. For the R_2 substituent, small substituents with simultaneous hydrophilic or hydrophobic properties are required to favor the interaction with the pyrophosphate moiety of NAD^+ and the methyl side chain of Ala198, respectively. The reported data provide structural guidance for the design of new and potent diphenyl ether-based inhibitors with high inhibitory activities against *M. tuberculosis* InhA.

Keywords *M. tuberculosis* · InhA · 3D-QSAR · MD simulation · Diphenyl ether inhibitors

This paper belongs to Topical Collection 9th European Conference on Computational Chemistry (EuCo-CC9)

Electronic supplementary material The online version of this article (doi:10.1007/s00894-014-2319-0) contains supplementary material, which is available to authorized users.

P. Kamsri · P. Pungpo (✉)
Department of Chemistry, Faculty of Science, Ubon Ratchathani University, 85 Sthollmark Rd., Warinchamrap, Ubonratchathani 34190, Thailand
e-mail: pompan_ubu@yahoo.com

A. Punkvang
Faculty of Liberal Arts and Sciences, Division of Science, Nakhon Phanom University, Nakhon Phanom 48000, Thailand

P. Saparpakorn · S. Hannongbua
Department of Chemistry, Faculty of Science, Kasetsart University, Chatuchak, Bangkok 10900, Thailand

S. Irle
Institute of Transformative Bio-Molecules (WPI-ITbM) and Department of Chemistry, Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan

Introduction

Tuberculosis (TB), caused by pathogenic bacterial species *Mycobacterium tuberculosis*, remains a major global health problem and ranks as the second-leading cause of death from an infectious disease worldwide, after the human immunodeficiency virus (HIV). The latest estimates included in World Health Organization (WHO) report were 8.6 million new TB cases and 1.3 million TB deaths in 2012 [1]. The enoyl-acyl carrier protein (ACP) reductase (InhA) catalyzes the NADH-specific reduction of α,β-unsaturated fatty acids bound to the enoyl-ACP, the last step of fatty acids biosynthesis in *M. tuberculosis* [2, 3], and is an attractive target to design novel antitubercular drugs [4–10]. Moreover, InhA has been identified as the primary target of the most effective first-line anti-TB drug, isoniazid (INH) [11–19]. INH is a prodrug that is activated by catalase-peroxidase (KatG) enzymes to form an acyl radical that binds covalently to nicotinamide adenine dinucleotide (NAD^+) at the position 4, producing an active INH-NAD adduct [20–25] that functions as a highly potent inhibitor of InhA [26, 27]. However, such high potency of

INH for tuberculosis treatment can be diminished if mutations arise in KatG, as found in previous clinical studies [18, 28]. Therefore, new inhibitors targeting the InhA without the activation process from KatG are required. The first inhibitor, triclosan, inhibiting the InhA directly has been reported [29]. Based on the mechanism action of triclosan, triclosan and diphenyl ether derivatives have been developed by using structure-based drug design [30–35]. A diphenyl ether derivative, 5-octyl-2-phenoxy phenol, shows the highest potent InhA inhibitor with IC_{50} of 5 nM [33]. Recently, molecular modeling and computer-aided molecular design approaches have been performed to develop the InhA inhibitors [36–48]. To gain insight into the structural requirement of highly potent diphenyl ether derivatives as the InhA inhibitors, three-dimensional quantitative structure–activity relationships (3D-QSAR) based on comparative molecular similarity indices analysis (CoMSIA) was performed. Moreover, molecular dynamics (MD) simulations were also performed to gain deeper insight into a fundamental basis of structural behavior, inhibitor–InhA interactions and thermodynamic properties. The molecular information obtained from both CoMSIA and MD simulations should be valuable for the design of new and better InhA inhibitors as anti-tubercular agents.

Materials and methods

Data sets for QSAR study

The 52 diphenyl ether derivatives [30, 31, 33, 35] listed in Table 1 were used to build the CoMSIA model. The experimentally obtained IC_{50} values of each compound for InhA inhibition were converted to the corresponding log ($1/IC_{50}$) values and used as dependent variables for the QSAR model. The chemical structures of these compounds were constructed using standard tools available in the GaussView 3.07 program [49] and were then fully optimized using the ab initio quantum chemical method (HF/3-21G) implemented in the Gaussian 09 program [50]. The compounds were divided into a training set of 43 compounds, and a test set of nine compounds for model development and validation, respectively. The test set was randomly selected based on a structural diversity and wide range of activity in the data sets.

Molecular docking calculations

The X-ray crystal structure of diphenyl ether complexed with InhA (PDB code 2X23) [34] was used as a template for molecular docking calculations. Docking calculations for all 52 diphenyl ether derivatives were carried out by the Autodock 4.02 program using the Lamarckian genetic algorithm (LGA) [51]. Docking parameters were used as default values, except for the number of docking runs, which was set to 50. The parameters of the docking calculations were validated by successfully reproducing the X-ray conformation of the ligand in the PDB structure 2X23, as well as its orientation in the binding pocket. The RMSD value between original and docked coordinates was lower than 1 Å and therefore acceptable. For all 52 candidate compounds, the ligand pose with the lowest final docked energy was selected as the best binding mode of these potential InhA inhibitors.

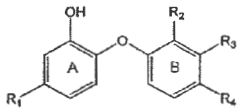
CoMSIA study

The binding mode of compound **21**, representing the best active compound for the InhA inhibition, was taken from the X-ray structure (PDB code 2B37) [33] and used as a template for molecular alignment. The pharmacophore alignment module with the GALAHAD fit implemented in SYBYL 8.0 program [52] was employed to align all compounds to the molecular template. SYBYL 8.0 molecular modeling software was then used to construct CoMSIA models. Five CoMSIA descriptors including steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor fields were calculated using an sp^3 carbon probe atom, with a formal charge of +1, which was placed at the intersections in a grid spacing of 2 Å. CoMSIA descriptors were set as independent variables and log ($1/IC_{50}$) values were used as dependent variables in the partial least square (PLS) analysis to derive a linear relationship between molecular descriptors and activities. The cross-validation was performed using the leave-one-out method with a 2.0-kcal/mol^{-1} column filter to minimize the influence of noisy columns. A final non-cross-validated analysis with the optimal number of components was sequentially performed and was then employed to analyze the results. The non-cross-validated correlation coefficient (r^2) and the leave-one-out (LOO) cross-validated correlation coefficient (q^2) were used to evaluate the predictive ability of the CoMSIA model. To estimate the predictive abilities of the best CoMSIA model, external validation using several statistical data was employed. According to Golbraikh and Tropsha [53], the best CoMSIA model is considerably acceptable if they satisfy all of the following criteria: $q^2>0.50$, $r^2>0.60$, and $0.85 \leq k \leq 1.15$.

MD simulations

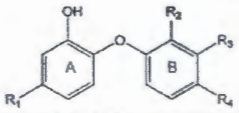
In a subsequent step, MD simulations were performed on compounds **17**, **18**, **19**, and **29**, which are representative compounds that cover a wide range from highly active (**17** and **29**) to less active compounds (**18**) among the candidate series in this study. Compound **19** was also included in the simulations to represent a moderate inhibitory activity.

Table 1 The chemical structures and activities for InhA inhibition of 52 diphenyl ether derivatives



CpdL	R_1	R_2	R_3	R_4	IC_{50}(nM)	Log(1/IC_{50})		
						Exp.	CoMSIA	Res.
1	Cl	Cl	H	Cl	1,100	5.96	6.03	−0.07
2	CH_3	Cl	H	Cl	800	6.10	6.12	−0.02
3	CH_2Cy	Cl	H	Cl	110	6.96	6.91	0.05
4[a]	CH_2CH_3	Cl	H	Cl	120	6.92	6.80	0.12
5	$(CH_2)_2CH_3$	Cl	H	Cl	91	7.04	6.84	0.20
6	$(CH_2)_3CH_3$	Cl	H	Cl	55	7.26	7.24	0.02
7	$(CH_2)_2CH(CH_3)_2$	Cl	H	Cl	63	7.20	7.27	−0.07
8	$CH_2CH(CH_3)CH_2CH_3$	Cl	H	Cl	130	6.89	6.78	0.11
9	CH_2(2-pyridyl)	Cl	H	Cl	29	7.54	7.39	0.15
10[a]	CH_2(3-pyridyl)	Cl	H	Cl	42	7.38	6.87	0.51
11	CH_2(4-pyridyl)	Cl	H	CN	75	7.12	6.98	0.14
12	o-CH_3-Ph	Cl	H	Cl	1,300	5.89	5.96	−0.07
13	m-CH_3-Ph	Cl	H	Cl	870	6.06	5.96	0.10
14	CH_2Ph	Cl	H	Cl	51	7.29	7.29	0.00
15	CH_2CH_2Ph	Cl	H	Cl	21	7.68	7.81	−0.13
16[a]	$(CH_2)_3Ph$	Cl	H	Cl	50	7.30	6.89	0.41
17	$(CH_2)_5CH3$	H	H	H	11	7.96	7.38	0.58
18	CH_2CH_3	H	H	H	2,000	5.70	6.33	−0.63
19	$(CH_2)_3CH_3$	H	H	H	80	7.10	7.47	−0.37
20	$(CH_2)_4CH_3$	H	H	H	17	7.77	7.78	−0.01
21	$(CH_2)_7CH_3$	H	H	H	5	8.30	8.23	0.08
22	$(CH_2)_{13}CH_3$	H	H	H	150	6.82	7.31	−0.49
23[a]	$(CH_2)_5CH_3$	NO_2	H	H	180	6.74	6.73	0.01
24	$(CH_2)_5CH_3$	H	NO_2	H	48	7.32	7.38	−0.05
25	$(CH_2)_5CH_3$	H	H	NO_2	90	7.05	6.99	0.06
26[a]	$(CH_2)_5CH_3$	NH_2	H	H	62	7.21	6.93	0.28
27	$(CH_2)_5CH_3$	H	NH_2	H	1,090	5.96	5.94	0.02
28	$(CH_2)_5CH_3$	H	H	NH_2	55	7.26	7.27	−0.01
29	$(CH_2)_5CH_3$	Br	H	H	10	8.00	7.93	0.07
30[a]	$(CH_2)_5CH_3$	CF_3	H	H	29.7	7.53	7.36	0.17
31	$(CH_2)_5CH_3$	F	H	H	12.1	7.92	7.92	0.00
32	$(CH_2)_5CH_3$	I	H	H	44.6	7.35	7.39	−0.04
33	$(CH_2)_5CH_3$	OH	H	H	48	7.32	7.29	0.03
34	$(CH_2)_5CH_3$	CN	H	H	235.6	6.63	6.72	−0.09
35	$(CH_2)_5CH_3$	Cl	H	H	49.5	7.31	7.51	−0.20
36[a]	$(CH_2)_5CH_3$	CH_3	H	H	50.7	7.29	7.14	0.15
37	$(CH_2)_5CH_3$	$NHCOCH_3$	H	H	1,550	5.81	5.88	−0.07
38	$(CH_2)_5CH_3$	H	H	$NHCONH_3$	1,300	5.89	5.87	0.02
39	$(CH_2)_5CH_3$	$NHCOCO_2H$	H	H	2,360	5.63	5.72	−0.09
40	$(CH_2)_5CH_3$	H	$NHCOCO_2H$	H	580	6.24	6.32	−0.08
41	$(CH_2)_5CH_3$	H	H		1,930	5.71	5.62	0.09
42[a]	$(CH_2)_5CH_3$	H	NHCO- isoxazole	H	1,220	5.91	6.08	−0.17

Table 1 (continued)



Cpd.	R_1	R_2	R_3	R_4	IC_{50}(nM)	Log(1/IC_{50})		
						Exp.	CoMSIA	Res.
43	$(CH_2)_5CH_3$	CH_2-*N*-CH_3-piperazine	H	H	1,315	5.88	5.76	0.12
44	$(CH_2)_5CH_3$	H	H	CH_2-*N*-CH_3-piperazine	306	6.51	6.53	−0.02
45	CH_2CH_2Ph	H	H	H	144.3	6.84	6.89	−0.05
46	CH_2CH_2Ph	CH_3	H	H	360.1	6.44	6.33	0.12
47	CH_2Ph	Cl	H	H	20.08	7.70	7.44	0.26
48	CH_2Ph	H	H	H	49.6	7.30	7.27	0.03
49	CH_2Ph	CH_3	H	H	56.4	7.25	7.59	−0.34
50	$CH_2CH_2CH_2OH$	CH_3	H	H	4,326	5.36	5.25	0.11
51	$OCH_2CH_2OCH_3$	H	H	H	253.1	6.60	6.55	0.05
52[a]	$O(CH_2)_6CH_3$	H	H	H	94.2	7.03	7.32	−0.29

[a] Test set

Complex InhA structures of these compounds as generated by the previous docking calculations were used as initial coordinates for MD simulations. The AMBER12 [54] software suite was used for all MD simulations to classically describe all relevant interactions within the system: InhA protein was described by the *ff03* force field [55] while NAD^+ and diphenyl ether inhibitors were described by the general AMBER force field (GAFF) [56, 57]. All missing hydrogen atoms of InhA were added using the LEaP module. To obtain the partial atomic charges of diphenyl ether derivatives and NAD^+, the geometry optimization and electrostatic potential calculation of each compound was first calculated at the HF/6-31G* level using the Gaussian 09 program [50]. Then, RESP partial charges [58–62] were assigned using the ANTECHAMBER module implemented in AMBER12. Each complex structure was solvated by TIP3P [63] waters in a truncated octahedral box extending up to 10 Å from each solute species. Five Na^+ cations were added to neutralize the charge in each system. Non-bonded cut-off was set to 10 Å. To relieve bad steric interactions that originated from addition of the water molecules and ions, the systems were first minimized with atomic positions of all solute species restraint (using a force constant of 500 kcal/mol^{-1} $Å^2$). Then, the whole system was fully minimized without restraining conditions. The solvated systems were gradually warmed up from 0 to 300 K in the first 20 ps followed by maintaining the temperature at 300 K during the last 10 ps. An integration time-step of 2 fs was used in a constant volume boundary. After minimization and heating, the position-restrained dynamics simulations were performed for 70 ps at 300 K under an isobaric condition to relax the positions of the solvent molecules. A weak force constant of 10 kcal/mol^{-1} $Å^2$ restraint on solute species was also applied for each simulation. Then, a 5-ns production MD simulation without restraints was performed on each system at a constant temperature of 300 K under isobaric condition. The Particle Mesh Ewald (PME) [64] was applied to treat the long-range electrostatic interactions with a periodic boundary condition during the MD simulations. The cut-off distance for the long-range van der Waals interaction was set to 8 Å. The SHAKE [65] method was applied to constrain the bond lengths of hydrogen atoms attached to heteroatoms. Coordinates and energy outputs during the MD simulation were collected every 2 ps. Finally, the root-mean-square deviations (RMSDs) of the InhA protein, NAD^+, and diphenyl ether ligand, respectively, were analyzed along the MD trajectory relative to the initial structures to determine the stability of the system. The binding free energies were calculated to evaluate the binding affinities of diphenyl ether derivatives in the InhA binding pocket.

Binding free energy calculation

The free energy of binding between InhA and diphenyl ether inhibitors were calculated using the Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA) [66–69] and Normal-mode [70] methods. For MM-PBSA calculation, 125 snapshots were generated

for each complex from the last 1 ns of MD trajectory with an interval of 8 ps. The binding free energies (ΔG_{bind}) were obtained using Eqs. (1–4).

$$\Delta G_{bind} = G_{com} - (G_{rec} + G_{ligand}) \quad (1)$$

$$\Delta G_{bind} = \Delta H - T\Delta S \quad (2)$$

$$\Delta H = \Delta G_{MM} + \Delta G_{solv} \quad (3)$$

$$\Delta G_{bind} = \Delta G_{MM} + \Delta G_{sol} - T\Delta S \quad (4)$$

where G_{com}, G_{rec}, and G_{ligand} are the free energies of the complex, InhA and the diphenyl ether inhibitors, respectively. In general, the binding free energy is composed of an enthalpic (ΔH) and an entropic contribution ($T\Delta S$). The enthalpic contribution (ΔH) contains the gas-phase molecular mechanics energy (ΔG_{MM}) and the solvation free energy (ΔG_{solv}) as shown in Eq. (3). The entropic contribution ($T\Delta S$) to the binding free energy was estimated using normal-mode analysis with AMBER Nmode module. Due to a highly computational cost in the entropy calculation, the residues around the ligand (less than 12 Å) were only considered as the receptor for normal-mode calculations [69, 71]. For this calculation, 50 snapshots were extracted from the last 1 ns of MD trajectory with an interval of 20 ps.

Results and discussion

MD simulation

System equilibration

Four MD simulations of compounds **17**, **18**, **19**, and **29** bound with InhA were performed for 5 ns to evaluate the structural stability of the complexes and their binding strength. The RMSDs for all atoms of three different solute species (InhA, NAD^+, and inhibitor) relative to the initial structure over the 5 ns of simulation times were analyzed and plotted in Fig. 1. The plateau characteristic of the RMSD plot over the simulation time is the criteria to indicate the equilibrium state of each solute species. Figure 1 shows that NAD^+ and compounds **17**, **18**, **19**, and **29** reach the equilibrium state at the early time. However, RMSDs of all compounds are more fluctuated, particularly compound **17**. InhA complexed with compounds **17**, **18**, **19**, and **29** reach the equilibrium state after 1.0 ns (Fig. 1a), 1.5 ns (Fig. 1b), 2.5 ns (Fig. 1c), and 1.0 ns (Fig. 1d), respectively. Moreover, to reveal the energy stability of each system, the receptor–ligand interaction energies of

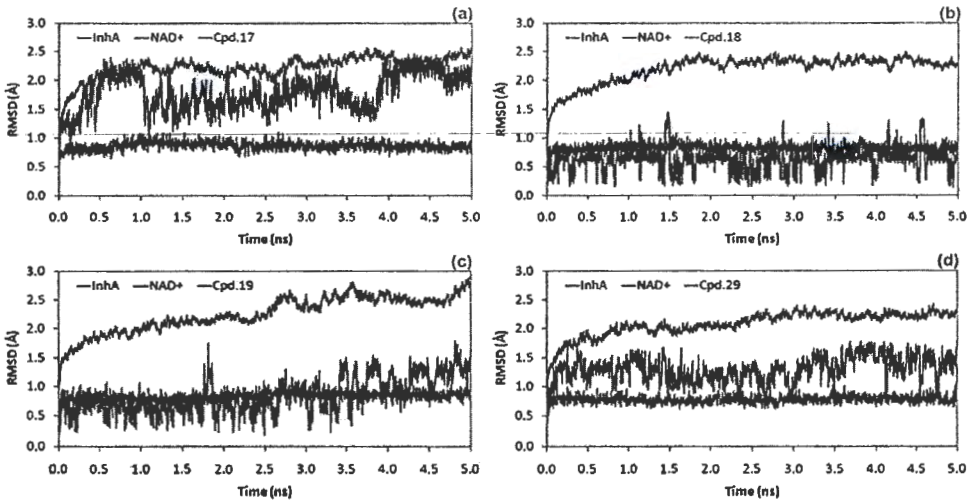


Fig. 1 RMSD plots of compounds **17** (a), **18** (b), **19** (c), and **29** (d) complexed with InhA

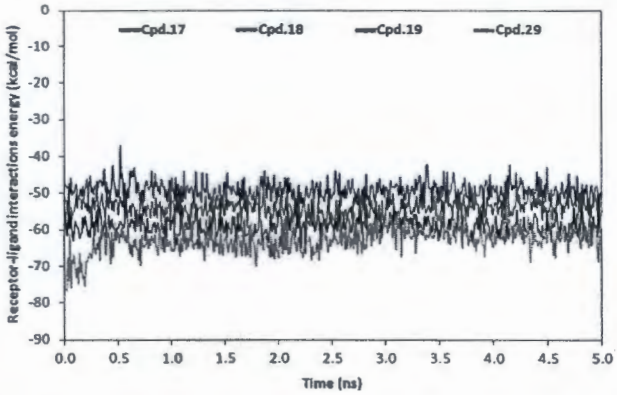


Fig. 2 Receptor–ligand interaction energies for the systems of compounds **17** (a), **18** (b), **19** (c), and **29** (d) over the 5 ns simulation

compounds **17**, **18**, **19**, and **29** over the 5-ns simulation time were calculated by MM-PBSA method. The receptor–ligand interaction energies of all compounds reach the equilibrium state at the beginning of the simulation time, except that of compound **29**, which reaches the equilibrium state after the 0.5 ns simulation time (Fig. 2). The average receptor–ligand interaction energies of compounds **17**, **18**, **19**, and **29** are −58.85±2.42, −49.27±2.55, −53.85±2.46, and −62.72±3.55 kcal/mol^{-1}, respectively. Based on the receptor–ligand interaction energy and RMSD plots, Compounds **17**, **18**, **19** and **29** complexed with InhA are sufficiently stable and the production simulations are reliable. Therefore, the subsequent free energy calculation and free energy decomposition analysis based on snapshots extracted from the stable state are reasonable.

Binding free energy calculations

The MM-PBSA method was employed to calculate the binding free energies of compounds **17**, **18**, **19**, and **29** in InhA and the results are shown in Table 2. The binding free energies (ΔG_{bind}) of compounds **17**, **18**, **19**, and **29** bound to the InhA pocket are calculated to be −15.02, −9.03, −13.90, and −15.40 kcal/mol^{-1}, respectively, which are in good agreement with those determined experimentally ($\Delta G_{exp.}$). Pearson correlation and Spearman rank correlation [72] were employed to determine the correlation between $\Delta G_{exp.}$ and ΔG_{bind}. The accepted values of correlation coefficient are in the range of −1 to 1. Based on these methods, the correlation between $\Delta G_{exp.}$ and ΔG_{bind} shows the correlation coefficient of Pearson correlation and Spearman rank correlation to be 0.98 and 1.00, respectively. Therefore, there is the correlation between $\Delta G_{exp.}$ and ΔG_{bind}.

Table 2 The binding free energies (kcal/mol^{-1}) calculated by the MM-PBSA method

Component	Diphenyl ether-InhA complexes			
	17	18	19	29
ΔG_{MM}	−58.77±2.59	−49.49±2.29	−52.90±2.57	−60.82±2.91
$\Delta G_{solv.}$	21.59±2.07	19.33±1.23	20.13±1.60	22.70±2.69
ΔH	−37.18±2.93	−30.16±2.24	−32.77±2.47	−38.06±3.21
$-T\Delta S$	22.16±0.85	21.13±1.17	18.87±1.06	22.66±0.57
$\Delta G_{bind.}$	−15.02±1.32	−9.03±.084	−13.90±1.31	−15.40±1.40
$\Delta G_{exp.}$[a]	−10.93	−7.83	−9.75	−10.99

[a] Derived from $\Delta G = RT \ln[IC_{50}]$, R represents the gas constant (1.988 cal/mol^{-1} K), T represents the temperature (300 K)

The binding modes of diphenyl ether derivatives in InhA

The binding modes of compounds **17**, **18**, **19**, and **29** bound with InhA pocket observed from the simulations are superimposed and illustrated in Fig. 3. In general, all compounds showed a similar binding mode and conformation: the OH group of the phenyl A ring lies in between the OH groups of Tyr158 and ribose fragment of NAD$^+$ to form the hydrogen bond interactions. The phenyl A ring forms the pi–pi interaction with pyridine amide ring of NAD$^+$. As the phenyl A ring bearing the R_1 substituent as the alkyl chain, it is placed in the hydrophobic pocket that is formed by Phe149, Met155, Pro156, Ala157, Tyr158, Pro193, Met199, Val203, Leu207, Ile215, and Leu218 (Fig. 3). Compounds **17** and **29** that hold the hexyl substituents at the R_1 position could form stronger hydrophobic interactions with Phe149, Met155, Pro156, Ala157, Tyr158, Pro193, Met199, Val203, Leu207, Ile215, and Leu218 when comparing these interactions with compounds **18** and **19** that have shorter alkyl substituents (containing ethyl and butyl, respectively), losing several hydrophobic interactions with Pro156, Ala157, Val203, Leu207,

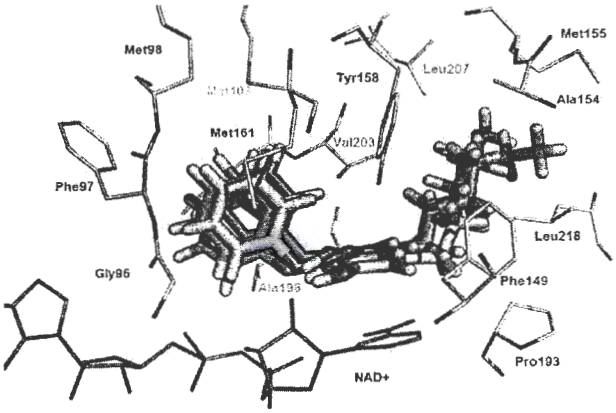


Fig. 3 The superimposition of compounds **17** (stick in *cyan color*), **18** (stick in *yellow color*), **19** (stick in *green color*), and **29** (stick in *pink color*) in the InhA pocket obtained from MD simulation

and Ile215. Therefore, the more hydrophobic interactions at the R_1 position of compounds **17** and **29** should account for better activities against InhA. The phenyl B ring containing the R_2, R_3, and R_4 substituents is surrounded by the pyrophosphate moiety of NAD^+, the hydrophilic backbones of Gly96, Met98, Phe97, and the hydrophobic side chains of Met103, Met161, Ile202, Val203, Ala198. The H and Br substituents at the R_2 position for compounds **17** and **29**, respectively, are closed to the methyl side chain of Ala198 and the pyrophosphate moiety of NAD^+ (Fig. 4). The Br substituent of compound **29** contributes greatly a hydrophobic interaction to the methyl side chain of Ala198 while the H substituent of compound **17** contributes a hydrophilic interaction to the ribose and pyrophosphate moieties of NAD^+. These results might explain why compounds **29** and **17** show the InhA inhibitory activities in the same level with IC_{50} of 10 and 11 nM, respectively. Accordingly, the R_2 substituent would also be hydrophobic or hydrophilic groups. For the R_3 position, the H substituents at this position for compounds **17**, **18**, **19**, and **29** form a hydrogen bond interaction with the carbonyl backbone of Gly96 and, besides the H substituent, other hydrogen bond donor substituents would also be possible. A similar H-bond interaction was also found for the R_4 substituent where all four compounds point to the NH and carbonyl backbone of Met98.

3D-QSAR study

CoMSIA model

The PLS results of CoMSIA models are summarized in Table 3. Ten CoMSIA models were constructed with various combinations of CoMSIA descriptors. Among all models,

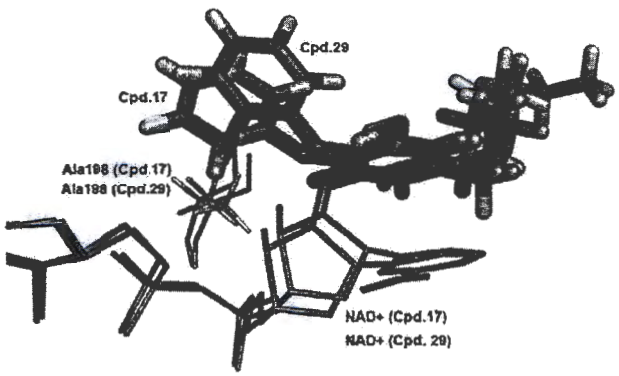


Fig. 4 The interactions of the R_2 substituents of compounds **17** and **29** with Ala198 and the pyrophosphate moiety of NAD^+

Table 3 Summary of statistical results of CoMSIA models

Models	Statistical data						Fraction
	q^2	r^2	s	SSE	F	N	
1.S/E	0.29	0.93	0.70	0.21	85.48	6	38.6/61.4
2.S/H	0.08	0.69	0.75	0.44	43.53	2	38.8/61.2
3.S/D	0.54	0.89	0.56	0.27	50.30	6	53.7/46.3
4.S/A	0.13	0.88	0.76	0.28	54.86	5	53.5/46.5
5.S/D/E	0.58	0.93	0.54	0.22	77.34	6	27.9/41.7/30.5
6.S/D/H	0.56	0.93	0.55	0.21	85.19	6	29.9/37.6/32.5
7.S/D/A	0.51	0.93	0.58	0.22	76.55	6	39.9/33.1/27.0
8. S/D/E/H	**0.60**	**0.95**	**0.52**	**0.19**	**104.17**	**6**	**19.0/32.5/23.8/24.8**
9. S/D/E/A	0.50	0.93	0.58	0.22	78.66	6	23.9/34.3/25.9/15.9
10.S/D/E/H/A	0.55	0.95	0.55	0.19	103.76	6	17.2/26.6/21.4/22.1/12.8

Bold values indicate the best CoMSIA model

N optimum number of components; *s* standard error of prediction; *SEE* standard error of estimate; *F* F-test value; *S* steric field; *E* electrostatic field; *H* hydrophobic field; *D* hydrogen donor field; *A* hydrogen acceptor field

model 8 composing the steric, hydrogen bond donor, electrostatic and hydrophobic fields is the best CoMSIA model, giving the best statistical parameters with a q^2 value of 0.60 and an r^2 value of 0.95. The predicted activities of 43 compounds in the training set and nine compounds in test set derived from the best CoMSIA model are summarized in Table 1. There is a good correlation between actual and predicted activities of the training set based on the best CoMSIA model, as depicted in Fig. 4. In order to assess the external predictive ability of this model, the InhA inhibitory activities of the test set were predicted. The predicted values of nine test-set compounds are within one logarithmic unit difference from the experimental values (Fig. 5). Therefore, the best CoMSIA model is reliable with highly predictive ability and could be utilized to predict the InhA activities for newly designed diphenyl ether inhibitors.

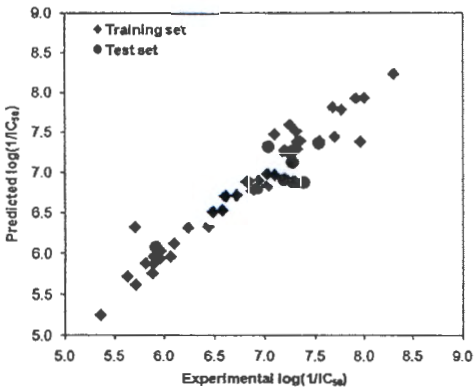


Fig. 5 Plots between the experimental and predicted activities of the training and test sets derived from the CoMSIA model

The predictive abilities of the best CoMSIA model were determined from the test set including nine compounds. For the best CoMSIA model, internal validation of leave-one-out cross-validated q^2 and predicted r^2 ($r^2_{pred.}$ or r^2) were found to be 0.64 and 0.70, respectively. The calculated square correlation coefficient values between the experimental and predicted values of the test-set compounds with intercept set at zero (r^2_0) and without intercept (r^2) were 0.56 and 0.73, respectively. The slope of regression line through the origin (k) of the best CoMSIA model was 1.02, which is close to 1. Based on the statistical results, the best CoMSIA model could be considered reliable.

CoMSIA contour maps

To reveal the importance of molecular descriptor fields on InhA inhibitory activities of diphenyl ether derivatives, CoMSIA contour maps were established. Figures 6 and 7 present the CoMSIA contour maps that reveal the influence of steric, electrostatic, hydrophobic, and hydrogen donor fields to the activity of diphenyl ether derivatives. Green and yellow contours indicate areas where favorable and unfavorable steric bulks are predicted to enhance the activities of diphenyl ether derivatives. Blue and red contours indicate regions where electropositive and electronegative groups lead to an increase of the InhA inhibitory activity, respectively. Magenta and white contours represent areas where the hydrophobic group and the hydrophilic group are predicted to favor the biological activities. The cyan and orange contours indicate regions that favor the hydrogen donor group and unfavor hydrogen donor group, respectively. The interpretation of

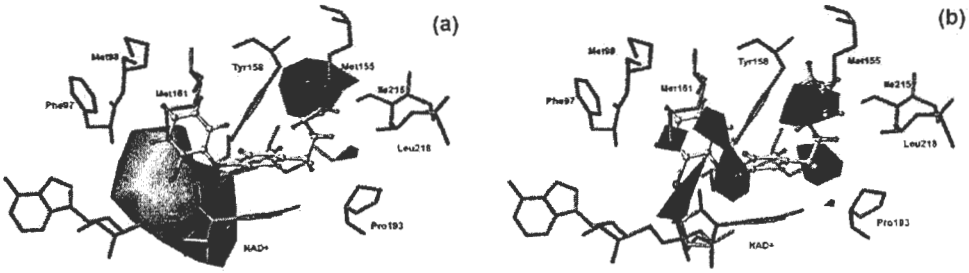


Fig. 6 CoMSIA steric (a) and electrostatic (b) contours in combination with compound **29** (ball and stick in atom type colors) in InhA binding pocket (stick in *greenblue*)

CoMSIA contour maps reveals the structural requirement of each substituent position in the scaffold of diphenyl ether derivatives helpful for rational design of novel and potent InhA inhibitors.

Structural requirement for the R_1 positions on the phenyl A ring

The appearance of cyan contours near the OH group of the phenyl A ring emphasizes the important role of this moiety to the InhA inhibitory activity of diphenyl ether derivatives (Fig. 7b). The C4 and C6 atoms of hexyl side chain of compound **29** are covered by green and red contours (Fig. 6a and b). Therefore, the R_1 substituent containing the bulky size and high electron density would be favorable for this region. In case of the R_1 substituent as the alkyl chain, the alkyl chain with the carbon atoms higher than two atoms should be preferable for the InhA inhibitory activity. As exemplified, compound **21**, containing an octyl group at the R_1 position, possesses the most active compound in this series, whereas compound **18** containing ethyl substituent exhibits much lower inhibitory activity than those of compounds **19**, **20**, **17**, and **21** bearing butyl, pentyl, hexyl, and octyl substituents at the R_1 position, respectively. Corresponding to the MD results, the longer alkyl chain at R_1 substituent could form hydrophobic interactions more than the shorter alkyl chain.

Structural requirement for the R_2, R_3, and R_4 positions on the phenyl B ring

The unfavorable hydrophobic white contour and the unfavorable steric yellow contour present near the R_2, R_3, and R_4 substituents (Figs. 6a and 7a). These results indicate that the small hydrophilic substituents at the R_2, R_3, and R_4 positions are required for the InhA inhibitory activity of diphenyl ether derivatives. Therefore, compounds **37**–**44** containing the bulky hydrophilic substituents at the R_2, R_3, and R_4 positions show poor activities for InhA inhibition with IC_{50} more than 360 nM. These suggestions are in agreement with the binding modes of compounds **17**, **18**, **19**, and **29** observed from the MD simulations that the R_2, R_3, and R_4 substituents are located near the pyrophosphate moiety of NAD^+, the hydrophilic backbones of Gly96 and Met98, respectively. Accordingly, the small substituent with hydrophilic property

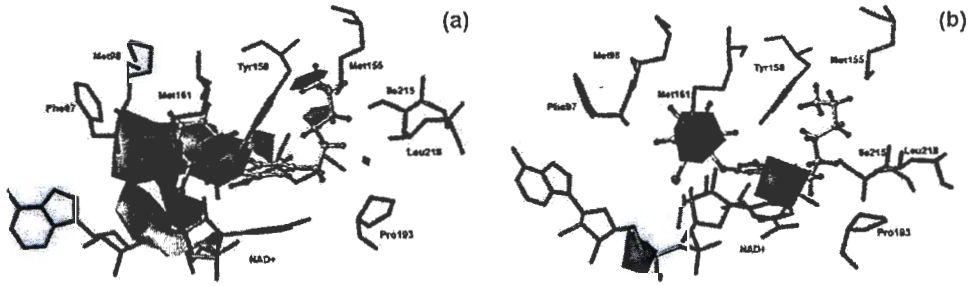


Fig. 7 CoMSIA hydrophobic (a) and hydrogen bond donor (b) contours in combination with compound **29** (ball and stick in atom type colors) in InhA binding pocket (stick in *greenblue*)

at the R_2, R_3 and R_4 substituents should be optimal for the InhA binding pocket. Moreover, the magenta and blue contours close to the R_2 substituent suggest additional structural requirement at this position, which should contain the hydrophobic property and less electron density. This suggestion is consistent with the MD results, which indicate that the R_2 position can be substituted with hydrophobic or hydrophilic groups so that the phenyl B ring could be favorable in binding with the methyl side chain of Ala198, and the pyrophosphate moiety of NAD^+, respectively. Apart from the hydrophobic properties, the R_2 substituent with the less electron density should be optimal for the pyrophosphate moiety of NAD^+ presenting the negative charge.

Conclusions

MD simulations were successfully applied to reliably predict binding modes, inhibitor–enzyme interactions, and binding free energies of diphenyl ether derivatives in the InhA binding pocket. The graphic interpretation of the obtained CoMSIA model reveals the key structural elements of diphenyl ether derivatives necessary for good InhA inhibitory activities. The structural requirements derived from the CoMSIA model correspond well with the binding interactions of diphenyl ether derivatives in the InhA pocket found in the MD simulations. The presented integrated results should be useful as guiding principles for the design of novel InhA inhibitors based on suitable modifications of the diphenyl ether scaffold.

Acknowledgments This research has been supported by the Thailand Research Fund (DBG5380006, MRG5680169, and RTA5380010), the National Research Council of Thailand and Higher Education Research Promotion. The financial support from Royal Golden Jubilee Ph.D. Program (PHD/004/2554) to P. Kamsri is gratefully acknowledged. Faculty of Science, Ubon Ratchathani University, University of Vienna, ASEA-Uninet (Austrian - South-East Asian University Partnership Network), NECTEC (National Electronics and Computer Technology Center) and BIOTEC (National center for genetic engineering and biotechnology) are gratefully acknowledged for supporting this research.

References

1. World Health Organization (2013) Global tuberculosis report. Available at http://apps.who.int/iris/bitstream/10665/91355/1/9789241564656_eng.pdf
2. Quemard A, Sacchettini JC, Dessen A, Vilcheze C, Bittman R, Jacobs WR Jr, Blanchard JS (1995) Enzymatic characterization of the target for isoniazid in *Mycobacterium tuberculosis*. Biochemistry 34(36):8235–8241
3. Rozwarski DA, Vilchèze C, Sugantino M, Bittman R, Sacchettini JC (1999) Crystal structure of the Mycobacterium tuberculosis enoyl-ACP reductase, InhA, in complex with NAD^+ and a C16 fatty acyl substrate. J Biol Chem 274(22):15582–15589
4. Agüero F, Al-Lazikani B, Aslett M, Berriman M, Buckner FS, Campbell RK, Carmona S, Carruthers IM, Chan AW, Chen F, Crowther GJ, Doyle MA, Hertz-Fowler C, Hopkins AL, McAllister G, Nwaka S, Overington JP, Pain A, Paolini GV, Pieper U, Ralph SA, Riechers A, Roos DS, Sali A, Shanmugam D, Suzuki T, Van Voorhis WC, Verlinde CL (2008) Genomic-scale prioritization of drug targets: the TDR targets database. Nat Rev Drug Discov 7(11):900–907
5. Campbell JW, Cronan JE Jr (2001) Bacterial fatty acid biosynthesis: targets for antibacterial drug discovery. Annu Rev Microbiol 55:305–332
6. Heath RJ, Rock CO (2004) Fatty acid biosynthesis as a target for novel antibacterials. Curr Opin Invest Drugs 5(2):146–153
7. White SW, Zheng J, Zhang YM, Rock CO (2005) The structural biology of type II fatty acid biosynthesis. Annu Rev Biochem 74: 791–831
8. Zhang YM, Lu YJ, Rock CO (2004) The reductase steps of the type II fatty acid synthase as antimicrobial targets. Lipids 39(11):1055–1060
9. Wen L, Chmielowski JN, Bohn KC, Huang JK, Timsina YN, Kodali P, Pathak AK (2009) Functional expression of *Francisella tularensis* FabH and FabI, potential antibacterial targets. Protein Exp Purif 65(1):83–91
10. Wright HT, Reynolds KA (2007) Antibacterial targets in fatty acid biosynthesis. Curr Opin Microbiol 10(5):447–453
11. Rozwarski DA, Grant GA, Barton DH, Jacobs WR Jr, Sacchettini JC (1998) Modification of the NADH of the isoniazid target (InhA) from *Mycobacterium tuberculosis*. Science 279(5347):98–102
12. Vilcheze C, Wang F, Arai M, Hazbon MH, Colangeli R, Kremer L, Weisbrod TR, Alland D, Sacchettini JC, Jacobs WR Jr (2006) Transfer of a point mutation in *Mycobacterium tuberculosis* InhA resolves the target of isoniazid. Nat Med 12(9):1027–1029
13. Dessen A, Quemard A, Blanchard JS, Jacobs WR Jr, Sacchettini JC (1995) Crystal structure and function of the isoniazid target of *Mycobacterium tuberculosis*. Science 267(5204):1638–1641
14. Johnsson K, Schultz PG (1994) Mechanistic studies of the oxidation of isoniazid by the catalase peroxidase from *Mycobacterium tuberculosis*. J Am Chem Soc 116(16):7425–7426
15. Lei B, Wei CJ, Tu SC (2000) Action mechanism of antitubercular isoniazid. Activation by *Mycobacterium tuberculosis* KatG, isolation, and characterization of InhA inhibitor. J Biol Chem 275(4):2520–2526
16. Johnsson K, King DS, Schultz PG (1995) Studies on the mechanism of action of isoniazid and ethionamide in the chemotherapy of tuberculosis. J Am Chem Soc 117(17):5009–5010
17. Zhang Y, Heym B, Allen B, Young D, Cole S (1992) The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*. Nature 358(6387):591–593
18. Banerjee A, Dubnau E, Quemard A, Balasubramanian V, Um KS, Wilson T, Collins D, de Lisle G, Jacobs WR Jr (1994) InhA, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*. Science 263(5144):227–230
19. Quemard A, Dessen A, Sugantino M, Jacobs WR Jr, Sacchetini JC, Blanchard JS (1996) Binding of catalase-peroxidase-activated isoniazid to wild-type and mutant *Mycobacterium tuberculosis* enoyl-ACP reductases. J Am Chem Soc 118(6):1561–1562
20. Saint-Joanis B, Souchon H, Wilming M, Johnsson K, Alzari PM, Cole ST (1999) Use of site-directed mutagenesis to probe the structure, function and isoniazid activation of the catalase/peroxidase, KatG, from *Mycobacterium tuberculosis*. Biochem J 338(Pt 3):753–760
21. Zhao X, Yu H, Yu S, Wang F, Sacchettini JC, Magliozzo RS (2006) Hydrogen peroxide-mediated isoniazid activation catalyzed by *Mycobacterium tuberculosis* catalase-peroxidase (KatG) and its S315T mutant. Biochemistry 45(13):4131–4140
22. Metcalfe C, Macdonald IK, Murphy EJ, Brown KA, Raven EL, Moody PC (2008) The tuberculosis prodrug isoniazid bound to activating peroxidases. J Biol Chem 283(10):6193–6200
23. Sinha BK (1983) Enzymatic activation of hydrazine derivatives. J Biol Chem 258(2):796–801

24. Nguyen M, Claparols C, Bernadou J, Meunier B (2001) A fast and efficient metal-mediated oxidation of isoniazid and identification of isoniazid-NAD(H) adducts. ChemBioChem 2(12):877–883
25. Heym B, Zhang Y, Poulet S, Young D, Cole ST (1993) Characterization of the katG gene encoding a catalase-peroxidase required for the isoniazid susceptibility of *Mycobacterium tuberculosis*. J Bacteriol 175(13):4255–4259
26. Timmins GS, Deretic V (2006) Mechanisms of action of isoniazid. Mol Microbiol 62(5):1220–1227
27. Johnsson K, Froland WA, Schultz PG (1997) Overexpression, purification, and characterization of the catalase-peroxidase KatG from *Mycobacterium tuberculosis*. J Biol Chem 272(5):2834–2840
28. De La Iglesia AI, Morbidoni HR (2006) Mechanisms of action of and resistance to rifampicin and isoniazid in *Mycobacterium tuberculosis*: new information on old friends. Rev Argent Microbiol 38(2):97–109
29. Parikh SL, Xiao G, Tonge PJ (2000) Inhibition of InhA, the enoyl reductase from *Mycobacterium tuberculosis*, by triclosan and isoniazid. Biochemistry 39(26):7645–7650
30. Freundlich JS, Wang F, Vilchèze C, Gulten G, Langley R, Schiehser GA, Jacobus DP, Jacobs WR Jr, Sacchettini JC (2009) Triclosan derivatives: towards potent inhibitors of drug-sensitive and drug-resistant *Mycobacterium tuberculosis*. ChemMedChem 4(2):241–248
31. am Ende CW, Knudson SE, Liu N, Childs J, Sullivan TJ, Boyne M, Xu H, Gegina Y, Knudson DL, Johnson F, Peloquin CA, Slayden RA, Tonge PJ, Ende CW, Knudson SE, Liu N, Childs J, Sullivan TJ, Boyne M, Xu H, Gegina Y, Knudson DL, Johnson F, Peloquin CA, Slayden RA, Tonge PJ (2008) Synthesis and in vitro antimycobacterial activity of B-ring modified diaryl ether InhA inhibitors. Bioorg Med Chem Lett 18(10):3029–3033
32. Boyne ME, Sullivan TJ, am Ende CW, Lu H, Gruppo V, Heaslip D, Amin AG, Chatterjee D, Lenaerts A, Tonge PJ, Slayden RA (2007) Targeting fatty acid biosynthesis for the development of novel chemotherapeutics against *Mycobacterium tuberculosis*: evaluation of A-ring-modified diphenyl ethers as high-affinity InhA inhibitors. Antimicrob Agents Chemother 51(10):3562–3567
33. Sullivan TJ, Truglio JJ, Boyne ME, Novichenok P, Zhang X, Stratton CF, Li HJ, Kaur T, Amin A, Johnson F, Slayden RA, Kisker C, Tonge PJ (2006) High affinity InhA inhibitors with activity against drug-resistant strains of *Mycobacterium tuberculosis*. ACS Chem Biol 1(1):43–53
34. Luckner SR, Liu N, am Ende CW, Tonge PJ, Kisker C (2010) A slow, tight-binding inhibitor of InhA, the enoyl-ACP reductase from *Mycobacterium tuberculosis*. J Biol Chem 285(19):14330–14337
35. Pan P (2012) Lead optimization and slow-onset inhibition of the enoyl-ACP reductase (InhA) from *Mycobacterium tuberculosis*. PhD Diss., Stony Brook University
36. Lu XY, Chen YD, Jiang YJ, You QD (2009) Discovery of potential new InhA direct inhibitors based on pharmacophore and 3D-QSAR analysis followed by in silico screening. Eur J Med Chem 44(9): 3718–3730
37. Kumar A, Siddiqi MI (2008) CoMFA based de novo design of pyrrolidine carboxamides as inhibitors of enoyl acyl carrier protein reductase from *Mycobacterium tuberculosis*. J Mol Model 14(10): 923–935
38. Andrade CH, Salum Lde B, Castilho MS, Pasqualoto KF, Ferreira EI, Andricopulo AD (2008) Fragment-based and classical quantitative structure-activity relationships for a series of hydrazides as antituberculosis agents. Mol Divers 12(1):47–59
39. Lu XU, Chen YD, You QD (2010) 3D-QSAR studies of arylcarboxamides with inhibitory activity on InhA using pharmacophore-based alignment. Chem Biol Drug Des 75(2): 195–203
40. Punkvang A, Saparpakorn P, Hannongbua S, Wolschann P, Beyer A, Pungpo P (2010) Investigating the structural basis of arylamides to improve potency against *M. tuberculosis* strain through molecular dynamics simulations. Eur J Med Chem 45(12):5585–5593
41. Punkvang A, Saparpakorn P, Hannongbua S, Wolschann P, Pungpo P (2010) Elucidating drug-enzyme interactions and their structural basis for improving the affinity and potency of isoniazid and its derivatives based on computer modeling approaches. Molecules 15(4):2791–2813
42. Punkvang A, Saparpakorn P, Hannongbua S, Wolschann P, Berner H, Pungpo P (2010) Insight into crucial inhibitor–enzyme interaction of arylamides as novel direct inhibitors of the enoyl ACP reductase (InhA) from *Mycobacterium tuberculosis*: computer-aided molecular design. Monatsh Chem 141(9):1029–1041
43. Pauli I, dos Santos RN, Rostirolla DC, Martinelli LK, Ducati RG, Timmers LF, Basso LA, Santos DS, Guido RV, Andricopulo AD, Norberto de Souza O (2013) Discovery of new inhibitors of *Mycobacterium tuberculosis* InhA enzyme using virtual screening and a 3D-pharmacophore-based approach. J Chem Inf Model 53(9): 2390–2401
44. Kinjo T, Koseki Y, Kobayashi M, Yamada A, Morita K, Yamaguchi K, Tsurusawa R, Gulten G, Komatsu H, Sakamoto H, Sacchettini JC, Kitamura M, Aoki S (2013) Identification of compounds with potential antibacterial activity against *Mycobacterium* through structure-based drug screening. J Chem Inf Model 53(5):1200–1212
45. Kumar UC, Bvs SK, Mahmood S, Sriram D, Kumar-Sahu P, Pulakanam S, Ballell L, Alvarez-Gomez D, Malik S, Jarp S (2013) Discovery of novel InhA reductase inhibitors: application of pharmacophore- and shape-based screening approach. Future Med Chem 5(3):249–259
46. Stigliani JL, Bernardes-Génisson V, Bernadou J, Pratviel G (2012) Cross-docking study on InhA inhibitors: a combination of Autodock Vina and PM6-DH2 simulations to retrieve bio-active conformations. Org Biomol Chem 10:6341–6349
47. Subba Rao G, Vijayakrishnan R, Kumar M (2008) Structure-based design of a novel class of potent inhibitors of InhA, the enoyl acyl carrier protein reductase from *Mycobacterium tuberculosis*: a computer modelling approach. Chem Biol Drug Des 72(5):444–449
48. Stigliani JL, Arnaud P, Delaine T, Bernardes-Génisson V, Meunier B, Bernadou J (2008) Binding of the tautomeric forms of isoniazid-NAD adducts to the active site of the *Mycobacterium tuberculosis* enoyl-ACP reductase (InhA): a theoretical approach. J Mol Graph Model 27(4):536–545
49. (2006) GaussView 03, Revision 3.07, Gaussian, Inc., Wallingford, CT
50. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Jr. JAM, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas O, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2009) Gaussian 09, revision B.01, Gaussian Inc., Wallingford CT
51. Morris GM, Goodsell DS, Huey R, Olson AJ (1996) Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. J Comput Aided Mol Des 10(4):293–304
52. Tripos International (2008) SYBYL v8.0, Tripos International, 1699 South Hanley Rd., St. Louis, MO
53. Golbraikh A, Tropsha A (2002) Beware of q^2! J Mol Graph Model 20:267–276
54. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Crowley M, Walker RC, Zhang W, Merz KM, Wang B, Hayik S, Roitberg A, Seabra G, Kolossvry I, Wong KF, Paesani F, Vanicek J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Yang L, Tan

C, Mongan J, Homak V, Cui G, Mathews DH, Seetin MG, Sagui C, Babin V, Kollman PA (2008) AMBER12. University of California, San Francisco

55. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24(16):1999–2012
56. Wang J, Wang W, Kollman PA, Case DA (2006) Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graphics Model 25(2):247–260
57. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general AMBER force field. J Comput Chem 25(9):1157–1174
58. Cornell WD, Cieplak P, Bayly CI, Kollman PA (1993) Application of RESP charges to calculation conformational energies, hydrogen bond energies, and free energies of salvation. J Am Chem Soc 115(21): 9620–9631
59. Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) A well-behaved electrostatic potential based method using charge-restraints for deriving charges: The RESP model. J Phys Chem 97(40):10269–10280
60. Wang J, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J Comput Chem 21(12):1049–1074
61. Gavernet L, Gonzalez Funes JL, Blanch LB, Estiu G, Maresca A, Supuran CT (2010) Affinity of sulfamates and sulfamides to carbonic anhydrase II isoform: experimental and molecular modeling approaches. J Chem Inf Model 50(6):1113–1122
62. Li W, Fu J, Cheng F, Zheng M, Zhang J, Liu G, Tang Y (2012) Unbinding pathways of GW4064 from human farnesoid X receptor as revealed by molecular dynamics simulations. J Chem Inf Model 52(11):3043–3052
63. Mahoney MW, Jorgensen WL (2000) A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. J Chem Phys 112(20):8910–8922
64. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. J Chem Phys 98(12):10089–10092
65. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comp Phys 23(3):327–341
66. Homeyer N, Gohlke H (2012) Free energy calculations by the molecular mechanics Poisson–Boltzmann surface area method. Mol Inf 31(2):114–122
67. Wang J, Hou T, Xu X (2006) Recent advances in free energy calculations with a combination of molecular mechanics and continuum models. Curr Comput Aided Drug Des 2(3):95–103
68. Wang J, Morin P, Wang W, Kollman PA (2001) Use of MM-PBSA in reproducing the binding free energies to HIV1 RT of TIBO derivatives and predicting the binding mode to HIV1 RT of Efavirenz by docking and MM-PBSA. J Am Chem Soc 123(22):5221–5230
69. Hou T, Wang J, Li Y, Wang W (2011) Assessing the performance of the MM/PBSA and MM/GBSA Methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. J Chem Inf Model 51(1):69–82
70. Kaledin M, Brown A, Kaledin AL, Bowman JM (2004) Normal mode analysis using the driven molecular dynamics method. II. An application to biological macromolecules. J Chem Phys 121(12):5646–5653
71. Xue W, Qi J, Yang Y, Jin X, Liu H, Yao X (2012) Understanding the effect of drug-resistant mutations of HIV-1 intasome on raltegravir action through molecular modeling study. Mol BioSyst 8:2135–2144
72. Zar JH (1998) Biostatistical analysis. Prentice Hall, Upper Saddle River

SAR and QSAR in Environmental Research, 2014
Vol. 25, No. 6, 473–488, http://dx.doi.org/10.1080/1062936X.2014.898690

Rational design of InhA inhibitors in the class of diphenyl ether derivatives as potential anti-tubercular agents using molecular dynamics simulations

P. Kamsri[a], N. Koohatammakun[a], A. Srisupan[a], P. Meewong[a], A. Punkvang[b], P. Saparpakorn[c], S. Hannongbua[c], P. Wolschann[d,e], S. Prueksaaroon[f], U. Leartsakulpanich[g] and P. Pungpo[a]*

[a]*Department of Chemistry, Faculty of Science, Ubon Ratchathani University, Ubonratchathani, Thailand;* [b]*Faculty of Liberal Arts and Sciences, Nakhon Phanom University, Nakhon Phanom, Thailand;* [c]*Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok, Thailand;* [d]*Institute for Theoretical Chemistry, University of Vienna, Vienna, Austria;* [e]*Department of Pharmaceutical Technology and Biopharmaceutics, Faculty of Life Sciences, University of Vienna, Vienna, Austria;* [f]*Department of Electrical and Computer Engineering Faculty of Engineering, Thammasat University, Phatumtani, Thailand;* [g]*National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency, Pathumthani, Thailand*

(*Received 8 October 2013; in final form 10 December 2013*)

A series of diphenyl ether derivatives were developed and showed promising potency for inhibiting InhA, an essential enoyl acyl carrier protein reductase involved in mycolic acid biosynthesis, leading to the lysis of *Mycobacterium tuberculosis*. To understand the structural basis of diphenyl ether derivatives for designing more potent inhibitors, molecular dynamics (MD) simulations were performed. Based on the obtained results, the dynamic behaviour in terms of flexibility, binding free energy, binding energy decomposition, conformation, and the inhibitor–enzyme interaction of diphenyl ether inhibitors were elucidated. Phe149, Tyr158, Met161, Met199, Val203 and NAD+ are the key residues for binding of diphenyl ether inhibitors in the InhA binding pocket. Our results could provide the structural concept to design new diphenyl ether inhibitors with better enzyme inhibitory activity against *M. tuberculosis* InhA. The present work facilitates the design of new and potentially more effective anti-tuberculosis agents.

Keywords: Diphenyl ether derivatives; MD simulations; InhA inhibitors; MM-PBSA; Anti-tuberculosis agents

1. Introduction

Tuberculosis (TB) caused by *Mycobacterium tuberculosis* (*M. tuberculosis*) remains a major global health problem. The World Health Organization reports that there were almost 8.7 million new cases and 1.4 million TB deaths in 2011. The high mortality rate of TB is caused by drug resistance and HIV co-infection [1]. Multidrug-resistant and extensively drug-resistant tuberculosis are generally thought to produce high mortality rates [2]. The most attractive target for drug design and discovery of new anti-tuberculosis agents is the enoyl-ACP reductase or InhA of *Mycobacterium tuberculosis* [3–8]. The InhA of *M. tuberculosis* catalyses the

*Corresponding author. Email: pompan_ubu@yahoo.com

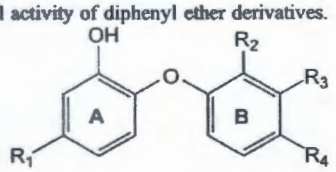
NADH-specific reduction of 2-*trans*-enoyl-ACP in the elongation cycle of the FAS II pathway (Scheme 1). This enzyme has been identified as the primary target of the most effective first-line anti-TB drug isoniazid (INH) [9–15]. As a prodrug of INH, InhA is inhibited by the active adduct of INH (INH-NAD) which is covalently formed between NAD^+ and the reactive acyl radical of INH activated by catalase-peroxidise (KatG) [16–22]. The active adduct (INH-NAD) is a highly potent inhibitor of InhA [23,24]. However, the high potency of INH for tuberculosis treatment was reduced by drug resistance from mutations in KatG [25,26]. To overcome INH resistance, new compounds which directly inhibit the InhA enzyme without requiring activation by KatG are to be seen as very promising new agents against tuberculosis. Diphenyl ethers, uncompetitive inhibitors of InhA, do not require activation by the mycobacterial KatG enzyme to generate the active form. Therefore, these inhibitors could circumvent the resistance of INH associated with the KatG mutations. A series of alkyl diphenyl ethers are potent inhibitors of InhA with K_i values less than 1 nM [27–32]. To understand the structural basis of diphenyl ether derivatives as a basis for designing more potent inhibitors, molecular dynamics (MD) simulations were performed in this work. Based on the obtained results, the dynamic behaviour, binding free energy, conformation and the inhibitor-enzyme interaction of diphenyl ether inhibitors of InhA will be provided. These results should facilitate the further modification of the diphenyl ether scaffold towards generating novel InhA inhibitors with improved inhibition potency.

2. Computational methods

2.1 *Structure and biological activity*

Four diphenyl ether derivatives (compounds 1–4) taken from the literature were used in this study [27,28,30,32]. The chemical structures and the inhibitory concentrations of the compounds required to inhibit InhA at 50% (IC_{50}) are summarized in Table 1. All chemical structures of these compounds were constructed using the standard tools available in the program GaussView 3.07 [33] and were then fully optimized using the HF/3-21G method implemented in the Gaussian 03 program [34].

Table 1. Structure and biological activity of diphenyl ether derivatives.



Compound	R_1	R_2	R_3	R_4	IC_{50} *(nM)*
1	Butyl	Cl	H	Cl	55
2	Isobutyl	Cl	H	Cl	96
3	Hexyl	H	NO_2	H	48
4	Hexyl	CH_3	H	NO_2	50
5	Benzyl	Cl	H	Cl	51
6	Octyl	H	H	H	5

2.2 *Molecular docking calculations*

To generate the binding modes of diphenyl ether derivatives in the InhA binding pocket for MD simulations, molecular docking calculations using Molecular Operating Environment (MOE) docking program [35] were used. The high-resolution X-ray crystal structure of diphenyl ether/InhA complex was obtained from Protein Data Bank (PDB code 2X23). For molecular docking parameters, alpha PMI placement strategies, affinity ΔG scoring function and 50 run were used. To ensure the reliability of the docking program, the diphenyl ether compound was removed and docked back into the InhA binding pocket. The root-mean-square deviation (RMSD) between the original and docked coordinates less than 1 Å was used to verify the reliability of docking program.

2.3 *MD simulations*

The initial coordinates for MD simulations of the diphenyl ether derivatives/InhA complexes were obtained from molecular docking calculations using the MOE program. AMBER12 [36] using the Amber03 force field [37] for InhA and the general AMBER force field (GAFF) force field [38,39] for NAD^+ and diphenyl ether inhibitors were employed for MD simulations. All missing hydrogen atoms of InhA were added using the LEaP module. To obtain the partial atomic charges of the diphenyl ether derivatives and NAD^+, the geometry optimization and electrostatic potential calculation of each compound was first calculated at the HF/6-31G* level using the Gaussian03 program. Then, RESP partial charges [40–44] were assigned using the ANTECHAMBER module implemented in AMBER12. Each complex structure was solvated by TIP3P [45] waters in an octahedral box extending up to 10 Å from each solute species. Five Na^+ cations were added to neutralize the charge of each system. The added water molecules and ions systems were minimized to relieve bad steric interactions. Non-bonded cutoff was set at 10 Å. A force of 500.0 kcal/mol was used to restrain the atom positions of all solute species. Thereafter, the whole system was minimized without restraint condition. Next, the solvated systems were gradually warmed up from 0 K to 300 K in the first 20 ps followed by maintaining the temperature at 300 K in the last 10 ps with 2 fs time steps in a constant volume boundary. The solute species in the solvated systems were restrained to their initial coordinates with a weak force constant of 10 kcal/mol $Å^2$ during the temperature warming. After minimization and heating, the position-restrained dynamics simulation using 2 fs time steps through 70 ps at 300 K under the isobaric condition was performed for each system to relax the positions of the solvent molecules. In this dynamics run, the positions of solute species were restrained with a weak force constant of 10 kcal/mol $Å^2$ during the position-restrained dynamics simulations. All the diphenyl ether/InhA complexed trajectories were run for 10 ns under the same conditions. To treat the long-range electrostatic interactions with a periodic boundary condition during the MD simulations, the Particle Mesh Ewald (PME) method [46] was applied. The cutoff distance for the long-range van der Waals interaction was set at 8 Å. The SHAKE [47] method was applied to constrain the bond lengths of hydrogen atoms attached to heteroatoms. Coordinates and energy outputs during MD simulation were printed every 2 ps. Finally, the RMSD of InhA, NAD^+ and diphenyl ether derivatives were calculated along the MD trajectory relative to the initial structures to determine the stability of the system. The binding free energy and interactions energy were calculated to study the key interactions of diphenyl ether in the InhA binding pocket.

2.4 *Binding free energy*

The binding free energy calculations between InhA and diphenyl ether inhibitors were calculated using the Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA) [48] and normal-mode [49] methods. The snapshot structures from the MD trajectory of the system for binding free energy calculations were taken from the last 2 ns of 10 ns MD simulations. 125 snapshots were extracted to calculate. The binding free energies (ΔG_{bind}) were obtained as shown in Equations (1) and (2).

$$\Delta G_{bind} = G_{com} - (G_{rec} + G_{ligand}) \quad (1)$$

$$\Delta G_{bind} = \Delta H - T\Delta S \quad (2)$$

$$\Delta H = \Delta G_{MM} + \Delta G_{solv} \quad (3)$$

where G_{com}, G_{rec} and G_{ligand} are the free energies of the complex, InhA and the diphenyl ether derivatives, respectively. In general, the binding free energy composes of an enthalpic (ΔH) and an entropic contribution ($-T\Delta S$). The enthalpic contribution (ΔH) contains the gas-phase molecular mechanics energy (ΔG_{MM}) calculated with a sander module and the solvation free energy (ΔG_{solv}) calculated with the PBSA program of the AMBER suite as shown in Equation (4).

$$\Delta G_{bind} = \Delta G_{MM} + \Delta G_{solv} - T\Delta S \quad (4)$$

ΔG_{MM} is divided into non-covalent van der Waals component (ΔG_{vdw}), electrostatic energies component (ΔG_{ele}) and bond, angle, dihedral energies (ΔG_{INT}) in Equation (5) [50].

$$\Delta G_{MM} = \Delta G_{vdw} + \Delta G_{ele} + \Delta G_{INT} \quad (5)$$

The entropy contribution ($-T\Delta S$) to the binding free energy was estimated using normal-mode analysis with the AMBER Nmode module. Due to the high computational cost in the entropy calculation, the residues around the ligand (less than 12 Å) were only considered for normal-mode calculations and 50 snapshots were used [51–53]. The contributions of entropy ($T\Delta S$) to binding free energy from changes of the translational, rotational and vibrational degrees of freedom were calculated as follows:

$$\Delta S = \Delta S_{translational} + \Delta S_{rotational} + \Delta S_{vibrational} \quad (6)$$

3. Results and discussion

3.1 *System equilibration*

To evaluate the reliable stability of the MD trajectories, the RMSDs for all atoms of InhA, NAD^+ cofactor and diphenyl ether derivatives relative to the initial minimized structure over the 10 ns of simulation times were calculated and plotted in Figure 1. There are three solute species in each MD system including InhA, NAD^+ and inhibitor. The plateau characteristic of the RMSD plot over the simulation time is the criteria to indicate the equilibrium state of each solute species. For the equilibrium state of each MD system, the RMSD plots of all solute species have to reach the plateau characteristic. InhA, NAD^+ and inhibitor in each system reach the equilibrium state at a different time (Figure 1). For the system of compound 1, NAD^+ and this compound reach equilibrium at an early time point, whereas InhA reaches the equilibrium state after 2 ns (Figure 1(a)). Therefore, after 2 ns the RMSD plots of all solute species reach the plateau characteristic, indicating the equilibrium state of this MD system. In

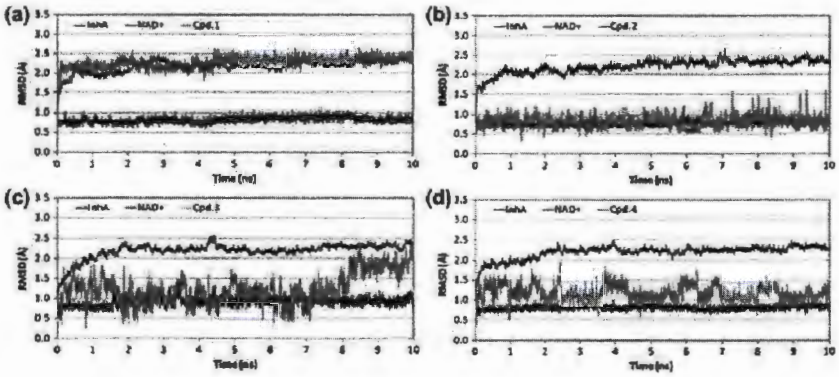


Figure 1. RMSDs of diphenyl ether derivatives, compounds 1 (a), 2 (b), 3 (c) and 4 (d) complexed with the InhA.

the case of compound 2, its MD system reaches equilibrium after 1 ns (Figure 1(b)). For compounds 3 and 4, the MD systems reach equilibrium after 2 ns (Figures 1(c) and 1(d)). The RMSD plots of these compounds over 10 ns show large fluctuations in the range of about 0.5–2.5 Å. This result could be accounted for by the greater flexibility of the long hexyl chain. Therefore, the data in terms of binding free energy, interaction energy and structure of each system after an equilibrium state were analysed.

3.2 *Structural flexibility of the InhA binding pocket*

The binding cavity volumes of InhA complexed with the diphenyl ether derivatives were calculated to study the flexibility of the InhA binding pocket using the computed atlas of surface topography of proteins (CASTp) method (http://sts.bioengr.uic.edu/castp/index.php) [54–56]. Amino acid residues within 10 Å from ligands-NAD^+ cofactor were used to calculate the binding cavity volumes of the InhA binding pockets. The binding cavity volumes of InhA complexed with six diphenyl ether derivatives are summarized in Table 2. Binding cavity volumes ranging from 1763 $Å^3$ to 2465 $Å^3$ of six inhibitors were found. Moreover, the binding cavity volume of *trans*-2-hexadecenoyl-(n-acetyl-cysteamine)- thioester substrate (PDB code

Table 2. The binding cavity volume of diphenyl ether inhibitors.

Compound	*Ligand Surface Area* ($Å^2$)	*Binding cavity volume* ($Å^3$)
1	546	2,318
2	530	2,042
3	574	2,465
4	611	2,039
5	575	1,763
6	612	1,876
trans-2-hexadecenoyl-(n-acetyl-cysteamine)- thioester substrate	663	1,902

1BVR) [57] is 1902 Å^3. In a previous study [58], the binding cavity volumes of direct InhA inhibitors obtained from the available X-ray structures of InhA inhibitors were found to range from 1597 Å^3 to 3047 Å^3. These results indicate that the binding pocket of InhA is flexible enough to bind with the substrate and diphenyl ether inhibitors. The binding cavity volume of compound 3 in the InhA binding pocket is higher than that of the highest compound 6 and *trans*-2-hexadecenoyl-(n-acetyl-cysteamine)- thioester substrate. In addition, the binding cavity volumes of low-potency compounds 1 and 2 are higher than *trans*-2-hexadecenoyl-(n-acetyl-cysteamine)- thioester substrate and the most potent diphenyl ether inhibitor. The most potent compound 6 shows a binding cavity volume comparable with that of the *trans*-2-hexadecenoyl-(n-acetyl-cysteamine)- thioester substrate. Compounds 1–4, creating InhA cavity volumes larger than that of the substrate, show poor activity in inhibiting InhA. To reveal the correlation between the InhA cavity volume, the molecular size of compounds 1–6 and their inhibitory activities, the molecular surface areas of these compounds were calculated using HYPERCHEM 7.51 [59], reported in Table 2. Apart from the *trans*-2-hexadecenoyl-(*n*-acetyl-cysteamine)- thioester substrate, possessing the largest molecular surface area with 663 Å^2, the molecular surface area of compound 6 with 612 Å^2 is the largest one. Compound 4 has a surface area of 611 Å^2, close to that of compound 6, but its cavity volume is larger. In case of compounds 1–3, these compounds show smaller surface areas than that of compound 6 but their pocket cavity volumes are larger. These results imply that compound 6, having a larger molecular size in its smaller cavity volume, could induce the InhA pocket to fit well, with better binding than compounds 1–4. Accordingly, compound 6 should produce more interactions with amino acid residues in the InhA pocket than compounds 1–4, leading to better inhibitory activity of this compound. In contrast, compound 5 has a smaller surface area than compound 6, but its cavity volume is close to that of compound 6. This result implies that the binding of compound 5 could not properly fit the InhA pocket, leading to loss of activity as compared with compound 6.

3.3 *Binding free energy calculations*

To evaluate the binding affinity of diphenyl ether derivatives in the InhA binding pocket, the binding free energies (ΔG_{bind}) were calculated using the MM-PBSA method. The binding free energies (ΔG_{bind}) were calculated by Equation (2). The entropic (-TΔS), enthalpic (ΔH) energy and binding free energy of the diphenyl ether/InhA complexes are listed in Table 3. The binding free energies of compounds 1, 2, 3 and 4 bound to the InhA are −10.3, −7.9, −11.6 and −13.7 kcal/mol, respectively. Moreover, the binding free energy of the X-ray crystal structures of compound 5/InhA and compound 6/InhA complexes were calculated to compare with those of compounds 1, 2, 3 and 4. The calculated and experimental binding free energies were compared as shown in Table 3. Moreover, a good linear correlation (r^2 = 0.78) between the experimental IC_{50} and calculated binding free energy is presented in Figure 2. It is notable that the calculated binding free energies of the selected compounds are in the correct order as compared with the IC_{50} values. The obtained results could be successfully used to validate the MD procedure in this study.

According to the energy components of the binding free energies listed in Table 2, the van der Waals energy of diphenyl ether derivatives in the InhA binding pocket provides the greatest contribution to the binding free energy due to the high hydrophobicity of the InhA binding pocket [58]. The van der Waals energy of diphenyl ether compounds 3, 4, 5 and 6 is lower than −50 kcal/mol. These results indicate that the increase of the hydrophobicity of

Table 3. Binding free energies in kcal/mol computed by the MM-PBSA method.

	Contribution							
Cpd.	ΔG_{ele}	ΔG_{vdw}	ΔG_{MM}	ΔG_{sol}	ΔH	$-T\Delta S$	ΔG_{bind}	ΔG_{exp} [a]
1	-10.5 ± 2.5	-45.2 ± 2.3	-55.6 ± 5.6	27.2 ± 3.5	-28.4 ± 3.0	18.2 ± 0.8	-10.3 ± 1.4	-10.0
2	-12.4 ± 2.9	-47.1 ± 2.2	-59.4 ± 3.0	31.2 ± 2.5	-28.2 ± 2.8	20.3 ± 0.8	-8.0 ± 0.8	-9.6
3	-12.9 ± 3.2	-55.8 ± 2.7	-68.7 ± 3.5	31.2 ± 4.4	-37.5 ± 4.0	25.9 ± 1.9	-11.6 ± 1.5	-10.0
4	-9.0 ± 2.7	-50.3 ± 2.5	-59.3 ± 3.0	24.3 ± 2.2	-34.9 ± 2.8	21.3 ± 1.1	-13.7 ± 1.3	-10.1
5	-12.5 ± 2.0	-51.2 ± 2.2	-63.6 ± 3.2	32.8 ± 3.2	-30.8 ± 2.8	19.0 ± 1.0	-11.8 ± 1.3	-10.0
6	-16.6 ± 3.0	-55.3 ± 2.4	-71.9 ± 3.4	24.3 ± 2.5	-36.1 ± 5.5	22.1 ± 1.4	-13.9 ± 1.5	-11.4

[a] derived from $\Delta G = RT \ln[IC_{50}]$, R represents the gas constant (1.988 cal/mol K), T represents the temperature (300 K).

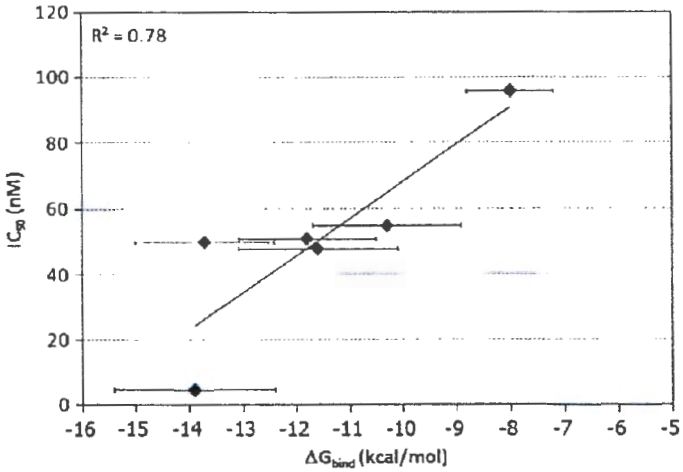


Figure 2. Correlation of experimental IC_{50} and calculated binding free energy using MM-PBSA method

inhibitors leads to an increase in the binding affinity of inhibitors in the InhA pocket. This result agrees with the experimental biological activity [27–32].

3.4 *Per-residue binding energy decomposition*

The binding energies between diphenyl ether derivatives with each residue in the InhA pocket were calculated by the MM-GBSA method to understand the key interactions for binding of protein–ligand complexes. Figure 3 displays the binding energy decomposition of diphenyl ether derivatives. The obtained results indicate that nine residues including Phe97, Phe149, Tyr158, Met161, Lys165, Ala198, Met199, Val203 and NAD^+ cofactor show lower interaction energies with diphenyl ether inhibitors. Figure 4 shows the contribution of the van der Waals and electrostatic energies on the binding of diphenyl ether derivatives in the InhA binding pocket. Most of the residues show the van der Waals energy lower than the electrostatic energy. This result indicates that the van der Waals interactions have an important role in the binding of diphenyl ether derivatives in the InhA binding pocket. It is important to note that all repulsive energies observed for each of the compounds are generated from electrostatic energy. In particular, the compounds with lower activity show greater repulsive energies.

3.5 *Binding interactions analysis*

To elucidate the dynamic behaviour of the diphenyl ether inhibitors in the InhA pocket, the binding modes and interaction of inhibitors in the InhA pocket obtained from MD simulations were analysed. As shown in Figures 5(a)–(d), the binding modes of all compounds in the InhA pocket obtained from MD simulations are of the same manner. For compound I, the phenyl B ring reveals a pi–pi interaction with the pyridine amide ring of NAD^+ cofactor. The hydrogen bond between the OH group of the phenyl A ring and NAD^+ cofactor is found.

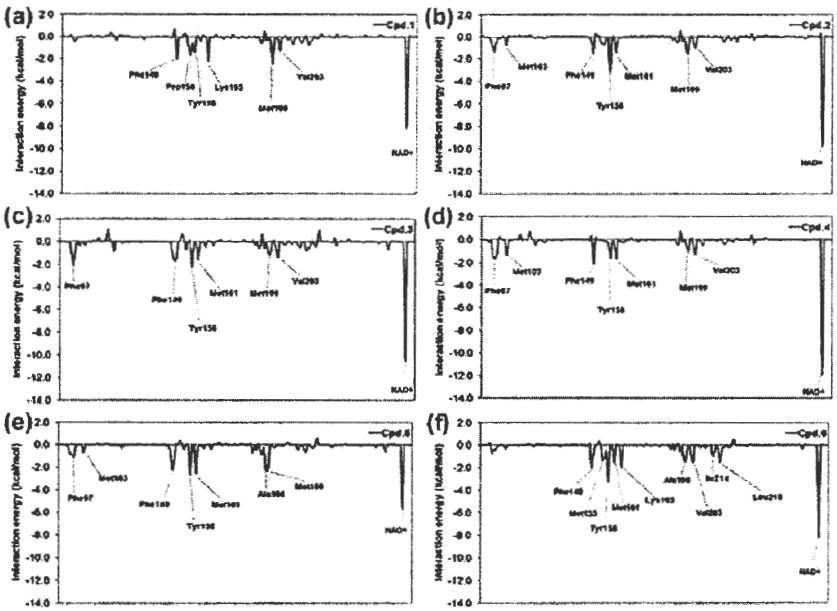


Figure 3. Per-residue binding energy decomposition of the selected diphenyl ether derivatives, compounds 1 (a), 2 (b), 3 (c) 4 (d), 5 (e) and 6 (f) using the MM-GBSA method.

The substituent R_2 with Cl atom causes a slight repulsive interaction with the phenyl ring of Phe149. For compound 2, the binding interactions between diphenyl ether with NAD^+ can be explained as follows; (i) hydrogen bond interaction between the –OH group of the phenyl A ring with the OH fragment of NAD^+; (ii) the pi–pi interaction of the phenyl A ring and the pyridine amide ring of NAD^+ cofactor; and (iii) a weak hydrogen bond of the Cl substituent and the CH of NAD^+ cofactor. With regard to compound 3, the crucial interactions of diphenyl ether compound 3 are hydrogen bond, pi–pi interactions and hydrophobic interactions with NAD^+ cofactor. The hydrogen bond interaction of the OH group on the phenyl A ring of diphenyl ether and the OH of the ribose fragment of NAD^+ cofactor is found. The pi–pi interactions between the phenyl A ring of diphenyl ether and the pyridine amide ring of NAD^+ cofactor can be formed, and the hydrophobic interactions between the phenyl B ring and all parts of NAD^+ cofactor can be observed. Importantly, the hydrogen bond interactions between the NO_2 substituent of diphenyl ether and the CH_2 of Met98 are found. Moreover, numerous hydrophobic interactions between the hexyl substituent attached to the phenyl A ring of diphenyl ether with Tyr158 are observed. Compared with compound 3, compound 4 also reveals similar interactions. However, only one hydrogen bond interaction between the NO_2 substituent of diphenyl ether and Met98 is formed. Weaker interactions between diphenyl ether with Tyr158 are observed.

Therefore, we can summarize the binding interaction of compounds 1–4 as follows. With regard to the phenyl B ring, the ring is surrounded by amino acid residues Met98, Phe97, Met161 and Ala198. The crucial interactions of this fragment are hydrogen bond interactions

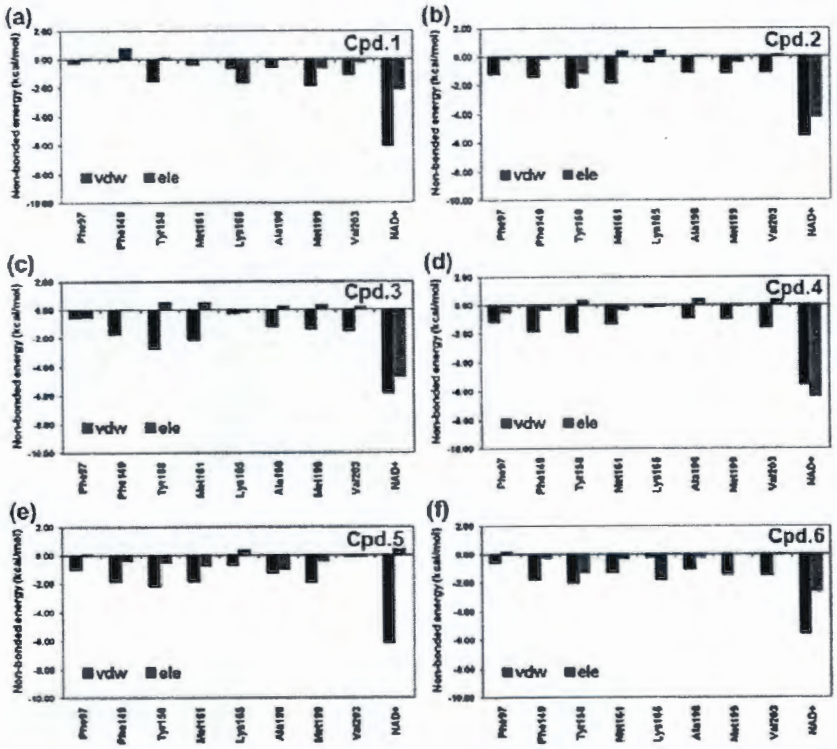


Figure 4. The plots of the decomposition energies in terms of van der Waals energy (vdw) and electrostatic energy (ele) for diphenyl ether derivatives of compounds 1 (a), 2 (b), 3 (c) 4 (d), 5 (e) and 6 (f).

and van der Waals interactions. Regarding the phenyl A ring bearing the alkyl chain, the R_1 substituent of each compound is placed in the pocket of Phe149, Tyr158, Met199 and Val203. These residues show van der Waals interaction energies lower than electrostatic interaction energies for all compounds (Figure 4). This result could be accounted for by hydrophobic interactions between the alkyl chain at the R_1 substituent with surrounding residues. The OH group at the phenyl A ring of compounds 1–4 lies among the OH groups of Tyr158 and the ribose fragment of NAD^+. Based on binding energy decomposition of each residue, NAD^+ shows the lowest interaction energy for all compounds (Figure 3), indicating the largest contribution of NAD^+ on binding of compounds 1–4 in the InhA pocket. All compounds can form two strong interactions with NAD^+; the first is the hydrogen bond between the OH group of the compound with the OH group of the ribose fragment. Another important interaction is the pi–pi interaction between the phenyl A ring of compounds with a pyridine amide ring of NAD^+ (Figure 5). Considering the contribution of NAD^+ in terms of van der Waals and electrostatic interaction energies on the binding of compounds 1–4 in the InhA binding pocket (Figure 4), both van der Waals and electrostatic interactions show large attractive energy for all compounds. These results indicate that the pi–pi interaction and the hydrogen

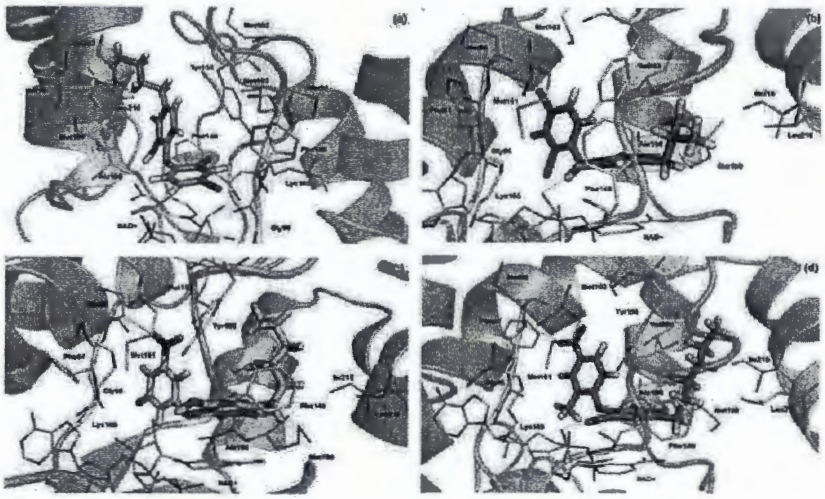


Figure 5. The MD structure averaged over the last 2 ns of compounds 1 (a), 2 (b), 3 (c) and 4 (d) in the InhA binding pocket.

bond interaction with NAD^+ play important roles in the binding of diphenyl ether derivatives in the InhA pocket. Accordingly, apart from the phenyl B ring playing important role for forming strong hydrogen bond interactions to Met98, the key fragment for binding of diphenyl ether derivatives in the InhA pocket is the phenyl A ring and the OH group, that generate the pi–pi interaction and the hydrogen bond interaction with NAD^+, respectively.

3.6 *Structural basis of diphenyl ether derivatives for rational inhibitor design*

Reliable binding modes of diphenyl ether derivatives in the InhA binding pocket were obtained from MD simulations. The superposition of MD structures of compounds 1–4 and the X-ray structures of compounds 5–6 is shown in Figure 6. As mentioned above, the key fragment for binding of diphenyl ether derivatives in the InhA pocket is the phenyl A ring and the OH group for generating the pi–pi interaction and the hydrogen bond interaction with NAD^+, respectively. The R_1 substituent on the phenyl A ring is oriented in the hydrophobic pocket of Phe149, Met155, Pro156, Ala157, Tyr158, Pro193, Met199, Val203, Leu207, Ile215 and Leu218 (Figure 6). Therefore, an R_1 substituent with highly lipophilic property, i.e. hexyl, heptyl and octyl, is optimal for forming hydrophobic interactions in this pocket. For the R_2, R_3 and R_4 substituents at the phenyl B ring, compound 6, the most active compound in this series, contains hydrogen atom at these positions. A hydrogen atom at the R_2 position of compound 6 could preferably form a hydrophobic interaction with the methyl side chain of Ala198 and pyrophosphate group of NAD^+. The CH_3 group attached to the same position of compound 4 located close to these residues could also possibly form hydrophobic interactions with Ala198 and pyrophosphate group of NAD^+. The Cl atom attached to the R_2 position with respect to compounds 2 and 5 generates a repulsive interaction with the pyrophosphate group of NAD^+. Accordingly, the hydrophobic substituents such as H and CH_3

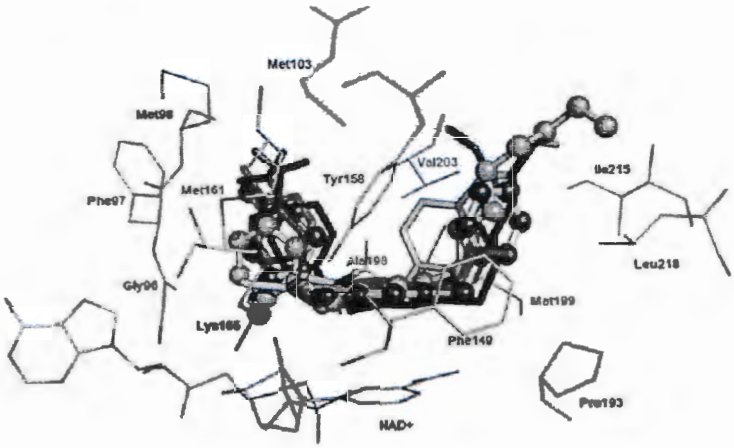
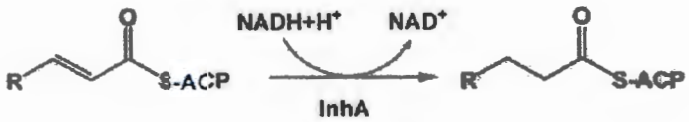


Figure 6. Superimposition of diphenyl ether derivatives, compounds 1-6, in the InhA binding pocket.



Scheme 1. The biochemical mechanism of the InhA enzyme.

should be optimal to interact with both the methyl side chain of Ala198 and the pyrophosphate group of NAD^+. In the case of the R_3 position, the H substituent at this position of compound 6, as well of compounds 1, 2, 4 and 5, points to the carbonyl backbone of Gly96. In contrast, the NO_2 substituent of compound 3 flips away from this direction (Figure 6). Therefore, the H substituent as a hydrogen bond donor substituent would be the most suitable for the R_3 position to possibly form a hydrogen bond interaction with the carbonyl backbone of Gly96. For the R_4 substituent, the H substituent at this position of compound 6, and Cl and NO_2 substituents of compounds 2, 5 and 4, respectively, is oriented to the NH backbone of Met98. Therefore, at the R_4 position, Cl and NO_2 substituents and a hydrogen bond acceptor substituent that can form a hydrogen bond interaction with the NH backbone of Met98 would be better for inhibitor-enzyme interaction than the H substituent. Moreover, the modification of the R_2, R_3 and R_4 substituents at the phenyl B ring should be of optimal size because the steric effect of each substituent may cause the loss of the pi–pi interaction and the hydrogen bond interaction of the phenyl A ring bearing the OH group with NAD^+.

4. Conclusions

MD simulations were successful in modelling reliable binding modes, binding energy, and InhA–ligand interactions of diphenyl ether in the InhA binding pocket. Based on MD

simulations, the binding cavity volumes of diphenyl ether derivatives in the InhA pocket are important for binding. The per-residue binding energy suggests that the van der Waals interaction shows a greater contribution to the binding of diphenyl ether derivatives in the InhA pocket than the electrostatic interaction. Based on the detailed MD simulations analysis, NAD^+ is the key residue for binding of diphenyl ether inhibitors in the InhA binding pocket. Therefore, the results obtained from this study should facilitate the further modification of a diphenyl ether scaffold for generating novel InhA inhibitors with improved inhibition potency.

Acknowledgements

This paper was presented at the 7th CMTPI conference in Seoul, 8–12 October 2013. This research was supported by the Thailand Research Fund (DBG5680003, RTA5380010) and the National Research Council of Thailand. The financial support from Royal Golden Jubilee Ph.D. Program (PHD/004/2554) to P. Kamsri is gratefully acknowledged. Faculty of science, Ubon Ratchathani University, University of Vienna, ASEA-Uninet, NECTEC and BIOTEC are gratefully acknowledged for supporting this research.

References

[1] World Health Organization, *Global tuberculosis report 2012*, Available at http://apps.who.int/iris/bitstream/10665/75938/1/9789241564502_eng.pdf.

[2] World Health Organization, *Multidrug and extensively drug-resistant TB (M/XDR-TB): 2010 global report on surveillance and response*, Available at http://whqlibdoc.who.int/publications/2010/9789241599191_eng.pdf.

[3] J.W. Campbell and J.E. Cronan Jr, *Bacterial fatty acid biosynthesis: Targets for antibacterial drug discovery*, Annu. Rev. Microbiol. 55 (2001), pp. 305–332.

[4] R.J. Heath and C.O. Rock, *Fatty acid biosynthesis as a target for novel antibacterials*, Curr. Opin. Invest. Drugs. 5 (2004), pp. 146–153.

[5] S.W. White, J. Zheng, Y.M. Zhang, and C.O. Rock, *The structural biology of type II fatty acid biosynthesis*, Annu. Rev. Biochem. 74 (2005), pp. 791–831.

[6] Y.M. Zhang, Y.J. Lu, and C.O. Rock, *The reductase steps of the type II fatty acid synthase as antimicrobial targets*, Lipids 39 (2004), pp. 1055–1060.

[7] L. Wen, J.N. Chmielowski, K.C. Bohn, J.K. Huang, Y.N. Timsina, P. Kodali, and A.K. Pathak, *Functional expression of* Francisella tularensis *FabH and FabI, potential antibacterial targets*, Protein Expr. Purif. 65 (2009), pp. 83–91.

[8] H.T. Wright and K.A. Reynolds, *Antibacterial targets in fatty acid biosynthesis*, Curr. Opin. Microbiol. 10 (2007), pp. 447–453.

[9] A. Quemard, J.C. Sacchettini, A. Dessen, C. Vilcheze, R. Bittman, W.R. Jacobs Jr, and J.S. Blanchard, *Enzymatic characterization of the target for isoniazid in* Mycobacterium tuberculosis, Biochemistry. 34 (1995), pp. 8235–8241.

[10] D.A. Rozwarski, G.A. Grant, D.H. Barton, W.R. Jacobs Jr, and J.C. Sacchettini, *Modification of the NADH of the isoniazid target (InhA) from* Mycobacterium tuberculosis, Science 279 (1998), pp. 98–102.

[11] C. Vilchèze, F. Wang, M. Arai, M.H. Hazbón, R. Colangeli, L. Kremer, T.R. Weisbrod, D. Alland, J.C. Sacchettini, and W.R. Jacobs Jr, *Transfer of a point mutation in* Mycobacterium tuberculosis *inhA resolves the target of isoniazid*, Nat. Med. 12 (2006), pp. 1027–1029.

[12] A. Dessen, A. Quémard, J.S. Blanchard, W.R. Jacobs Jr, and J.C. Sacchettini, *Crystal structure and function of the isoniazid target of* Mycobacterium tuberculosis, Science 267 (1995), pp. 1638–1641.

[13] B. Lei, C.J. Wei, and S.C. Tu, *Action mechanism of antitubercular isoniazid. Activation by* Mycobacterium tuberculosis *KatG, isolation, and characterization of inhA inhibitor*, J. Biol. Chem. 275 (2000), pp. 2520–2526.

[14] K. Johnsson, D.S. King, and P.G. Schultz, *Studies on the mechanism of action of isoniazid and ethionamide in the chemotherapy of tuberculosis*, J. Am. Chem. Soc. 117 (1995), pp. 5009–5010.

[15] A. Quémard, A. Dessen, M. Sugantino, W.R. Jacobs Jr, J.C. Sacchetini, and J.S. Blanchard, *Binding of catalase-peroxidase-activated isoniazid to wild-type and mutant* Mycobacterium tuberculosis *enoyl-ACP reductases*, J. Am. Chem. Soc. 118 (1996), pp. 1561–1562.

[16] B. Saint-Joanis, H. Souchon, M. Wilming, K. Johnsson, P.M. Alzari, and S.T. Cole, *Use of site-directed mutagenesis to probe the structure, function and isoniazid activation of the catalase/peroxidase, KatG, from* Mycobacterium tuberculosis, Biochem. J. 338 (1999), pp. 753–760.

[17] X. Zhao, H. Yu, S. Yu, F. Wang, J.C. Sacchettini, and R.S. Magliozzo, *Hydrogen peroxide-mediated isoniazid activation catalyzed by* Mycobacterium tuberculosis *catalase-peroxidase (KatG) and its S315T mutant*, Biochemistry 45 (2006), pp. 4131–4140.

[18] C. Metcalfe, I.K. Macdonald, E.J. Murphy, K.A. Brown, E.L. Raven, and P.C. Moody, *The tuberculosis prodrug isoniazid bound to activating peroxidases*, J. Biol. Chem. 283 (2008), pp. 6193–6200.

[19] B.K. Sinha, *enzymatic activation of hydrazine derivatives*, J. Biol. Chem. 258 (1983), pp. 796–801.

[20] M. Nguyen, C. Claparols, J. Bernadou, and B. Meunier, *A fast and efficient metal-mediated oxidation of isoniazid and identification of isoniazid-NAD(H) adducts*, Chem. Bio. Chem. 2 (2001), pp. 877–883.

[21] B. Heym, Y. Zhang, S. Poulet, D. Young, and S.T. Cole, *Characterization of the katG gene encoding a catalase-peroxidase required for the isoniazid susceptibility of* Mycobacterium tuberculosis, J. Bacteriol. 175 (1993), pp. 4255–4259.

[22] K. Johnsson and P.G. Schultz, *Mechanistic studies of the oxidation of isoniazid by the catalase peroxidase from* Mycobacterium tuberculosis, J. Am. Chem. Soc. 116 (1994), pp. 7425–7426.

[23] G.S. Timmins and V. Deretic, *Mechanisms of action of isoniazid*, Mol. Microbiol. 62 (2006), pp. 1220–1227.

[24] K. Johnsson, W.A. Froland, and P.G. Schultz, *Overexpression, purification, and characterization of the catalase-peroxidase KatG from* Mycobacterium tuberculosis, J. Biol. Chem. 272 (1997), pp. 2834–2840.

[25] A.I. De La Iglesia and H.R. Morbidoni, *Mechanisms of action of and resistance to rifampicin and isoniazid in* Mycobacterium tuberculosis: *New information on old friends*, Rev. Argent Microbiol. 38 (2006), pp. 97–109.

[26] A. Banerjee, E. Dubnau, A. Quemard, V. Balasubramanian, K.S. Um, T. Wilson, D. Collins, G. de Lisle, and W.R. Jacobs Jr, *InhA, a gene encoding a target for isoniazid and ethionamide in* Mycobacterium tuberculosis, Science. 263 (1994), pp. 227–230.

[27] J.S. Freundlich, F. Wang, C. Vilchèze, G. Gulten, R. Langley, G.A. Schiehser, D.P. Jacobus, W.R. Jacobs Jr, and J.C. Sacchettini, *Triclosan derivatives: Towards potent inhibitors of drug-sensitive and drug-resistant* Mycobacterium tuberculosis, Chem. Med. Chem. 4 (2009), pp. 241–248.

[28] C.W. Am Ende, S.E. Knudson, N. Liu, J. Childs, T.J. Sullivan, M. Boyne, H. Xu, Y. Gegina, D.L. Knudson, F. Johnson, C.A. Peloquin, R.A. Slayden, and P.J. Tonge, *Synthesis and in vitro antimycobacterial activity of B-ring modified diaryl ether InhA inhibitors*, Bioorg. Med. Chem. Lett. 18 (2008), pp. 3029–3033.

[29] M.E. Boyne, T.J. Sullivan, C.W. Am Ende, H. Lu, V. Gruppo, A.G. Amin, D. Chatterjee, A. Lenaerts, P.J. Tonge, and R.A. Slayden, *Targeting fatty acid biosynthesis for the development of novel chemotherapeutics against* Mycobacterium tuberculosis: *Evaluation of A-ring-modified diphenyl ethers as high-affinity InhA inhibitors*, Antimicrob. Agents Chemother. 51 (2007), pp. 3562–3567.

[30] T.J. Sullivan, J.J. Truglio, M.E. Boyne, P. Novichenok, X. Zhang, C.F. Stratton, H.J. Li, T. Kaur, A. Amin, F. Johnson, R.A. Slayden, C. Kisker, and P.J. Tonge, *High affinity InhA inhibitors with activity against drug-resistant strains of* Mycobacterium tuberculosis, ACS Chem. Biol. 1 (2006), pp. 43–53.

[31] S.R. Luckner, N. Liu, C.W. am Ende, P.J. Tonge, and C.A. Kisker, *A slow, tight-binding inhibitor of InhA, the enoyl-ACP reductase from* Mycobacterium tuberculosis, J. Biol. Chem. 285 (2010), pp. 14330–14337.

[32] P. Pan, *Lead optimization and slow-onset inhibition of the Enoyl-ACP reductase (InhA) from* Mycobacterium tuberculosis, Ph.D. Diss., Stony Brook University, 2012.

[33] *GaussView 03, Revision 3.07*, Gaussian, Inc., Wallingford, 2006.

[34] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery, T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, and J.A. Pople, *Gaussian 03*, Gaussian Inc., Wallingford, 2004.

[35] *Molecular Operating Environment (MOE)*, Chemical Computing Group Inc., Canada, 2012.

[36] D.A. Case, T.A. Darden, T.E. Cheatham, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, R.C. Walker, W. Zhang, K.M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvry, K.F. Wong, F. Paesani, J. Vanicek, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, and P.A. Kollman, *AMBER12*, University of California, San Francisco, 2008.

[37] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P.A. Kollman, *A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations*, J. Comput. Chem. 24 (2003), pp. 1999–2012.

[38] J. Wang, W. Wang, P.A. Kollman, and D.A. Case, *Automatic atom type and bond type perception in molecular mechanical calculations*, J. Mol. Graphics Model. 25 (2006), pp. 247–260.

[39] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, and D.A. Case, *Development and testing of a general AMBER force field*, J. Comput. Chem. 25 (2004), pp. 1157–1174.

[40] W.D. Cornell, P. Cieplak, C.I. Bayly, and P.A. Kollman, *Application of RESP charges to calculation conformational energies, hydrogen bond energies, and free energies of salvation*, J. Am. Chem. Soc. 115 (1993), pp. 9620–9631.

[41] C.I. Bayly, P. Cieplak, W.D. Cornell, and P.A. Kollman, *A well-behaved electrostatic potential based method using charge-restraints for deriving charges: The RESP model*, J. Phys. Chem. 97 (1993), pp. 10269–10280.

[42] J. Wang, P. Cieplak, and P.A. Kollman, *How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?* J. Comput. Chem. 21 (2000), pp. 1049–1074.

[43] L. Gavernet, J.L.G. Funes, L.B. Blanch, G. Estiu, A. Maresca, and C.T. Supuran, *Affinity of sulfamates and sulfamides to carbonic anhydrase II isoform: Experimental and molecular modeling approaches*, J. Chem. Inf. Model. 50 (2010), pp. 1113–1122.

[44] W. Li, J. Fu, F. Cheng, M. Zheng, J. Zhang, G. Liu, and Y. Tan, *Unbinding pathways of GW4064 from Human Farnesoid X Receptor as revealed by molecular dynamics simulations*, J. Chem. Inf. Model. 52 (2012), pp. 3043–3052.

[45] M.W. Mahoney and W.L. Jorgensen, *A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions*, J. Chem. Phys. 112 (2000), pp. 8910–8922.

[46] T. Darden, D. York, and L. Pedersen, *Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems*, J. Chem. Phys. 98 (1993), pp. 10089–10092.

[47] J.-P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-Alkanes*, J. Comp. Phys. 23 (1977), pp. 327–341.

[48] N. Homeyer and H. Gohlke, *Free energy calculations by the molecular mechanics Poisson–Boltzmann surface area method*, Mol. Inf. 31 (2012), pp. 114–122.

[49] M. Kaledin, A. Brown, A.L. Kaledin, and J.M. Bowman, *Normal mode analysis using the driven molecular dynamics method. II. An application to biological macromolecules*, J. Chem. Phys. 121 (2004), pp. 5646–5653.

[50] J. Wang, T. Hou, and X. Xu, *Recent advances in free energy calculations with a combination of molecular mechanics and continuum models*, Curr. Comput. Aided Drug Des. 2 (2006), pp. 95–103.

[51] J. Wang, P. Morin, W. Wang, and P.A. Kollman, *Use of MM-PBSA in reproducing the binding free energies to HIV1 RT of TIBO derivatives and predicting the binding mode to HIV1 RT of Efavirenz by docking and MM-PBSA*, J. Am. Chem. Soc. 123 (2001), pp. 5221–5230.

[52] T. Hou, J. Wang, Y. Li, and W. Wang, *Assessing the performance of the MM/PBSA and MM/GBSA Methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations*, J. Chem. Inf. Model. 51 (2011), pp. 69–82.

[53] W. Xue, J. Qi, Y. Yang, X. Jin, H. Liu, and X. Yao, *Understanding the effect of drug-resistant mutations of HIV-1 intasome on raltegravir action through molecular modeling study*, Mol. BioSyst. 8 (2012), pp. 2135–2144.

[54] J. Liang, H. Edelsbrunner, and C. Woodward, *Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design*, Protein Sci. 7 (1998), pp. 1884–1897.

[55] J. Liang, H. Edelsbrunner, P. Fu, P.V. Sudhakar, and S. Subramaniam, *Analytical shape computing of macromolecules II: Identification and computation of inaccessible cavities inside proteins*, Proteins 33 (1998), pp. 18–29.

[56] J. Liang, H. Edelsbrunner, P. Fu, P.V. Sudhakar, and S. Subramaniam, *Analytical shape computing of macromolecules I: Molecular area and volume through alpha shape*, Proteins 33 (1998), pp. 1–17.

[57] D.A. Rozwarski, C. Vilchèze, M. Sugantino, R. Bittman, and J.C. Sacchettini, *Crystal structure of the* Mycobacterium tuberculosis *enoyl-ACP reductase, InhA, in complex with NAD+ and a C16 fatty acyl substrate*, J. Biol. Chem. 274 (1999), pp. 15582–15589.

[58] I. Pauli, R.N. Dos Santos, D.C. Rostirolla, L.K. Martinelli, R.G. Ducati, L.F. Timmers, L.A. Basso, D.S. Santos, R.V. Guido, A.D. Andricopulo, and O. Norberto de Souza, *Discovery of new inhibitors of* Mycobacterium tuberculosis *InhA enzyme using virtual screening and a 3D-pharmacophore-based approach*, J. Chem. Inf. Model. 53 (2013), pp. 2390–2401.

[59] *HyperChem (TM) Professional 7.51*, Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA.

molecular informatics

Special Issue EuroQSAR

DOI: 10.1002/minf.201200132

Key Structural Features of Azanaphthoquinone Annelated Pyrrole Derivative as Anticancer Agents Based on the Rational Drug Design Approaches

Pharit Kamsri,[a] Auradee Punkvang,[b] Nipawan Pongprom,[a] Apinya Srisupan,[a] Patchreenart Saparpakom,[c] Supa Hannongbua,[d] Peter Wolschann,[d, e] and Pompan Pungpo*[a]

Abstract: Azanaphthoquinone annelated pyrrole derivatives have been developed and synthesized with a continuous attempt to develop novel DNA intercalating agents as anti-cancer compounds with lower organ toxicity. With the remarkable antiproliferative activity of synthesized azanaphthoquinone annelated pyrrole derivatives, a structurally novel scaffold of these compounds is appropriated for further development of novel anti-cancer agents. Therefore, in the present study, 3D QSAR study (CoMSIA) was applied on 28 azanaphthoquinone annelated pyrrole derivatives to evaluate the structural requirement of these compounds. The resulting CoMSIA model is satisfied with r^2 of 0.99 and q^2 of 0.65. The interpretation of CoMSIA contours reveals the significant importance of steric, electrostatic, hydrophobic and hydrogen acceptor descriptors on the activities of azanaphthoquinone annelated pyrrole derivatives. Remarkably, the structural requirement of six substituent positions on the azanaphthoquinone annelated pyrrole scaffold was elucidated here. This result is the useful concept for design of new and more active azanaphthoquinone annelated pyrrole derivatives. Moreover, MD simulations using AMBER program were performed to model the binding of azanaphthoquinone annelated pyrrole derivatives in the intercalation site of the DNA duplex. Based on MD simulations, the information in terms of ligand-DNA interaction, complex structure and binding free energy was provided in this work. Therefore, the integrated results are informative for further modification of azanaphthoquinone annelated pyrrole scaffold leading to gain novel azanaphthoquinone annelated pyrrole derivatives possessing better antiproliferative activity.

Keywords: QSAR · MD simulations · Anti-cancer · Intercalating agent · Azanaphthoquinone

1 Introduction

Mitoxantrone displays the important role for fighting a variety of cancers and is widely used as synthetic DNA intercalating agent [1–3]. However, the fighting of cancer using mitoxantrone and related anthraquinone compounds is limited by their cardiotoxicity and drug resistance which are the major drawback of these anti-cancer agents [4]. With the aim to search for analogs with lower cardiotoxicity, the new analogs of mitoxantrone were developed leading to discover an aza-anthracene-9,10-dione [5]. This compound showed a better therapeutic index and lower cardiotoxicity as compared with mitoxantrone [6]. With the continuous effort to develop novel DNA intercalating agents, compounds based on the azanaphthoquinone annelated pyrrole scaffold were developed as anti-cancer agents [7–14]. The evaluation for cytotoxic activity against different human cancer cell lines shows the promising activities of azanaphthoquinone annelated pyrrole derivatives. Moreover, lead compounds of this derivative show better antiproliferative effects than paclitaxel and doxorubicin on multidrug resistant cell lines [9]. However, the major drawback of these compounds is the easy metabolic cleavage of the oxime group [10]. To overcome this disadvantage of azanaphthoquinone annelated pyrrole derivatives, compounds containing a piperidinyl carbinol instead of the oxime

[a] P. Kamsri, N. Pongprom, A. Srisupan, P. Pungpo
Department of Chemistry, Ubon Ratchathani University
85 Satholimark Rd., Warinchamrap, Ubonratchathani, 34190, Thailand
phone/fax: +66 45 353400 4124/+66 45 288379
*e-mail: pompan_ubu@yahoo.com

[b] A. Punkvang
Division of Science, Faculty of Liberal Arts and Sciences, Nakhon Phanom University
Nakhon Phanom 48000, Thailand

[c] P. Saparpakorn, S. Hannongbua
Department of Chemistry, Kasetsart University
Chatuchak, Bangkok, 10900, Thailand

[d] P. Wolschann
Department of Pharmaceutical Technology and Biopharmaceutics, Faculty of Life Sciences, University of Vienna
Althanstrasse 14, Vienna, Austria

[e] P. Wolschann
Institute for Theoretical Chemistry, University of Vienna
Waehringerstrasse 17, Vienna, Austria

group were developed [10]. The replacement of the oxime group with a piperidinyl carbinol could increase the stability of compounds from the metabolic cleavage. However, the series of synthesized compounds containing a piperidinyl carbinol display only moderate activity [10]. Therefore, with an attempt to improve the cytotoxic activity against cancer cell lines of azanaphthoquinone annelated pyrrole derivatives, a ligand-based design method using 3D QSAR approach was employed in this work. The obtained results should provide the powerful guideline for designing novel and highly effective anti-cancer agents in class of azanaphthoquinone annelated pyrrole derivatives. Moreover, molecular dynamics (MD) simulations using AMBER12 program [15] were performed to model the binding of azanaphthoquinone annelated pyrrole derivatives in the intercalation site of the DNA duplex fragment, $d(CGTACG)_2$.

2 Methodology

2.1 Data Sets and Biological Activities

Twenty-eight azanaphthoquinone annelated pyrrole derivatives taken from one laboratory were used in this work to ensure that all experimental data were determined under consistent assay conditions [9–14]. The general frame of these compounds is shown in Figure 1. Chemical structures and antiproliferative activities on the human cancer cell

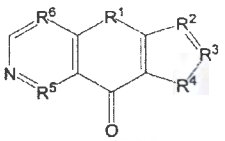


Figure 1. General frame of azanaphthoquinone annelated pyrrole derivatives.

line of cervical carcinoma (KB/HeLa) of azanaphthoquinone annelated pyrrole derivatives are listed in Table 1. All chemical structures of these compounds were constructed using the standard tools available in the program GaussView 3.07 [16] and were then fully optimized using the HF/3-21G method implemented in the Gaussian 03 program [17]. For QSAR study, 28 azanaphthoquinone annelated pyrrole derivatives were divided into a training set of 24 compounds and a test set of 4 compounds for final model development and model validation, respectively. The representatives of the test set were manually selected and are covering the utmost range of activity and structural diversity of compounds in the data set.

Table 1. Chemical structures and antiproliferative activities on cervical carcinoma of azanaphthoquinone annelated pyrrole derivatives.

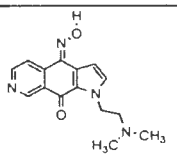
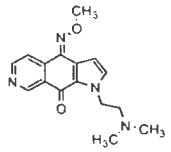
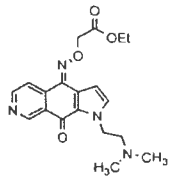
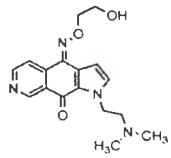
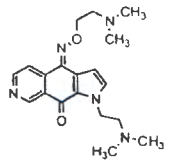
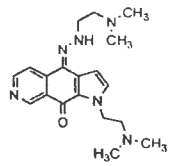
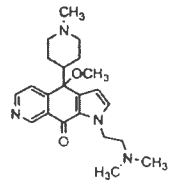
Compound	Structure	log ($1/IC_{50}$)
1		4.92
2		5.92
3		4.92
4		6.18
5		6.17
6		6.55
7[a]		4.45

Table 1. (Continued)

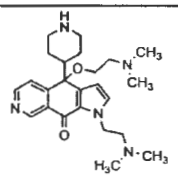
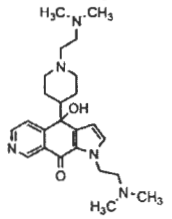
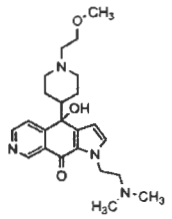
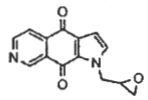
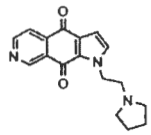
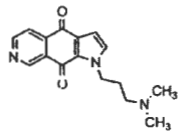
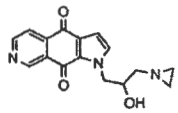
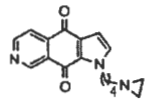
Compound	Structure	log (1/IC_{50})
8		5.05
9		4.75
10		4.67
11		6.61
12		8.10
13		6.85
14		5.60
15		5.73

Table 1. (Continued)

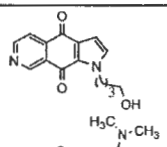
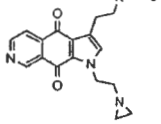
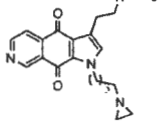
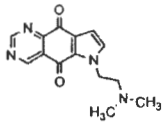
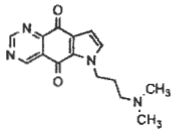
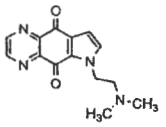
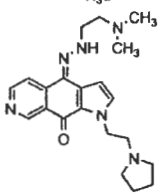
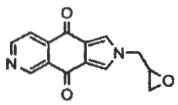
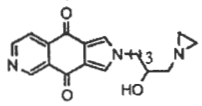
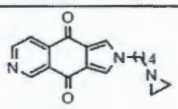
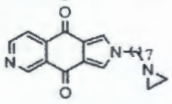
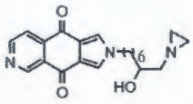
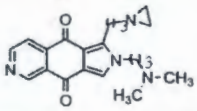
Compound	Structure	log (1/IC_{50})
16		5.13
17[a]		5.54
18[a]		5.85
19		6.19
20[a]		6.01
21		5.30
22		4.58
23		5.49
24		5.66

Table 1. (Continued)

Compound	Structure	log (1/IC_{50})
25		5.55
26		5.68
27		5.55
28		5.56

[a] Test set compound.

2.2 3D QSAR Technique

CoMSIA (Comparative Molecular Similarity Indices Analysis) [18] was used to elucidate the relationship between the structures and the activities of azanaphthoquinone annelated pyrrole derivatives. Molecular modeling software of SYBYL 8.0 [19] was used to calculate CoMSIA models. The molecular alignment for the set up of appropriate CoMSIA models was carried out by the SYBYL pharmacophore alignment module GALAHAD (Genetic Algorithm with Linear Assignment for Hypermolecular Alignment of Datasets) [20]. Five CoMSIA similarity index descriptors including steric, electrostatic, hydrophobic, hydrogen donor and hydrogen acceptor fields were derived with the grid spacing of 2 Å. There are no energy cutoffs for CoMSIA calculations. To generate a contour map with prominent molecular features in the CoMSIA study, an attenuation factor of 0.3 was used.

To derive a linear relationship between molecular descriptors and activities, the partial least square (PLS) approach was employed, in which CoMSIA descriptors were set as independent variables and log (1/IC_{50}) values were used as dependent variables. The cross-validation was performed using the leave-one-out method with a 2.0 kcal/mol column filter to minimize the influence of noisy columns. A final non cross-validated analysis with the optimal number of components was sequentially performed and was then employed to analyze the results. The r^2 and q^2 values were used to evaluate the predictive ability of CoMSIA models.

2.3 MD Simulations

To obtain accuracy and reliability of the binding mode information, five compounds covering the range of the most active to the least active compounds in the series studied were selected for MD simulations. Compound **12** is represented as the most active compound, whereas compounds **7** and **22** are representative compounds possessing weak inhibitory activities in the dataset. Moreover, compounds **11** and **15** possessing moderate activities were also selected. The X-ray crystal structure with the pdb code of 2GB9 was used as the initial coordinates of $d(CGTACG)_2$. The initial coordinates of the selected compounds complexed with $d(CGTACG)_2$ were taken from molecular docking calculations using GOLD Program [21]. AMBER12 using the Amber99 force field [22] for DNA duplex and the general Amber force field (GAFF) parameters [23] for the selected compounds was employed for MD simulations. Each complex structure was solvated by TIP3 waters [24] in an octahedral box extending up to 10 Å from each solute species, $d(CGTACG)_2$ and the selected compounds. 12 Na^+ cations were added to neutralize the charge of each system. The added water molecules and ions in the solvated systems were relaxed using the Sander program to relieve bad steric interactions. Non-bonded cutoff was set at 10 Å. The force of 500.0 kcal/mol was used to restrain the atom positions of all solute species. Thereafter, the whole system was minimized without restraint condition. The systems were then gradually warmed up from 0 K to 300 K in the first 20 ps followed by maintaining the temperature at 300 K in the last 10 ps with 2 fs time steps in a constant volume boundary. The solute species in the solvated systems were restrained to their initial coordinates with a weak force constant of 10 kcal/mol $Å^2$ during the temperature warming. Afterward, the position-restrained dynamics simulation using 2 fs time steps through 70 ps at 300 K under the isobaric condition was performed for each system to relax the positions of the solvent molecules. In this dynamics run, the positions of solute species were restrained with a weak force constant of 10 kcal/mol $Å^2$ during the position-restrained dynamics simulations. Finally, 20 ns MD simulations without the position restraints were performed under the same conditions. During the dynamics simulations, a non-bonded cutoff distance of 8 Å was applied to handle electrostatic interactions in periodic boxes by the Particle Mesh Ewald method [25]. The SHAKE method [26] was applied to constrain the bond lengths of hydrogen atoms attached to heteroatoms. Coordinates and energy outputs during molecular dynamics simulation were printed every 2 ps. MD trajectories were evaluated in terms of the root mean square deviation (*RMSD*), complex structure and binding free energy.

2.4 Binding Energy Calculation

The binding free energies of all selected compounds bound to d(CGTACG)$_2$ were calculated using the Molecular Mechanics/Poisson–Boltzmann Surface Area method (MM-PBSA method) [27] implemented in AMBER 12 package. In this work, 500 snapshots evenly from the last 10 ns on the MD trajectory with an interval of 20 ps were used in the MM-PBSA calculations. The binding free energies (ΔG_{bind}) were obtained as shown in Equations 1 and 2.

$$\Delta G_{bind} = G_{com} - (G_{rec} + G_{ligand}) \quad (1)$$

$$\Delta G_{bind} = \Delta H - T\Delta S \quad (2)$$

where G_{com}, G_{rec} and G_{ligand} are the free energies of the complex, DNA and the ligand, respectively. In general, the binding free energy composes of an enthalpic (ΔH) and an entropic contribution ($-T\Delta S$). The enthalpic contribution (ΔH) contains the gas-phase molecular mechanics energy (ΔG_{MM}) calculated with a sander module and the salvation free energy (ΔG_{solv}) calculated with the PBSA program of the AMBER suite. ΔG_{MM} is divided into noncovalent van der Waals component (ΔG_{vdW}), electrostatic energies component (ΔG_{ele}) and bond, angle, dihedral energies (ΔG_{INT}). The entropy contribution ($-T\Delta S$) to the binding free energy was estimated using normal-mode analysis with AMBER NMODE module. To save the computational time, 100 snapshots evenly from the last 10 ns on the MD trajectory with an interval of 100 ps were used to estimate the contribution of the entropy.

3 Results

3.1 CoMSIA Models

The training set chosen contains 24 compounds with activity range from 4.58 to 8.10 in log(1/IC_{50}) units. Only one compound, compound 12, has a high activity value. If it is removed from the dataset, the activity range narrows. It is important to note that compound 12 shows great influence on the CoMSIA model. Removing this compound from the train set results in statistically unsatisfied model (unpublished data). Therefore, compound 12 was included as the training set in this study. Table 2 lists the statistical parameters of CoMSIA models obtained from the PLS analysis. With the highest q^2 of 0.65, the CoMSIA model including steric, electrostatic, hydrophobic and hydrogen acceptor fields was selected as the best CoMSIA model. The contribution of steric, electrostatic, hydrophobic and hydrogen acceptor fields is 12%, 18%, 42% and 28%, respectively, indicating that the hydrophobic field shows greatest influence on the activity of azanaphthoquinone annelated pyrrole derivatives. The selected CoMSIA model has high power to estimate the activities of training set with r^2 of 0.99 and q^2 of 0.65. In order to assess the predictive ability of this CoMSIA model, the activities of the test set compounds were predicted. Experimental and predicted activities (log(1/IC_{50}) for the training set and test set are reported in Table 3, while distribution of experimental and predicted values for the training and the test sets according to the best CoMSIA model is represented in Figure 2. The calculated data of compounds in training set fit well with experimental results with error less than 0.1 for all compounds and the prediction error for all tested compounds are less than 1.0 Therefore, the best CoMSIA model could be utilized to predict the activities for new designed azanaphthoquinone annelated pyrrole derivatives.

3.2 CoMSIA Contour Maps

Figures 3 and 4 present the CoMSIA contour maps which reveal the influence of steric, electrostatic, hydrophobic and hydrogen acceptor fields to the activity of azanaphthoquinone annelated pyrrole derivatives. Favorable and unfavorable steric regions are represented in green and

Table 2. Statistical results of various CoMSIA models with different combined fields. q^2: leave-one-out cross-validated correlation coefficient, r^2: non-cross-validated correlation coefficient, N: optimum number of components, s: standard error of prediction, SEE: standard error of estimate, F: F-test value; S: steric field, E: electrostatic field, H: hydrophobic field, D: hydrogen donor field, A: hydrogen acceptor field.

CoMSIA model	Statistical parameters						Fraction
	q^2	r^2	N	s	SEE	F	
S/E	0.31	0.89	4	0.73	0.3	38	44/56
S/H	0.44	0.96	4	0.66	0.19	101	21/79
S/A	0.44	0.99	6	0.70	0.08	386	37/63
S/D	0.02	0.38	1	0.81	0.65	14	25/75
S/E/H	0.61	0.99	6	0.58	0.08	394	14/25/61
S/E/A	0.61	0.99	6	0.58	0.04	1819	25/30/45
S/E/D	0.12	0.45	1	0.77	0.61	18	18/27/55
S/E/H/A[a]	0.65	0.99	6	0.55	0.04	1335	12/18/42/28
S/E/H/D	0.42	0.99	6	0.71	0.09	316	12/22/51/16
S/E/H/A/D	0.54	0.99	6	0.63	0.07	510	9/15/33/16/27

[a] The best CoMSIA model.

Table 3. Experimental and predicted antiproliferative activities of compounds in the training set and the test set.

Compound	log (1/IC_{50})		
	Experimental	Predicted[b]	Residue
1	4.92	4.93	−0.01
2	5.92	5.96	−0.04
3	4.92	4.94	−0.02
4	6.18	6.18	0.00
5	6.17	6.13	0.04
6	6.55	6.57	−0.02
7[a]	4.45	5.02	−0.57
8	5.05	5.06	−0.01
9	4.75	4.77	−0.02
10	4.67	4.65	0.02
11	6.61	6.53	0.07
12	8.10	8.11	−0.01
13	6.85	6.86	−0.01
14	5.60	5.63	−0.03
15	5.73	5.81	−0.08
16	5.13	5.17	−0.04
17[a]	5.54	6.30	−0.76
18[a]	5.85	5.06	−0.57
19	6.19	6.17	0.02
20[a]	6.01	5.44	0.57
21	5.30	5.33	−0.03
22	4.58	4.53	0.05
23	5.49	5.52	−0.03
24	5.66	5.65	0.01
25	5.55	5.55	0.00
26	5.68	5.69	−0.01
27	5.55	5.49	0.06
28	5.56	5.52	0.04

[a] Test set. [b] Calculated by CoMSIA model (S/E/H/A).

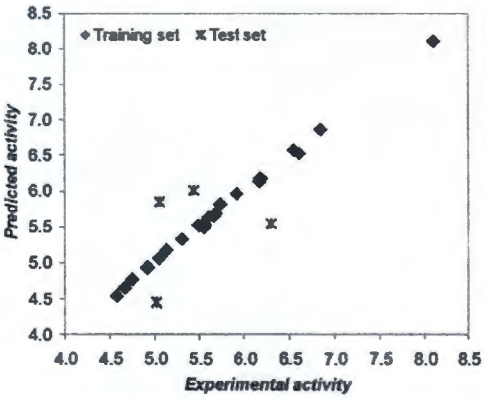


Figure 2. Plot of the experimental and predicted activities of the training and test sets derived from the best CoMSIA model.

yellow contours, respectively, while blue and red contours indicate the regions which favor positive and negative charges, respectively. The magenta and white contours represent the favorable and unfavorable hydrophobic regions, respectively. The cyan and orange contours indicate regions that favor the hydrogen acceptor group and unfavor hydrogen acceptor group, respectively. The interpretation of CoMSIA contour maps reveals the structural requirement of each substituent position in azanaphthoquinone annelated pyrrole scaffold helpful for rational design of novel and potent azanaphthoquinone annelated pyrrole derivatives.

3.3 Structural Requirement for the R^1 Position

Among all selected CoMSIA descriptors, a large yellow, contour overlapping with a blue contour locate near the R^1 position shown in Figure 3. These contours suggest that this position prefers the small substituent which possesses low electron density. This suggestion is supported by all compounds presenting the R^1 substituent as the carbonyl oxygen showing the activities above five logarithmic units. In particular, the R^1 substituent of the most active compound, compound **12**, is the carbonyl oxygen. Therefore, the carbonyl oxygen should be optimal substituent for the R^1 position.

3.4 Structural Requirement for the R^2 and R^3 Positions

The R^3 position is buried in a large unfavorable hydrophobic region, white contour, indicating that the hydrophobic substituent should be not presented at this position (Figure 4). Accordingly, the presence of hydrophobic groups at the R^3 position of compounds **23–28** might be one factor responsible for lower activities of these compounds as compared to that of compound **12**. Additionally, a large favorable steric contour locates near the R^3 position (Figure 3a). Therefore, the introduction of bulky substituents possessing hydrophilic properties onto this position could enhance the activity of azanaphthoquinone annelated pyrrole derivatives. In case of the R^2 position, no structural requirement is suggested from the best CoMSIA. Most of the compounds in the training set contain the same substituent at the R^2 position (the CH moiety). Only compound **28**, bearing (aziridine-1-yl)butyl at the R^2 position, is different from those of other compounds. That means the substituent at this position does not significantly contribute to the binding affinity of the compounds. As exemplified by the comparison of the inhibitory activities of compound **15** (R^2: H, log (1/IC_{50}) = 5.73), and that of compound **18** (R^2: $-CH_2CH_2NMe_2$, log (1/IC_{50}) = 5.85), the bulkier substituent attached to the R^2 position does not confer to the inhibitory activities of both compounds.

3.5 The Structural Requirement for the R^4 Position

As shown in Figure 3, a large green contour corresponds to the location of the group attached to the R^4 substituent. However, the tolerated steric requirement of this region is shown by a yellow contour located on the opposite side of the favorable steric region. It is indicated that steric occu-

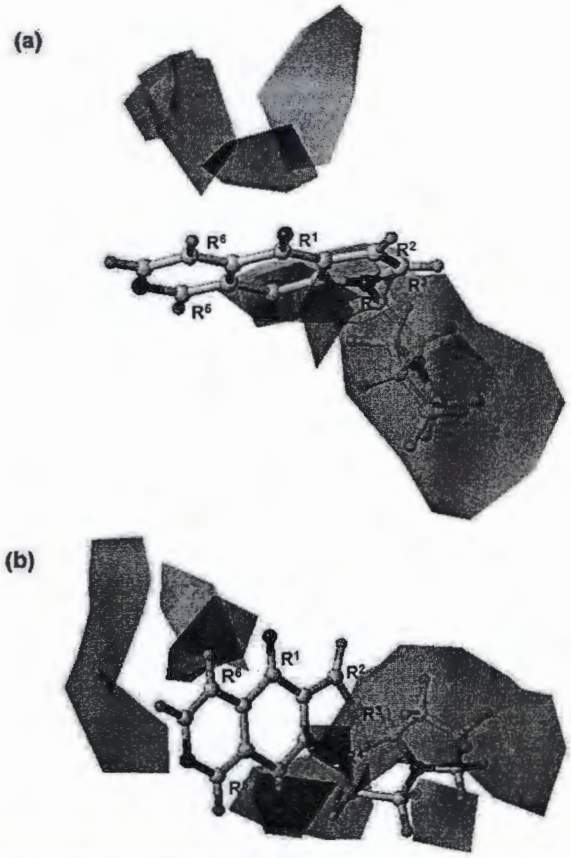


Figure 3. Steric and electrostatic contours obtained from the best CoMSIA model in combination with the most active compound (compound 12), (a) and (b) are the steric and electrostatic contours with the horizontal and vertical plane of azanaphthoquinone annelated pyrrole scaffold, respectively.

pancy with bulky groups would increase the binding affinity, but the size of the substituent should be optimum and not be too large. In addition, a predominant feature of hydrophobic contour, magenta area, in the proximity of the R^4 substituent (Figure 4) predicts favorable hydrophobic substituents. The reliability of the suggestions derived from the CoMSIA contour maps is verified by compound 12, the highest active compound, with optimum bulky group and preferably hydrophobic property of $-NCH_2CH_2$-pyrrolidine substituent attached to the R^4 substituent. Compounds, such as compounds 23–27, occupying the small R^4 substituent i.e. hydrogen atom, lose to fill the bulkier favorable region, showing lower the activities than that of compound 12. On the other hand, compounds 13 and 15, occupying the R^4 substituent with too large substituents, display significantly reduced biological activities compared to compound 12. Besides, lipophilic substituents attached to the similar position of compounds 11, 14 and 16 are the reason why these compounds exhibit weak inhibitory activities compared to that of compound 12.

3.6 Structural Requirement for the R^5 Position

The orange contour is placed near the R^5 position indicating that this position disfavors the hydrogen acceptor substituent (Figure 4). This finding is supported by the lower

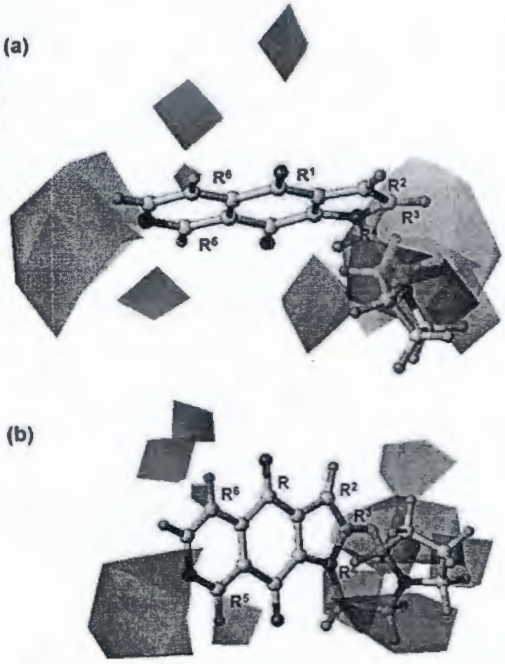


Figure 4. The hydrophobic and hydrogen acceptor contours obtained from the best CoMSIA model in combination with the most active compound (compound 12), (a) and (b) are the hydrophobic and hydrogen acceptor contours with the horizontal and vertical plane of azanaphthoquinone annelated pyrrole scaffold, respectively.

activity of compound **21** bearing a nitrogen atom at the R^5 position, as compared to that of compound **19** containing the CH group.

3.7 Structural Requirement for the R^6 Position

The orange contour locates near the R^6 position indicating that this position disfavors the hydrogen acceptor substituent (Figure 4). This finding is explaining why compound **20** bearing a nitrogen atom at the R^6 position exhibits lower potency than compound **13**.

3.8 MD Simulations

3.8.1 Structural Stability

In order to investigate the structural stability during MD simulations, the *RMSDs* as a function of the simulation time of each complex with respect to the starting structure were analyzed as shown in Figure 5. The *RMSDs* of the complex structures of the selected compounds, **7**, **11**, **12**, **15** and **22** bound to $d(CGTACG)_2$ reach the plateau characteristic at 8 ns, 2 ns, 8 ns, 4 ns and 6 ns, respectively. These results in-

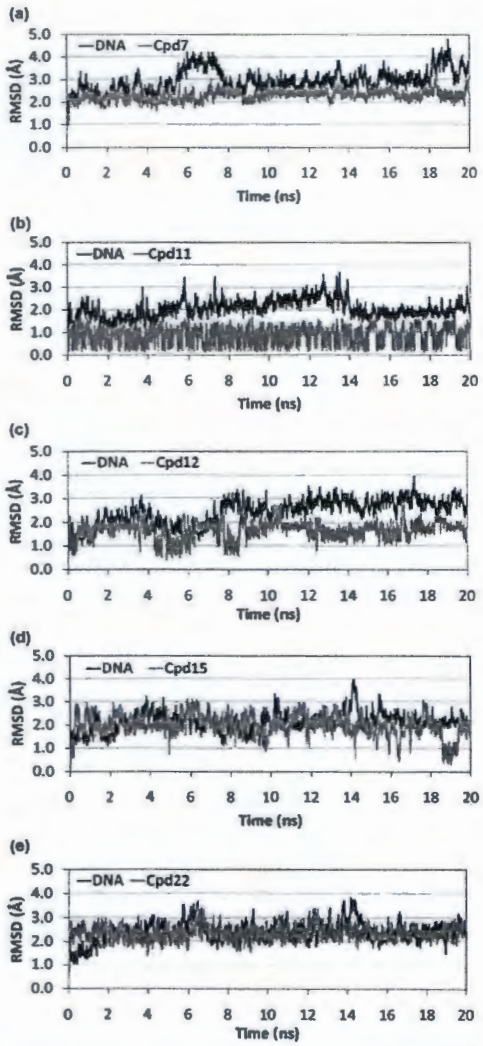


Figure 5. *RMSD* plots for the compound 7(a), 11(b), 12(c), 15(d) and 22(e)/$d(CGTACG)_2$.

dicate that each complex structure reaches an equilibrium state after that time. Therefore, the information in terms of energy and structure of each system were analyzed over an equilibrium state.

3.8.2 The Binding Free Energy

To gain quantitative insights into the affinity for binding of the azanaphthoquinone derivatives in the intercalation binding site, the binding free energies of the selected compounds were calculated by the Molecular Mechanics/Poisson–Boltzmann Surface Area (MM-PBSA) method [27]. The comparison between the experimental binding free energies (ΔG_{exp}) and the calculated binding free energies (ΔG_{cal}) of compounds **7**, **11**, **12**, **15** and **22** is shown in Table 4. The correlation of experimental IC_{50} and calculated free binding free energy is presented in Figure 6. It is notable that the calculated binding free energies of all selected compounds are in the correct order as compared with the IC_{50} values. The obtained results could be successfully used to validate the MD procedure in this study. This result shows the reliability of the MD simulations.

3.8.3 Structural Analysis

The binding mode analysis of the selected azanaphthoquinone annelated pyrrole derivatives/ $d(CGTACG)_2$ complexes started with compound **12** (IC_{50}=0.008 μM), which is the most active compound of the series studies. Compound **12** inserts at the C1G2(A)/C5G6(B) of $d(CGTACG)_2$ as shown in Figure 7. The horizontal plane of azanaphthoquinone annelated pyrrole scaffold is bound perpendicular to the horizontal plane of the CG base pairs. The interactions of azanaphthoquinone annelated pyrrole scaffold of this compound and the CG base pairs were observed by π–π stacking interactions between quinone and pyrrole moieties with the purine and pyrimidine ring of C1G2(A) and C5G6(B).Compound **12** display extensively hydrogen bond contacts between: (i) the oxygen carbonyl of quinone ring at the R^1 substituent with CH group of C1(A) and G2(A) deoxyribose (ii) the CH group at the alkyl group of the R^4 substituent and NH_2 group of C1(A). Notably, compound **12** is engaged in an additional hydrogen-π interaction between CH of the pyrrolidini-1-yl-ethyl linker and pyrimidine ring of the C5(B). Moreover, hydrophobic interactions of pyrrolidine ring at the R^4 substituent with C5(B) and G6(B) of the DNA major groove were observed. The key structural features derived are in consistence with the CoMSIA interpretaion. Numerous crucial interactions observed for compound **12** should be accounted for displaying the excellent binding free energy (−11.1 kcal/mol).

Compounds **11** and **15** (IC_{50}=2.512 and 1.862 μM, respectively) are represented as moderate active compounds. Figure 8a shows the binding interactions of compound **11** bound to C1G2(A)/C5G6(B) of $d(CGTACG)_2$. Compound **11** is parallel to the long axes of the CG base pairs duplex to form the π–π stacking interaction with pyrimidine and purine rings of cytosine and guanine, respectively. The R^4 substituent protrude into the major groove of the DNA duplex with C1G2(A). Compound **11** could form hydrogen

Table 4. The ΔG_{exp} and ΔG_{cal} of the selected azanaphthoquinone derivatives bound to $d(CGTACG)_2$.

Cpd.	IC_{50} (μM)	ΔH	$-T\Delta S$	ΔG_{cal}	ΔG_{exp} [a]
7	35.481	−20.0	16.5	−3.5	−6.1
22	26.282	−25.8	20.4	−5.4	−6.3
15	1.862	−22.3	15.9	−6.4	−7.8
11	0.245	−22.9	15.5	−7.4	−9.0
12	0.008	−25.0	16.3	−8.7	−11.1

[a] derived from $\Delta G = RT \ln[\text{activity}]$, where activity is the antiproliferative activity of compounds **7**, **22**, **15**, **11** and **12** on cervical carcinoma expressed in IC_{50} [13]. R represents the gas constant (1.988 cal/mol K), T represents the temperature (300 K).

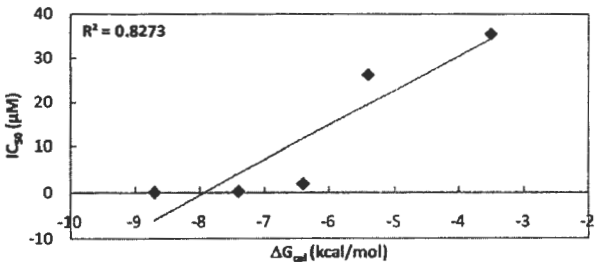


Figure 6. Correlation of experimental IC_{50} and calculated free binding free energy using MM-PBSA method

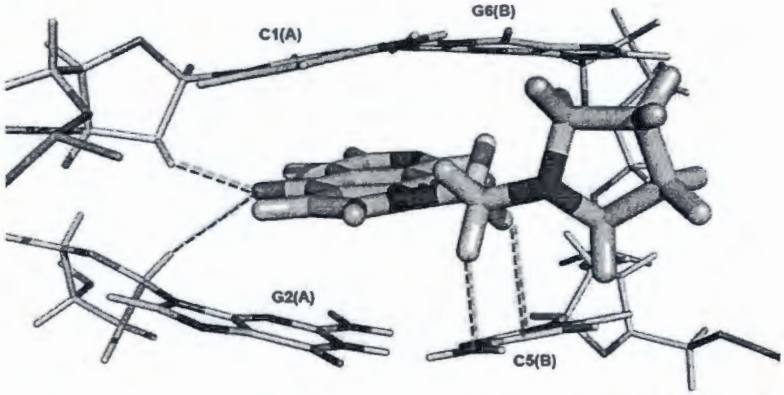


Figure 7. The structure of compound 12/d(CGTACG)$_2$ complex averaged over the last 5 ns of the simulation time. For clarity, only the structure in the intercalating part, C1G2(A)/C5G6(B) is shown. Carbon atoms of DNA are colored by gray. Carbon atoms of compound 12 are colored by yellow. H-bond are colored in red. Hydrogen–π is colored in green.

bond interactions between: (i) the CH group of methyl linker at the R^4 substituent with oxygen carbonyl of C1(A) (ii) the oxygen ether group of oxirane ring at the R^4 substituent with pyrimidine ring of C1(A) (iii) the CH group at chiral carbon of oxirane ring at the R^4 substituent and nitrogen atom of G2(A). However, any hydrogen-π interaction and hydrophobic interactions concerning the R^4 substituent with the base pairs of the DNA duplex are missing. This is the appropriate explanation why compound 11 exhibited lower inhibitory activity compared to compound 12.

The binding interactions of compound 15 bound to of d(CGTACG)$_2$ is presented in Figure 8b. The azanaphthoquinone annelated pyrrole scaffold of compound 15 interact with C1C2(A)/C5G6(B) in the intercalation binding site of d(CGTACG)$_2$ are similar to that of compound 12. The azanaphthoquinone annelated pyrrole axes of these compounds are perpendicular to the long axes of the CG base pairs Therefore, the scaffold of compound 15 inserts in the CG steps of the DNA duplex to form the π–π stacking interactions between quinone and pyrrole moieties with the purine and pyrimidine ring. Only two hydrogen bond interactions between the CH of butyl linker at the R^4 substituent of compound 15 and the nitrogen atom of purine base G6(B) were observed. Similar to compound 11, any hydrogen–π interaction and hydrophobic interactions concerning the R^4 substituent with the base pairs of the DNA duplex were observed. Moreover, the conformation of the DNA helix strand of compound 15/d(CGTACG)$_2$ complex was significantly changed. The number and the quality of interactions between the R^4 substituent and base pairs of the DNA duplex decreased could explain their moderate to weak inhibitory activities of compounds 11 and 15 compared to that of compound 12.

Compounds 22 and 7 (IC_{50} = 26.282 μM and 35.481 μM, respectively), are representative compounds possessing weak inhibitory activities in this analogues. Interestingly, the chemical structure of compound 22 is highly similar to that of compound 12, the most active compound, except the the steric hindrance of the R^1 substituent. To investigate the effect of the R^1 substituent on the inhibitory activities of the compounds, the binding mode of compound 22/ d(CGTACG)$_2$ was compared to that of compound 12. In contrary, the horizontal plane of azanaphthoquinone annelated pyrrole scaffold of compound 22 is parallel only to the horizontal plane of the C5(B)G2(A) base pair (Figure 9a).

The CH_2CH_2-pyrrolidine side chain at the R^4 position of this compound protrudes into the minor groove of the DNA duplex, whereas bulky substituent, $-NNHCH_2CH_2NMe_2$ group, at the R^1 position protrudes into the major groove. Because of the insertion of compound 22, bearing too bulky substituent at the R^1 substituent, the hydrogen bonds between the C1(A) base and the G6(B) base were broken leading to open up of the C1(A)G6(B) base pair of the DNA duplex. The broken base pair results in the loss of the π–π interactions between the azanaphthoquinone annelated pyrrole scaffold and the CG base pairs of compound 22. In case of compound 7, the least active compound in the dataset which containing bulky substituent at the R^1 substituent, the binding mode of compound 7/ d(CGTACG)$_2$ complex is similar to that of compound 22 (Figure 9b). The C1(A)G6(B) base pair is opened up because of the insertion of compound 7. This finding explains why compounds 22 and 7 display poor binding free energy (−6.3 kcal/mol and −3.5 kcal/mol, respectively), as compared to that of compound 12 (−11.1 kcal/mol). Therefore, these obtained results imply that the presence of bulky

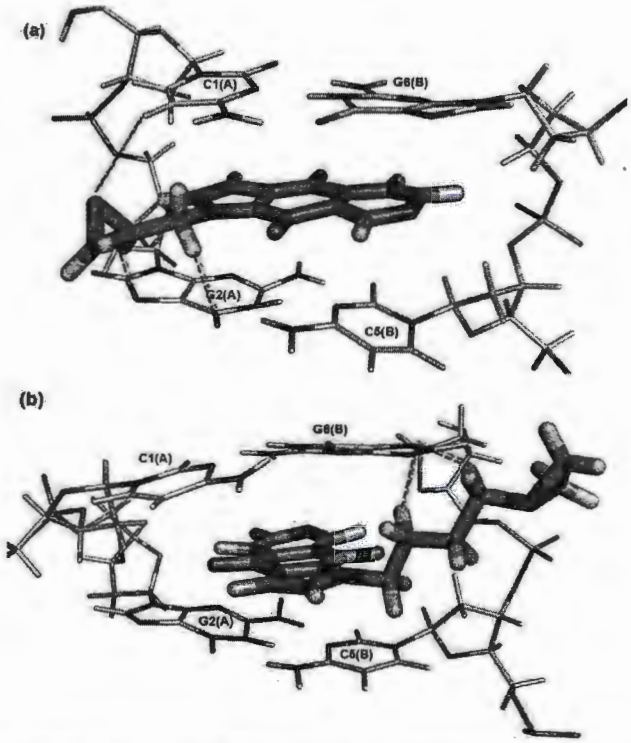


Figure 8. The structure of compound 11/d(CGTACG)$_2$ complex (a) and compound 15/d(CGTACG)$_2$ complex (b) averaged over last 5 ns of the simulation time. For clarity, only the structure in the intercalating part, C1G2(A)/C5G6(B) is shown. Carbon atoms of DNA are colored by gray. Carbon atoms of compound 11 are colored by pink. Carbon atoms of compound 15 are colored by orange. H-bond are colored in red.

substituents at both R^1 and R^4 positions such as compounds 7 and 22 diminishes the binding affinity of azanaphthoquinone annelated pyrrole derivatives in the CG step of DNA duplex. There analyses are in well consistence with the CoMSIA suggestion that the small substituent is preferred for the R^1 position.

3.8.4 A Comparison Between the CoMSIA Model and MD Analyses

In order to verify the correspondence of the structural requirements derived from the 3D-QSAR model with the MD analysis, the CoMSIA contour maps were superimposed to the equilibrium MD conformation of compound 12 shown in Figure 10. The CoMSIA and MD analyses clearly indicate the similar suggestions for the importance of the R^4 substituents to enhance the inhibitory activities of compounds in the dataset. As previously discussed, the orientation of bulky and hydrophobically favored substituents of the R^4 position bound into the base pairs of the DNA duplex is one of the key characteristics of the compound. The steric contour map highlights the importance of bulky substituent of the pyrrolidini-1-yl-ethyl group at the R^4 position which could be involved in hydrogen bond interaction and hydrogen–π interaction with the key DNA base pairs, shown in Figure 10a. As shown in Figure 10b, the hydrophobic map points out the beneficial presence of the hydrophobic substituents at the R^4 position, enhancing hydrophobic contacts with C5(B) and G6(B) of the DNA major groove. Moreover, an unfavorable steric contour located in the vicinity of the R^1 position proves to match with the binding site topology by representing small substituent i.e. the carbonyl group as shown in compound 12. Satisfactory agreement obtained from the CoMSIA model and MD anal-

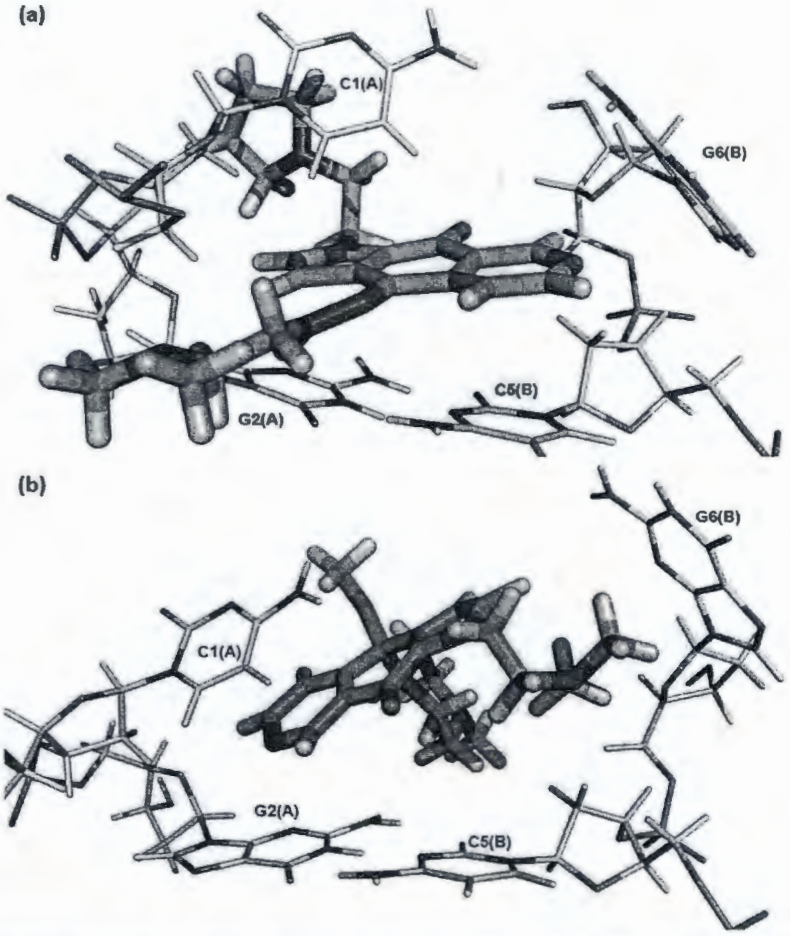


Figure 9. The structure of compound 22/d(CGTACG)$_2$ complex (a) and compound 7/d(CGTACG)$_2$ complex (b) averaged over last 5 ns of the simulation time. For clarity, only the structure in the intercalating part C1G2(A)/C5G6(B) is shown. Carbon atoms of DNA are colored by gray. Carbon atoms of compound 22 are colored by cyan. Carbon atoms of compound 7 are colored by green.

yses may provide insight into crucial structural features effecting ligand receptor interactions and their binding affinities and thus can provide guideline for novel inhibitor design of azanaphthoquinone annelated pyrrole derivatives possessing better antiproliferative activity.

4 Conclusions

The CoMSIA model based on pharmacophore alignment has high power to predict the activities of azanaphthoquinone annelated pyrrole derivatives with r^2 of 0.99 and q^2 of 0.65. By the interpretation of CoMSIA contour maps, the key structural elements required for a better antiprolifera-

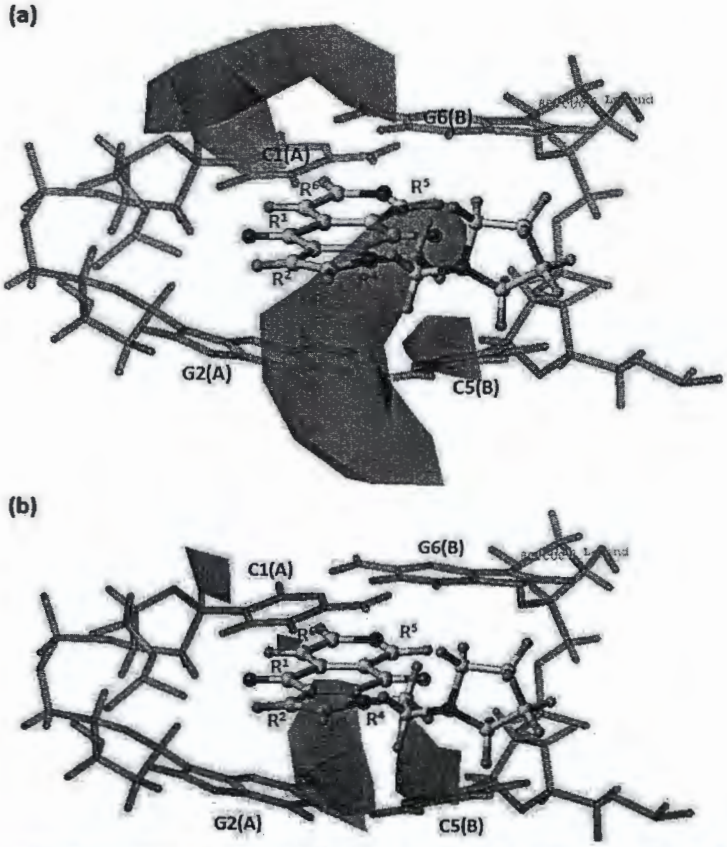


Figure 10. Superimposition of the average structure from the last 5 ns of the MD simulation (bond and stick) and the CoMSIA steric (a), and hydrophobic contour (b). For clarity, only the structure in the intercalating part C1G2(A)/C5G6(B) (stick) is shown. Carbon atoms of DNA are colored by light blue. Carbon atoms of compound 12 are colored by atom type. Sterically favored steric areas are represented by green and disfavored steric areas are represented by yellow. Hydrophobically favored areas are represented by magenta and disfavored hydrophobic areas are represented by white.

tive activity of azanaphthoquinone annelated pyrrole derivatives were clearly elucidated for six substituent positions on azanaphthoquinone annelated pyrrole scaffold. Moreover, MD simulations using AMBER12 program were successful to model the reliable binding modes of azanaphthoquinone annelated pyrrole derivatives in the CG step of $d(CGTACG)_2$. MD trajectory analysis in terms of complex structure and binding free energy provides the insight into the crucial ligand-DNA interaction and the key structural feature favorable for binding affinity of azanaphthoquinone annelated pyrrole derivatives in the intercalation site of DNA duplex. Furthermore, the finding obtained from MD simulations is supported to the CoMSIA guideline for designing new compounds with the improved activity. Therefore, the combined results obtained from this study should facilitate the further modification of azanaphthoquinone annelated pyrrole scaffold toward to generate novel DNA intercalating agents as anti-cancer compounds in class of azanaphthoquinone annelated pyrrole derivatives with improved inhibition potency, target selectivity and specificity.

Acknowledgements

The financial support from Royal Golden Jubilee Ph.D. Program (PHD/004/2554) to P. Kamsri is gratefully acknowledged. This research was supported by the *National Research Council of Thailand* and the Thailand Research Fund (DBG5380006, RTA53800010). Faculty of science, *Ubon Ratchathani University, University of Vienna* and *ASEA-Uninet* are gratefully acknowledged for supporting this research.

References

[1] C. J. Dunn, K. L. Goa, *Drugs Aging* **1996**, *9*, 122–147.
[2] L. R. Wiseman, C. M. Spencer, *Drugs Aging* **1997**, *10*, 473–485.
[3] J. Mazerski, S. Martelli, E. Borowski, *Acta Biochim. Pol.* **1998**, *45*, 1–11.
[4] W. H. Frishman, H. M. Sung, H. C. Yee, L. L. Liu, D. Keefe, A. I. Einzig, J. Dutcher, *Curr. Probl. Cancer* **1997**, *21*, 301–360.
[5] A. Engert, R. Herbrecht, A. Santoro, P. L. Zinzani, I. Gorbatchevsky, *Clin. Lymphoma Myeloma* **2006**, *7*, 152–154.
[6] E. Cavalletti, L. Crippa, P. Mainardi, N. Oggioni, R. Cavagnoli, O. Bellini, F. Sala, *Invest. New Drugs* **2007**, *25*, 187–195.
[7] H. Spreitzer, A. Pichler, W. Holzer, M. Kratzel, R. Slanz, A. Koulouri, P. Krenn, U. Parrer, P. Szieberer, *Heterocycles* **2001**, *54*, 111–121.
[8] H. Spreitzer, C. Puschmann, *Chem. Mon.* **2007**, *138*, 517–522.
[9] K. Shanab, N. Pongprom, E. Wulz, W. Holzer, H. Spreitzer, P. Schmidt, B. Aicher, G. Muller, E. Gunther, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 6091–6095.
[10] N. Pongprom, G. Muller, P. Schmidt, W. Holzer, H. Spreitzer, *Monatsh. Chem.* **2009**, *140*, 309–313.
[11] K. Shanab, E. Schirmer, H. Knafl, E. Wulz, W. Holzer, H. Spreitzer, P. Schmidt, B. Aicher, G. Muller, E. Gunther, *Bioorg. Med. Chem. Lett.* **2010**, *20*, 3950–3952.
[12] K. Shanab, E. Schirmer, E. Wulz, B. Weissenbacher, S. Lassnig, R. Slanz, G. Fosleitner, W. Holzer, H. Spreitzer, P. Schmidt, B. Aicher, G. Muller, E. Gunther, *Bioorg. Med. Chem. Lett.* **2011**, *21*, 3117–3121.
[13] N. Pongprom, P. Pungpo, *KKU Res. J.* **2012**, *17*, 195–202.
[14] N. Pongprom, H. Bachitsch, A. Bauchinger, H. Ettefagh, T. Haider, M. Hofer, H. Knafl, R. Slanz, M. Waismeyer, F. Wieser, H. Spreitzer, *Monatsh. Chem.* **2010**, *141*, 53–62.
[15] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, P. A. Kollman, *AMBER12*, University of California, San Francisco, **2008**.
[16] *GaussView 03*, Revision 3.07, Gaussian, Inc., Wallingford, **2006**.
[17] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B.Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M.W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, *Gaussian 03*, Gaussian Inc., Wallingford, **2004**.
[18] G. Klebe, U. Abraham, T. Mietzner, *J. Med. Chem.* **1994**, *37*, 4130–4146.
[19] *SYBYL 8.0*, Tripos, Inc., **2007**.
[20] *GALAHAD*, Tripos, St. Louis, MO, **2007**.
[21] a) G. Jones, P. Willett, R. C. Glen, *J. Mol. Biol.* **1995**, *245*, 43–53; b) G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, *J. Mol. Biol.* **1997**, *267*, 727–748; c) J. C. Cole, J. W. M. Nissink, R. Taylor, in *Virtual Screening in Drug Discovery* (Eds: B. Shoichet, J. Alvarez), Taylor & Francis, CRC Press, Boca Raton, Florida, USA **2005**; d) M. L. Verdonk, G. Chessari, J. C. Cole, M. J. Hartshorn, C. W. Murray, J. W.M. Nissink, R. D. Taylor, R. Taylor, *J. Med. Chem.* **2005**, *48*, 6504–6515; e) N. M. O'Boyle, S. C. Brewerton, R. Taylor, *J. Chem. Inf. Model.* **2008**, *48*, 1269–1278; f) O. Korb, T. Stützle, T. E. Exner, *J. Chem. Inf. Model*, **2009**, *49*, 84–96; g) S. C. Brewerton, *Curr. Opin. Drug Discov. Devel.* **2008**, *11*, 356–364.
[22] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, *J. Comp. Chem.* **2003**, *24*, 1999–2012.
[23] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174.
[24] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–936.
[25] T. Darden, D. York, L. J. Pedersen, *Chem. Phys.* **1993**, *98*, 10089–10092.
[26] J. P. Ryckaert, G. Ciccotti, H. J.C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.
[27] J. Srinivasan, T. E. Cheatham III, P. Cieplak, P. A. Kollman, D. A. Case, *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409.

Received: October 30, 2012
Accepted: April 23, 2013
Published online: June 11, 2013

CONFERENCE PROCEEDINGS

1. **P. Kamsri**, A. Punkvang, D. Roba3, W. Sippl, P. Saparpakorn, S. Hannongbua and P. Pungpo, Computer-based Inhibitor Design for M. tuberculosis PknG: Integrations of MD simulations and 3D-QSAR Study, Pure and Applied Chemistry International Conference (PACCON2016), 2016, Accepted for publication.

2. **P. Kamsri**, P. Meewong, A. Punkvang, P. Saparpakorn, S. Hannongbua, W. Zhu, Z. Chen and P. Pungpo, Identification of Thai natural products as anti-tuberculosis agents through structure based virtual screening, The 19th International Annual Symposium on Computational Science and Engineering (ANSCSE19), 2015, 219-224.

3. **P. Kamsri**, A. Punkvang, P. Saparpakorn, S. Hannongbua, S. Irle and P. Pungpo, Key structural and binding free energies of heterocyclic B-ring diphenyl ether derivatives as newly potent InhA inhibitors, Pure and Applied Chemistry International Conference (PACCON2015), 2015, 335-338.

4. **P. Kamsri**, A. Srisupan, P. Meewong, A. Punkvang, P. Saparpakorn, S. Hannongbua, P. Wolschann, U. Leartsakulpanich, S. Prueksaaroon and P. Pungpo, Discovery of novel InhA inhibitors as anti-tuberculosis agents using ligand and structure based virtual screening, Pure and Applied Chemistry International Conference (PACCON2014), 2014, 200-203.

5. **P. Kamsri**, N. Pongprom, A. Punkvang, P. Saparpakorn, S. Hannongbua, S. Prueksaaroon and P. Pungpo, Molecular dynamics simulations of the single substituent of azanaphthoquinone annelated pyrrole derivatives as anti-cancer agents in DNA duplex, The 18th International Annual Symposium on Computational Science and Engineering (ANSCSE18), 2014, 16-22.

6. **P. Kamsri**, A. Srisupan, A. Punkvang, P. Saparpakorn, S. Hannongbua, P. Wolchann, S. Prueksaaroon and P. Pungpo, Insight into the binding mode of the potential bi-substrate InhA inhibitors as anti-tuberculosis agents: Molecular dynamics simulations, Pure and Applied Chemistry International Conference (PACCON2013), 2013, 936-939.

7. **P. Kamsri**, A. Punkvank, N. Pongprom, A. Srisupan, P. Saparpakorn, S. Hannongbua, P. Wolchann, H. Spreitzer, S. Prueksaaroon and P.Pungpo, Insight into the key structural feature of anti-cancer agents in class of azanaphthoquinone

annelated pyrrole derivatives as based on QSAR approaches, The 17th International Annual Symposium on Computational Science and Engineering (ANSCSE17), 2013, 32-40.

8. **P. Kamsri**, A. Punkvang, A. Srisupan, P. Saparpakorn, S. Hannongbua, P. Wolchann, S. Prueksaaroon, N. Pongprom and P. Pungpo, Molecular modeling of azanaphthoquinone annelated pyrrolo hydrazone as anticancer agent in DNA duplex using molecular dynamics simulations, The 38th Congress on Science and Technology of Thailand, 2012, 1-4.

9. **P. Kamsri**, A. Punkvang, A. Srisupan, P. Saparpakorn, S. Hannongbua, P. Wolchann, S. Prueksaaroon and P. Pungpo, Insight into the binding modes of bi-substrate inhibitors of anti-tuberculosis agents as enoyl-ACP reductase through molecular dynamics simulations, The 7th International Symposium of The Protein Society of Thailand, 2012, 148-153.

10. **P. Kamsri**, A. Punkvang, D. Kasemsri, A. Srisupan, K. Kun-asa, P. Saparpakorn, Supa Hannongbua, P. Wolschann, S. Prueksaaroon and P. Pungpo, Computer aided molecular design of novel diarylpyrimidines with cyano and hydroxymethy linker as HIV-1 NNRTIs using QSAR studies, The 16th International Annual Symposium on Computational Science and Engineering (ANSCSE16), 2012, 115-122.

11. **P. Kamsri**, A. Punkvang, K. Kun-asa, D. Kasemsee, P. Saparpakorn, S. Hannongbua, P. Wolschann, S. Prueksaaroon and P. Pungpo, The binding mode investigation of potential bi-substrate InhA inhibitors as anti-tuberculosis agents using molecular modeling, Pure and Applied Chemistry International Conference (PACCON2012), 2012, 987-989.

12. **P. Kamsri**, A. Punkvang, K. Kun-asa, P. Saparpakorn, S. Hannongbua, P. Wolschann, S. Prueksaaroon and P. Pungpo, Insight into the binding interactions of anti-mycobacterial agents in class of B-ring modified alkyl diphenyl ether based on molecular modeling, The 37th Congress on Science and Technology of Thailand, 2011, 1-5.

13. **P. Kamsri**, A. Punkvang, K. Kun-asa, A. Kumkong, P. Saparpakorn, S. Hannongbua, P. Wolschann, S. Prueksaaroon and P. Pungpo, Molecular modeling and quantum chemical calculations of high potency anti-tuberculosis agents in class of

triclosan and diphenyl ether derivatives as InhA enzyme of *Mycobacterium tuberculosis,* 14th Asian Chemical Congress 2011 (14 ACC), 2011, 326-332.

14. **P. Kamsri**, A. Punkvang, K. Kun-asa, P. Saparpakorn, S. Hannongbua, P. Wolschann, S. Prueksaaroon and P. Pungpo, Elucidating structural and dynamical properties of a TMC278 analog as a highly potent HIV-1 RT inhibitor through molecular dynamic simulations, The 6th International Symposium of The Protein Society of Thailand, 2011, 217-222.

15. **P. Kamsri**, A. Punkvang, K. Kun-asa, A. Kumkong, P. Saparpakorn, S. Hannongbua and P. Pungpo, The Investigation in the binding mode of triclosan and its derivatives to enoyl-ACP reductase of *Mycobacterium tuberculosis* using molecular docking calculations, The 5th UBU Research Conference Proceeding, 2011, 386-393.

CURRICULUM VITAE

NAME	Mr. Pharit Kamsri
EDUCATION	Ubon Ratchathani University, 2007-2010, B.Sc. (Chemistry)
SCHOLARSHIPS	Young Scientist and Technologist Programme: YSTP, NSTDA, 2010
	The Royal Golden Jubilee Ph.D (RGJ-Ph.D.) Program (PHD0004/2554), TRF, 2011-2015
AWARDS	Bangkok Bank Young Chemist Award (14ACC), Thailand, 2011
	Poster Awards, The 7th International Symposium of the Protein Society of Thailand, Thailand, 2012
	Distinguish RGJ Ph.D. Student Research Awards, the RGJ Ph.D. Congress XVI, Thailand Research Fund, Thailand, 2015
VISITTINGS	May 10–October 22, 2013. Institute for Theoretical Chemistry, University of Vienna, Austria (Prof. Dr. Peter Wolschann)
	April 13-September 30, 2014. Quantum Chemistry Group, Institute of Transformative Bio-Molecules (WPI-ITbM) and Department of Chemistry, Graduate School of Science, Nagoya University, Japan (Prof. Dr. Stephan Irle)
	October 15-December 15, 2014, Shanghai Institute of Meteria Medica, Chinese Academy of Science, China (Prof. Dr. Weiliang Zhu)
	September 2-15, 2015, Shanghai Institute of Meteria Medica, Chinese Academy of Science, China (Prof. Dr. Weiliang Zhu)
	September 20-October 31, 2015, Institute of Pharmacy, Martin Luther University of Halle-Wittenberg, Germany (Prof. Dr. Wolfgang Sippl)