

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA
RECHERCHE SCIENTIFIQUE

UNIVERSITE MOHAMED KHIDER BISKRA

Faculté des sciences exactes et
des sciences de la nature et de la vie

Département des Sciences de la matière

THESE

Présentée par

Khamouli Saida

En vue de l'obtention du diplôme de :

DOCTORAT EN SCIENCES

Option :

Chimie informatique et pharmaceutique

Intitulée:

*Study of the structure, QSAR properties and molecular
docking of some amino-pyrimidine derivatives as Novel
Mycobacterium tuberculosis inhibitors for drug design*

Soutenue le : 18/12/2019

Devant la commission d'Examen

Mr. Touhami LANEZ	Prof.	Université d'El-Oued	Président
Mr. Salah BELAIDI	Prof.	Université de Biskra	Directeur de thèse
Mr. Lotfi BELKHIRI	Prof.	Université de Constantine	Examineur
Mme. Dalal HARKATI	MC/A	Université de Biskra	Examinatrice

ACKNOWLEDGMENTS

First of all, I pray to Allah, the almighty for providing me this opportunity and granting me the capability to proceed successfully.

*I would like to express my great thanks to my supervisor Prof. **BELAIDI Salah**, his wide knowledge and his logical way of thinking have been of great value for me. His understanding, encouraging and personal guidance have provided a good basis for this thesis.*

*I wish to extend my sincere thanks to Mr. **LANEZ Touhami**, Professor at the University of El-Oued, for accepting to chair the dissertation of my thesis.*

*I would also like to thank the committee members, Mr. **BELKHIRI Lotfi** Professor at the University of Constantine and Mrs. **HARKATI Dalal MC/A** at the University of Biskra for agreeing to examine and judge the work.*

I would like to thank all the members of group of computational and pharmaceutical chemistry, LMCE Laboratory at Biskra University, with whom I had the opportunity to work. They provided a friendly and cooperative atmosphere at work and also useful feedback and insightful comments on my work.

DEDICATION

This thesis is profoundly dedicated to all.....

*my beloved family (my mother, my father, my husband, my
children, my sibling and my all friends).*

my thankful your endless love, Sypports and Advice.

Table of contents

LIST OF FIGURES.....	I
LIST OF TABLES.....	II
LIST OF ABBREVIATIONS.....	III
General Introduction	2

CHAPTER I: The disease of tuberculosis

1. Tuberculosis.....	6
1.1 Introduction.....	6
1.2 TB Pathogenesis.....	7
1.3 Symptoms.....	9
1.4 Treatment and Prevention of Tuberculosis.....	9
1.4.1 Treatment of Tuberculosis.....	10
1.4.2 Mechanisms of Drug Resistance.....	10
1.4.3 Prevention of the Tuberculosis	11
2. Mycobacterium tuberculosis.....	12
2.1 General Characteristics.....	12
2.2 Taxonomy.....	12
2.3 Classification.....	13
2.4 Morphology.....	13
2.5 Cell-wall.....	14
3. Serine/threonine protein kinases (STPKs).....	15
5. Challenges in the Development of New Drugs.....	16
Reference.....	18

CHAPTER II: Computational methods in drug discovery

1. Introduction.....	24
2. Computer-aided drug design.....	25
2.1 Introduction.....	25
2.2 CADD applications in drug discovery and development.....	25
2.3 Classification of CADD.....	27
2.3.1 Ligand-based drug design (LBDD).....	27
2.3.2 Structure-based drug discovery (SBDD).....	28

2.4	Molecular mechanics.....	29
2.5	Quantum mechanics.....	29
3.	Quantitative structure-activity relationships (QSAR).....	31
3.1	Introduction.....	31
3.2	History of QSAR.....	32
3.3	Tools and Techniques of QSAR.....	33
3.3.1	Biological Parameters.....	33
3.3.2	Molecular Descriptors.....	34
3.3.3	Statistical methods.....	36
3.3.4	Multiple linear regressions (MLR).....	36
3.4	Validation of QSAR Models.....	37
3.4.1	Internal Validation.....	38
3.4.2	External validation.....	40
4.	Molecular Docking.....	41
4.1	Introduction.....	41
4.2	Search Algorithms.....	41
4.3	Scoring functions.....	43
	Reference.....	44

**CHAPTER III: Qualitative Structure-Activity Relationships and 2D-QSAR Modeling
of PKNB Inhibition by Amino-pyrimidines Derivatives**

1	Introduction.....	51
2	Materials and methods.....	52
2.1	Data Set.....	52
2.2	Molecular descriptors.....	52
2.3	Model validation.....	53
3	Results and discussion.....	54
3.1	Structure Activity Relationship (SAR) and drug likeness.....	54
3.2	Quantitative Structure-Activity Relationships Studies.....	61
3.2.1	Internal validation.....	63
3.2.2	External Validation.....	67
4	Conclusion.....	68
	Reference.....	69

CHAPTER III: In silico Molecular docking and ADMET study of amino-pyrimidine derivatives as mycobacterium tuberculosis PKnB

1	Introduction.....	74
2	Material and methods.....	75
	2.1 Protein Preparation Structure.....	75
	2.2 Ligands structure optimization.....	76
	2.3 Molecular Docking.....	76
	2.4 Drug Likeness and Pharmacokinetic Properties of selected ligands.....	77
3	Results and discussion.....	78
	3.1 Validation of Model structures.....	78
	3.2 Prediction of binding sites.....	79
	3.3 Protein-Ligand Docking.....	80
	3.4 Drug-likeness and ADMET Properties of selected ligands.....	89
	3.4.1 Drug-likeness of selected ligands.....	89
	3.4.2 ADMET study of selected ligands.....	90
4	Conclusion.....	92
	Reference.....	93
	General Conclusion.....	98

LIST OF FIGURES

Figure I.1: Estimated TB incidence rates, 2017.....	7
Figure I.2: TB pathogenesis.....	8
Figure I.3: Morphological variations in <i>M. tuberculosis</i> . (A). Thin section transmission electron micrograph of <i>Mtb</i> . (B). Scanning electron microscope shows shape variation in <i>Mtb</i> at exponential phase of growth.....	14
Figure I.4: The mycobacterial cell wall.....	15
Figure II.1: CADD in drug discovery.....	27
Figure II.2: Flow chart of CADD processes.....	28
Figure III.1: 2D Structure of amino-pyrimidine.....	51
Figure III.2: Structure of amino- pyrimidine derivatives.....	56
Figure III.3: Predicted plot versus experimentally observed anti-tubercular activity.	66
Figure III.4: Plots of the residual against the experimental values of pIC50.....	66
Figure IV.1: Three-dimensional crystal structure of the target protein.....	75
Figure IV.2: Ramachandran plot for 2FUM protein generated by Procheck.....	78
Figure IV.3: The five cavities MVD- detected cavities in PknB, PDB code; 2FUM.....	79
Figure IV.4: Hydrogen bonds interaction between ligands L9 (A), L12 (B), L21(C) and MIX(D) and residues of active site of 2FUM.....	86
Figure IV.5: LigPlot generated for the best poses obtained with ligands L9 (A), L12 (B), L21 (C) and receptor 2FUM,.....	88

LIST OF TABLES

Table II.1: Types of biological data utilized in QSAR analysis.....	34
Table III.1: Drug-likeness parameters of amino-Pyrimidine derivatives.....	59
Table III.2: Values of molecular descriptors used in the regression analysis.....	61
Table III.3: Experimental and predicted activities pIC50 of the molecules under study.	64
Table III.4: Cross-validation parameters.....	65
Table IV.1: Cavity information of enzyme (2FUM).....	80
Table IV.2: The structure of amino-pyrimidine derivatives with docking scores.....	81
Table IV.3: Prediction interactions.....	88
Table IV.4: Physicochemical property of anti-tubercular ligands.....	89
Table IV.5: Drug likeness prediction using PreADMET Server.....	90
Table IV.6: ADME properties of ligands.....	90
Table IV.7: Toxicological properties of anti-tubercular agents.....	92

LIST OF ABRÉVIATIONS

2D	2 Dimensional
3D	3 Dimensional
ADME	Absorption, Distribution, Metabolism, and Excretion
AMBER	Assisted Model Building with Energy Refinement
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
ATP	Adenosine Triphosphate
BCG	Bacille Calmette-Guérin
Caco-2	Human colon adenocarcinoma
CADD	Computer-Assisted Drug Design
DFT	Density Functional Theory
DNA	Deoxyribonucleic Acid
ΔE	HOMO–LUMO Gaps
HE	Hydration Energy
HF	Hartree-Fock
HIV	Human Immunodeficiency Virus
HOMO	Highest Occupied Molecular Orbital
HIV	Human Immunodeficiency Virus
HTS	High-throughput screen
IC50	Half maximal Inhibitory Concentration
KatG	catalase-peroxidase
LBDD	ligand-based Drug Design
LDA	Local Density Approximations
LE	Ligand Efficiency
LipE	Lipophilic Efficiency
LOO	leave-one-out cross-validation
LTBI	latent tuberculosis infection
LUMO	Lowest Unoccupied Molecular Orbital
MDR-TB	Multidrug-Resistant Tuberculosis
MLR	Multiple Linear Regression
MM	Molecular Mechanics
MMFF	Merck Molecular Force Field

LIST OF ABRÉVIATIONS

MP2	Møller-Plesset level 2
MPO	Multi-Parameter Optimization
MW	Molecular Weight
NBO	Natural Bond Orbital
NMR	Nuclear Magnetic Resonance
PM3	Parameterized Model number 3
Pol	Polarizability
PRESS	Predicted Residual Sum of Squares
PSA	Polar Surface Area
PZA	Pyrazinamide
QSAR	Quantitative Structure–Activity Relationship
QSPR	Quantitative Structure–Property Relationship
RCSB	Research collaboratory for structural Bioinformatics
RIF	Rifampicin
RMSE	Root-Mean Squared Error
RNA	Ribonucleic Acid
rpoB	RNA polymerase beta
rRNA	ribossomal Ribonucleic Acid
SAG	Surface Area Grid
SAR	Structure–Activity Relationships
SCF	Self-Consistent-Field procedure
SE	standard error
SSY	Sum of The Squares of The Response Value
STPKs	Serine/threonine protein kinases
STR	Streptomycin
SVM	Support Vector Machines
TB	Tuberculosis
VdW	Van der Waals
WHO	World Health Organization
XDRTB	Extensively Drug Resistant Tuberculosis

General Introduction

Tuberculosis (TB) is one of the oldest human diseases and also one of the major diseases that endangers human's health. With the drug-resistant TB especially multidrug-resistant TB (MDR-TB) and extensively drug-resistant TB (XDR-TB) dissemination, the treatment of TB faced up more difficulties [1].

Thereby, search for new anti-TB drug with high activity at low concentration is urgent need which can be achieved using Computer-aided drug design. Computer-aided drug design is a powerful tool for designing new drug or modifying the existing one. These methods are dependent on bioinformatics tools, applications and databases, simple, non expensive, and accelerate the process of designing novel and potent molecules with desired biological activity [2, 3]. The significant gain in knowledge and structure information of both biological macromolecules and small molecules facilitate CADD to be extended and broadly applied to almost every stage in the drug discovery and development stage, from target identification and validation, to lead discovery and optimization, and preclinical tests.

Molecular mechanics MD simulation has become one of the most influential tools to predict the conformation of small molecules, as well as for modeling, conformational changes within a biological target upon binding by small molecules [4, 5]. Semi-empirical methods such as *ab initio* methods or density functional theory *DFT* are most often used to provide the expected optimized parameters for molecular mechanics calculations and to estimate important electronic properties (electrostatic potential, polarizability, *etc.*) of the drug candidate that influence its binding affinity [6].

In this thesis, we have used different CADD approaches to investigate the detailed structural mechanisms of the target proteins and search for inhibitors, the detailed structural mechanisms of the target proteins and search for inhibitors.

Ligand-based drug design (LBDD) is an appropriate method to apply if there are experimentally active compounds that bind to the biological target of interest. Quantitative structure–activity relationship (QSAR) can be used to derive a correlation between theoretically calculated properties of molecules and their experimentally obtained biological activity. The resulting correlation derived from QSAR can in turn be further used to predict the activity of new analogues.

Structure-based drug design (SBDD) virtual screening involves using the known 3D structure of a target protein to predict binders, through the process of the docking. Docking is the computational determination of a binding pose and "binding free energy" of a ligand to a receptor. Determination of the binding pose and "binding free energy" is crucial for understanding the important ligand-receptor interactions and mechanism of action, thus valuable in the design of new drug. In molecular docking thousands of possible poses of a ligand inside a receptor cavity are tried and evaluated; the pose with the lowest energy score is predicted as the "best match", the binding pose [7].

In silico tools, it is possible to accelerate the drug-discovery process by modeling the most relevant ADMET properties [8]. A molecule could be too toxic, too quickly eliminated from the body, possess fast metabolic reaction, unstable, too challenging to synthesize in large volume, or too expensive to produce.

This manuscript consists of four chapters. :

The first chapter represents a bibliographic review describing the tuberculosis (Treatment and Prevention of Tuberculosis, *Mycobacterium tuberculosis*, Serine/threonine protein kinase B).

The second chapter describes the different Computational methods in drug discovery (Quantitative structure-activity relationships (QSAR), Molecular Docking).

The third chapter comprises a various 'multi-parameter optimization' (MPO) approaches such as lipophilicity profile, rules of thumb and calculated metrics methods used to highlight the Structure Activity/Property Relationships of a bioactive series of aminopyrimidines, and we have studies Quantitative relationship (QSAR) between physicochemical properties and biological activity (work published in: *Journal of Bionanoscience*, Vol 11(4), 31–309, 2017 and *Journal of Turkish Computational and Theoretical Chemistry (TC&TC)*, Vol 2(2), 16–27, 2018).

The fourth chapter we have studies Structure-based virtual screening using docking analysis and (ADMET) absorption, distribution, metabolism and toxicity study forward few of them as plausible lead molecule or a novel class of drugs with enhanced pharmacological properties (work published in: press in the journal : *Journal of Fundamental and Applied Sciences*, Vol 11(2), 914-939, 2019).

Reference

1. S. Tang, Q.Zhang ,J.Yu, Y. Liu., W.Shaet *al.*, “Extensively Drug-Resistant Tuberculosis, China,” *Emerg Infect Dis*, vol. 17, no. 3, pp. 558–560, Mar. 2011.
2. V.H.Masand, K.N. Patil, G.M. Nazerruddin, R.D. Jawarkar, S.O. Bajaj, “Correlation potential of Wiener index vis-a-vis molecular refractivity: anti-malarial activity of xanthone derivatives,”*Org.Chem Indian J.*, vol. 6, no. 1, pp. 30–38. 2010.
3. V. S. Rao and K. Srinivas, “Modern drug discovery process: An in silico approach,” *JBSA*, vol. 3, no. 5, pp. 89–94, Jun. 2011.
4. P. L. Freddolino, C. B. Harrison, Y. Liu, and K. Schulten, “Challenges in protein folding simulations: Timescale, representation, and analysis,” *Nat Phys*, vol. 6, no. 10, pp. 751–758, Oct. 2010.
5. M. H. Baig *et al.*, “Insight into the effect of inhibitor resistant S130G mutant on physico-chemical properties of SHV type beta-lactamase: a molecular dynamics study,” *PLoS One*, vol. 9, no. 12, p. e112456, 2014.
6. P. Comba, T. W. Hambley, and B. Martin, *Molecular Modeling of Inorganic Compounds*, 3 edition. Weinheim: Wiley-VCH, 2009.
7. S. J. Teague, “Implications of protein flexibility for drug discovery,” *Nat Rev Drug Discov*, vol. 2, no. 7, pp. 527–541, Jul. 2003.
8. J. Augen, “The evolving role of information technology in the drug discovery process,” *Drug Discov. Today*, vol. 7, no. 5, pp. 315–323, Mar. 2002.

Chapter I:

The disease of tuberculosis

1. Tuberculosis

1.1 Introduction

Tuberculosis (TB) is a contagious but curable infectious disease caused by the pathogen, *Mycobacterium tuberculosis* (*Mtb*) and is again becoming a major cause of mortality worldwide. The *Mtb* and its pathogenic strains cause infection mainly in the oxygen-rich macrophages of the lungs. The main causes for the resurgence of this once nearly eradicated infectious diseases are the reduction in the emphasis on TB control programs, the declined socioeconomic standards and also the emergence of immune deficiency states like, AIDS. Despite the fact that TB has been recognized for thousands of years and its etiological agent has been identified since the earliest days of medical microbiology, TB continues to loom as one of the largest infectious diseases, with enormous global burden of morbidity and mortality. TB thrives in impoverished or malnourished communities; individuals weakened by immunological deficiencies and situations where healthcare delivery is poor [1, 2]. According to global tuberculosis report by World Health Organization (WHO), in 2017, an estimated 10 million people developed TB and 1.3 million deaths among HIV-negative people and almost 300000 deaths among HIV-positive people.

The geographic incidence of infection and therefore disease burden varies greatly. Most of the estimated number of cases in 2017 occurred in the South-East Asia Region (44%), the African Region (25%) and the Western Pacific Region (18%); smaller proportions of cases occurred in Eastern Mediterranean Region (7,7%), Region of the Americas (2,8%) and European Region (2,7%). The 30 high TB burden countries accounted for 87% of all estimated incident cases worldwide, and eight of these countries accounted for two thirds of the global total: India (27%), China (9%), Indonesia (8%), the Philippines (6%), Pakistan (5%), Nigeria (4%), Bangladesh (4%) and South Africa (3%) of the global total shows estimated TB incidence rates by country in 2017 (Figure I.1) [1].

Furthermore, the development of multidrug-resistant tuberculosis (MDR-TB) and extensively drug-resistant tuberculosis (XDR-TB) strains, together with the spread of risk factors such as human immunodeficiency virus (HIV), acquired immunodeficiency

syndrome (AIDS) and diabetes [3], continue making *Mtb* a health concern in developed countries and strengthened the urge to develop new treatment infection strategies.

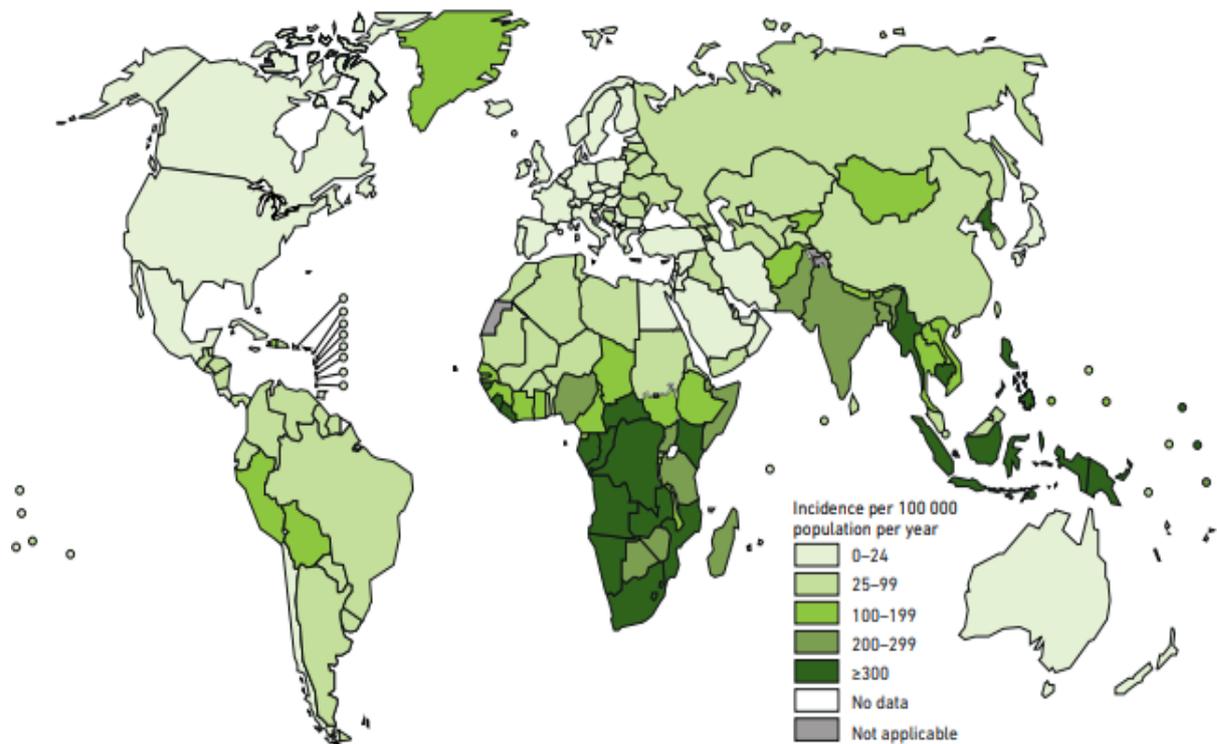


Figure I.1: Estimated TB incidence rates, 2017.

1.2 TB Pathogenesis

Mtb infection occurs when few tubercle bacilli dispersed in the air from a patient with active pulmonary TB reach the alveoli of the host. Here, *Mtb* is quickly phagocytized by professional alveolar macrophages that most often can kill the entering bacteria thanks to the innate immune response (Figure I.2.) [4]. If the bacilli can survive this first line of defense, it starts actively replicating in macrophages; diffuse to nearby cells including epithelial and endothelial cells, reaching in few weeks of exponential growth a high bacterial burden. During these early steps of infection, *Mtb* can diffuse to other organs through the lymphatics and by haematogenous dissemination where it can infect other cells [5].

Thereafter, once the adaptive immune response kicks in, migration to the site of primary infection of neutrophils, lymphocytes and other immune cells form a cellular infiltrate that later assume the typical structure of a granuloma [6]. Fibrotic components cover the granuloma that becomes calcified such that bacilli remain encapsulated inside and protected by the host immune response. This primary lesion, classically termed the Ghon complex [7], was thought to be the “sanctuary” of *Mtb* during latent infection, with bacilli persisting in a dormant, non-metabolically active state, for years, decades, or most often for lifetime. In this scenario, when, during latent infection, for unknown reasons, bacilli would start replicating inside this primary lesion, active disease would ensue [8].

A major corollary of this hypothesis, with relevant pathophysiological and clinical implications, was that reactivation of TB originated from this very primary site of infection. This hypothesis was challenged since the early 20th century, when it was shown that viable and infective bacilli were found in unaffected portion of lung tissues of infected guinea pigs or human necropsy rather than from the central core of the tuberculous lesions [9, 10].

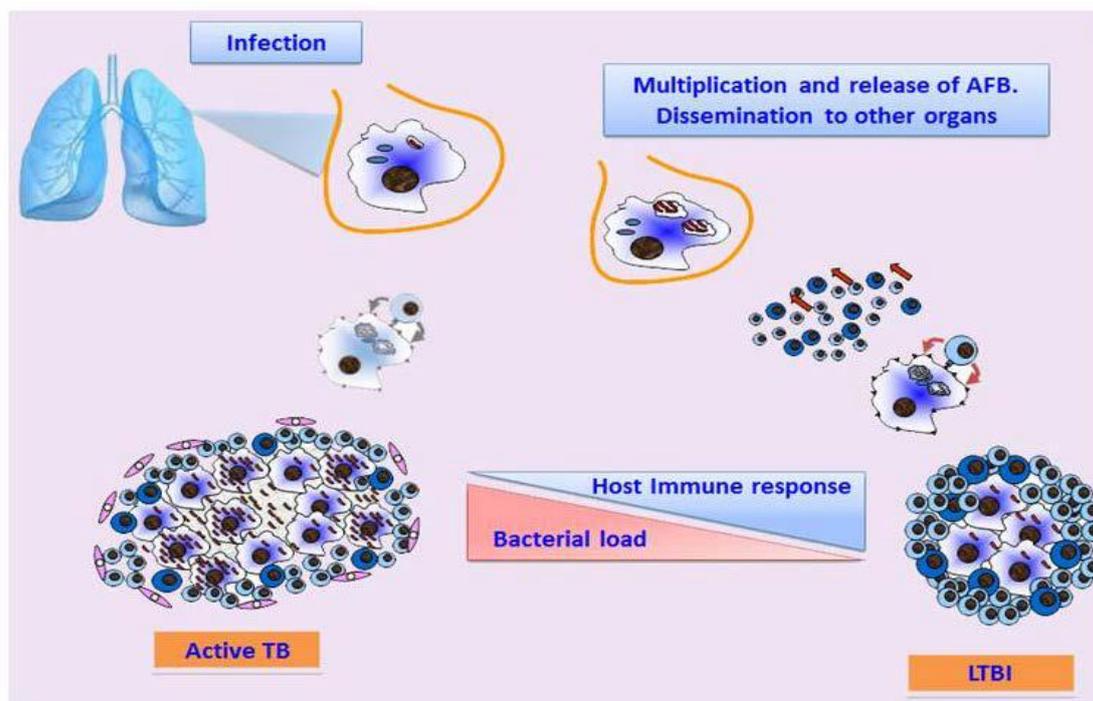


Figure I.2: TB pathogenesis.

1.3 Symptoms

Many of the worst symptoms of active tuberculosis arise as a direct result of the extensive tissue damage that the bacteria do to your lungs . Typical signs and symptoms of active tuberculosis include a bad dry cough lasting for more than three weeks that may cause you to cough up bloody sputum. You are also likely to experience night sweats, fever and weight loss (the reason why historically tuberculosis was called “consumption”) as your body tries to fight off the infection.

For about one in five people, the infection is so severe that cavities begin to form within the lung tissue. If these areas start to bleed, tuberculosis bacteria are able to enter the blood stream. If this happens, they can travel to other parts of your body, causing additional symptoms. A tuberculosis infection outside of the lungs is called an extra-pulmonary tuberculosis infection and most commonly occurs in your lymphatic system, your genitourinary system, and/or in your bones and joints. However, in some cases extra-pulmonary tuberculosis is disseminated, which means the infection has spread widely throughout whole body [11].

1.4 Treatment and Prevention of Tuberculosis

1.4.1 Treatment of Tuberculosis

Treatment of TB is used not only to cure the disease but also to interrupt the transmission and to prevent relapse (most relapses occur within 6-12 months after the end of therapy) [12]. In 1994, the WHO introduced the Directly Observed Treatment Short Course (DOTS), a strategy for the detection and treatment of TB, in which patients are observed to take each dose of anti-TB medication, until the end of therapy [13]. Monthly sputum specimens are then taken until 2 consecutive specimens are negative. However, the effectiveness of DOTS is facing new challenges due to the increase of MDRTB and the emergence of XDRTB. This lead to a new strategy called DOTS-plus, a comprehensive management initiative built upon the DOTS strategy with the goal of preventing further development and spread of MDRTB [14, 15].

Treatment can be divided into first-line and second-line drugs : the first-line drugs used are INH (Isoniazid), RIF (Rifampicin), pyrazinamide (PZA), ethambutol (EMB) and

streptomycin (STR) and the second-line drugs include fluoroquinolones, aminoglycosides such as kanamycin and amikacin, cyclic peptides like capreomycin, D-cycloserine, ethionamide, and r-amino salicylic acid. Each treatment regimen for pulmonary TB caused by susceptible organisms has an initial 2 months intensive phase with INH, RIF, PZA and EMB, followed by a continuation phase with INH and RIF for 4 to 7 months. STR can be used as an interchangeable drug with EMB in the initial phase of treatment. However, STR is only recommended to be interchangeable with EMB when the organism is known to be susceptible to the drug or the patient is from a community in which STR resistance is unlikely [16,13]

1.4.2 Mechanisms of Drug Resistance

Drug resistance is a natural phenomenon and the mechanisms via which bacteria develop resistance are diverse and complex. *M. tuberculosis* resistance occurs due to its highly hydrophobic cell wall (Figure 1.2) that is impermeable to most drugs and the resistance determinates encoded by its genome [17, 18].

The resistance to RIF is due to mutations in the *rpoB* gene which is responsible for producing the beta subunit of the DNA-dependent RNA polymerase. However, there is a small fraction of resistant bacteria that does not exhibit this mutation suggesting there are other mechanisms of RIF resistance. INH is a pro-drug that is activated by an enzyme called KatG and it inhibits an enzyme that is involved in fatty acid biosynthesis called InhA. Resistant to INH occurs due to mutations in the either *katG* or *inhA* genes [19]. Streptomycin resistance occurs due to mutations in the 16s rRNA genes *rrs* and *rpsL*, the gene encoding ribosomal protein S12 [20]. ETH resistance is acquired via the mutations in the *embB* gene that encodes for the arabinosyl transferases that is inhibited by ETH [21].

Resistance to fluoroquinolones is due to specific mutations in the *gyrA* and *gyrB* genes that encode for DNA gyrase A and B subunits respectively [22]. Aminoglycosides have the same mechanism of resistance as described for streptomycin. Since the MOA of polypeptides has not been characterized, the mechanism of resistance is not fully understood; however it also changes the 16s rRNA and thus may confer cross-resistance from streptomycin [19]. Despite the impact of global tuberculosis and the prevalence of drug resistance, no new classes of drugs have been approved for Tb treatment by the US Food and Drug Administration since 1972 [23]. There is an urgent need for new and

alternate drugs to target Tb with novel MOA to prevent cross-resistance from current anti-tubercular drugs [24].

1.4.3 Prevention of the Tuberculosis

Beside the treatment of both open and latent form of tuberculosis, i.e. intervention on infected patients, there is a question of prevention of mycobacterial infections. The combination of a new highly effective pre-exposure vaccine with a more effective preventive therapy would potentially have a dramatic effect on incidence. A highly effective post-exposure vaccine may have the same effect as a preventive treatment. Unfortunately, these methods are not yet available [25].

BCG vaccine is currently the most widely used only available vaccine against TB. It has been administered up to three billion people without serious complications. Although it displays some advantages (easy administration, it can confer immunity for a long period, a very efficient adjuvant for immunity induction, a low cost), recent molecular analysis have revealed that genetic modifications formed sub strains along time and a complete virulence acquisition is also suspected for the vaccine. Another problem is that there can be a gradual loss of T-memory cell population during in the course of time. Regrettably, BCG vaccination has had the limited impact on the global burden of TB. Although vaccine has been effective to prevent severe and fatal forms of TB in young children (e.g. meningeal form), it does not protect against pulmonary TB in adults sufficiently. Since the efficacy varies from zero to 80 %, the new improved vaccine is searched constantly [26, 27].

A second approach consists of thorough prevention leading to the reduction of such factors, which increase the risk of progression from infection to disease; HIV, the malnutrition, smoking, diabetes mellitus and alcohol misuse are individual risk factors that can double or triple the risk of development of active tuberculosis. Indoor air pollution is an additional possible causal factor. A wide range of disorders (e.g. silicosis, malignant diseases and chronic systemic illnesses) and immunosuppressive treatment are established risk factors for TB; some other common situations were proposed, but very little research has been done to test these hypotheses (chronic helminthes infections, depression, mental illness, pregnancy and the postpartum, outdoor air pollution) [25].

2. Mycobacterium tuberculosis

2.1 General Characteristics

Tuberculosis (TB) is caused by the infectious agent known as *Mycobacterium tuberculosis* (MTB). This rod-shaped bacterium, also called Koch's bacillus, was discovered by Dr. Robert Koch in 1882 [28]. MTB is a small, slow-growing bacterium that can live only in people. It is not found in other animals, insects, soil, or other nonliving things. MTB is an aerobic bacterium, meaning it needs oxygen to survive. For this reason, during active tuberculosis disease, MTB complexes are always found in the upper air sacs of the lungs. The bacterium is a facultative intracellular parasite, usually of macrophages, and has a slow generation time, 15-20 hours, and a physiological characteristic that may contribute to its virulence. The bacteria usually attack the lungs, but MTB bacteria can attack any part of the body such as the kidney, spine, and brain. If not treated properly, disease can be fatal. It is transmitted from person to person via droplets from the throat and lungs of people with the active respiratory disease.

2.2 Taxonomy

In the Bacteria Kingdom, Genus *Mycobacterium* belongs to the Phylum *Actinobacteria*, Order *Actinomycetales* (Suborder *Corynebacterineae*) and Family *Mycobacteriaceae*.

2.3 Classification

Mycobacteria can be separated into different groups according to their phenotypic characteristics. The most used classification system was created by Runion in 1959 and establishes four groups based on culture characteristic, the incubation period required for growth and the development of pigmentation in the presence/absence of light [29,30]. The first three groups include slow growth mycobacteria that require longer periods of incubation (*e.g. M. tuberculosis*) and the fourth group comprises rapid growth mycobacteria that form colonies within seven days of incubation (*e.g. M. smegmatis*):

o Group 1 (Photochromogens) – mycobacteria that produce nonpigmented colonies

When grown in the dark and pigmented colonies only after exposure to light (*e.g.:* *M. kansasii*, *M. marinum*).

o Group 2 (Scotochromogens) – mycobacteria that produce yellow to orange

colonies when grown in the dark (e.g.: *M. gordonae*, *M. xenopi*).

o Group 3 (Non-chromogens) – mycobacteria nonpigmented in the light or dark.

However, older cultures can develop a yellow pigmentation (e.g. *M. tuberculosis*, *M. avium*, *M. bovis*, *M. ulcerans*).

o Group 4 – rapid growth mycobacteria that have a slight yellow pigment that does not intensify after light exposure (e.g. *M. fortuitum*, *M. chelonae*).

2.4 Morphology

Mycobacteria are typically rod-shaped, non-spore forming, aerobic bacteria, classified as acid-fast bacilli. The dimensions of the bacilli have been reported to be 1-10 μm in length (usually 3-5 μm), and 0.2 -0.6 μm width. Variable morphology can be observed when grown on solid media and some species exist as shorter cocci-bacilli or curved rods on artificial media [31].

The reported morphological variation in *M. tuberculosis* are classified in two categories; i) those which are frequently seen at exponential phase of growth that is rod, V, Y-shape, branched or buds, and ii) those that are seen occasionally under stress or environmental conditions which are round, oval, ultra-virus, spore like, and cell wall defiant or L-forms (Figure I.3) [32].

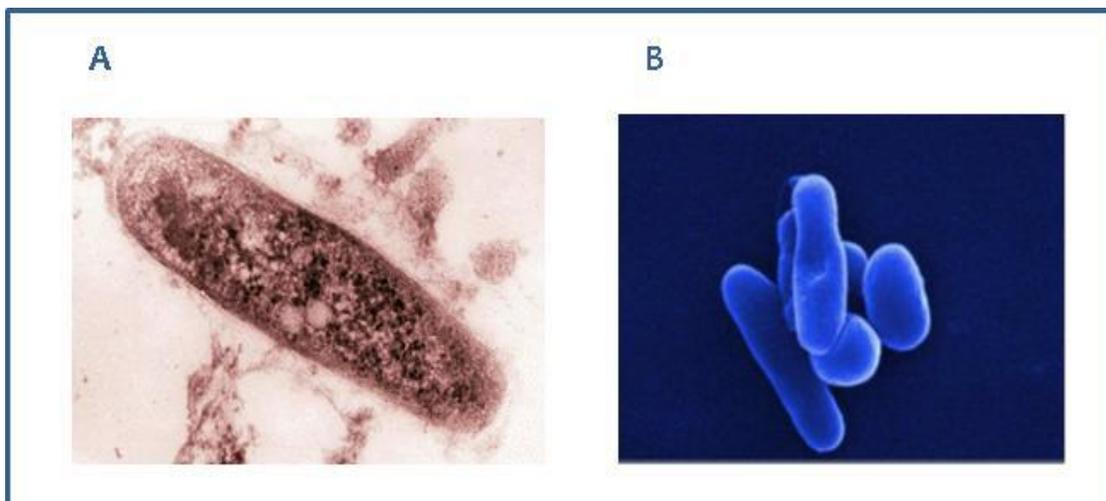


Figure I.3: Morphological variations in *M. tuberculosis*. (A). Thin section transmission electron micrograph of Mtb. (B). Scanning electron microscope shows shape variation in Mtb at exponential phase of growth.

2.5 cell-wall

The cell wall structure of Mtb (Figure I.4), unique among prokaryotes, is associated with the pathogenicity of Mtb [34-36]. The richness in high molecular weight lipids represents the complexity of the cell wall [38-42]. Unusual impermeable properties of Mtb cell wall are thought to be advantageous for the bacilli in stressful conditions of osmotic shock and the polymers, covalently linked with peptidoglycan and trehalose dimycolate, provide a thick layer involved in Mtb resistance to antibiotics and the host defense mechanisms [43].

The mycobacterial cell wall consists of a capsule, a core and an inner membrane (Figure I.4). The capsule is made up of free lipids and mycolates such as the phosphatidylinositol mannosides (PIMs) and lipoarabinomannan (LAM). The core consists of peptidoglycan (PG) and mycolic acid connected by an arabinogalactan (AG) polysaccharide, collectively called the mycolyl arabinogalactan–peptidoglycan (mAGP) complex. And lastly the inner membrane which consists of a lipid bilayer. The presence of such a large amount of lipid in the cell wall makes it acid-fast. This is evident when the cells are stained with Ziehl-Neelsen stain. They resist decolorization with acid alcohol and retain the primary carbol fuchsin stain giving it the characteristic red/pink color.

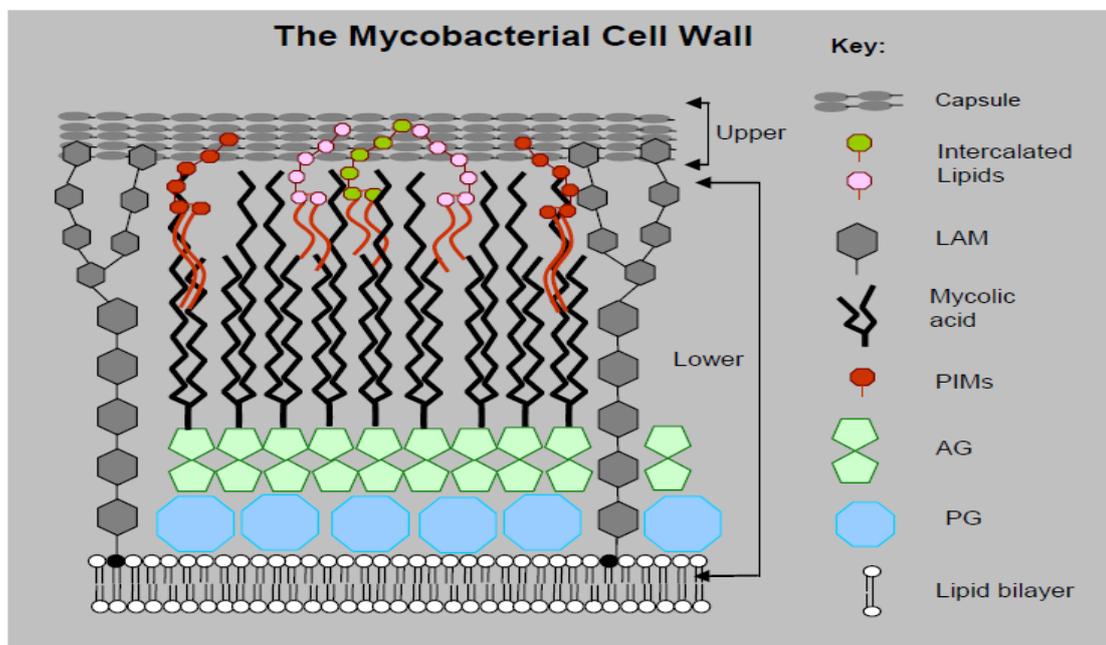


Figure I.4: The mycobacterial cell wall.

3. Serine/threonine protein kinases (STPKs)

Protein phosphorylation/dephosphorylation is a critical regulatory mechanism by which extracellular cues are transduced into cellular responses. Traditionally, two-component systems comprising a histidine kinase sensor and the associated response regulator were thought to be responsible for phosphorylation-mediated signaling in prokaryotes [44, 45]. Two-component pathways are rare in eukaryotes, and phosphotransferbased signaling predominantly involves phosphorylation/ dephosphorylation on serine, threonine, or tyrosine residues, often in a cascade.

The presence of a eukaryotic-like serine/ threonine protein kinase (STPK) in a prokaryote was first reported in 1991 (*pkn1* in *Myxococcus xanthus*) [46]. These “eukaryotic-like” STPKs play important roles in bacterial cellular processes, including cell division, cell wall synthesis, cell metabolism, and dormancy exit. Analysis of the *Mycobacterium tuberculosis* genome sequence suggested the presence of 11 putative eukaryotic-like STPKs and three protein phosphatases [47]. Except for PknG and PknK, all of these kinases were predicted to have a transmembrane domain [47, 48]. All of the kinases possessed the protein kinase “signature” motifs, including 11 conserved subdomains as per Hanks’ criteria, and amino acid sequence alignment of these STPK family members revealed that 15 catalytically important residues were conserved across all of them [48,49].

The *M. tuberculosis* STPKs affect key mycobacterial processes: signal transduction mediated by PknA and PknB plays an important role in determining cell shape, morphology, and possibly cell division ; PknG and PknH influence *M. tuberculosis* virulence, adaptation, and growth within the host [50]; and PknF affects cell division, growth rate, morphology, and glucose transport [51]. Recently, Ortega *et al.* [52] proposed a role for PknB as a replication switch in response to hypoxia. They demonstrated that PknB activity is necessary for reactivation of cells from the hypoxic state.

4. Challenges in the Development of New Drugs

Along with the socioeconomic and host factors that underlie the serious global burden of TB, a fundamental problem that hinders more effective TB control is the ability of *MTB* to persist in the host and to develop drug resistance, often because of poor adherence to lengthy therapy. Despite a resurgence of TB, development of new drugs to treat the disease has stagnated in the face of numerous scientific and economic obstacles. Showing of the superiority of new agents constitutes the most convincing clinical evidence of drug efficacy, but in the case of drug-sensitive disease this may be infeasible given the high efficacy rates of existing regimens, the need for extended follow-up, and the large number of participants required supporting statistical conclusions [53].

Primarily, overall funding for TB research in general, and drug discovery in particular, remains alarmingly inadequate. TB research is funded in competition with all other areas of biomedicine and is clearly not receiving funds commensurate with the global dimension of the disease and the probability that untreatable forms of TB will become increasingly widespread [54]. Besides, the TB drug market is associated with insufficient profit opportunity or investment return to instigate pharmaceutical industries to develop new drugs [55].

Secondly, it is the lack of access to information, pharmaceutical expertise, compounds, and research tools. There would be great value, for example, in a publicly accessible database that collected thorough information about screenings of compounds and about analyses that indicate which targets in *MTB* appear to be “druggable.” Considering the limited resources for TB drug development, it is critical to avoid repetitive efforts, particularly multiple independent journeys to a dead-end. However, the molecular mechanisms responsible for mycobacterial dormancy, persistence, and drug resistance are not yet fully understood [56].

Thirdly, there are a number of constraints that have companies from investing in new anti-TB drugs. The research is expensive, slow and difficult and requires specialized facilities for handling *MTB*. There are few animal models that closely mimic the human TB disease. Development time of any anti-TB drug will be long. In fact minimum six

Month therapy will require with a follow up period of one year or more [57]. Lastly, the challenge of TB drug Research and Development is the long timeline of clinical trials. Phase II studies for TB drugs typically require at least two years, and pivotal trials a minimum of three years from beginning patient enrollment to finalized study reports. Furthermore, the fact that people must be treated with a combination of four drugs, rather than with a single drug, means that to replace the current regimen with a totally new three- or four-drug regimen by testing the substitution of one drug at a time into the standard regimen will require not a minimum of six years, but at least four times six years - over two decades - just for the clinical phase of development [58].

Reference

1. World Health Organization. Global tuberculosis report 2018. Geneva, World Health Organization, Nov2018.
2. N. Arya, M.K. Raut, S.G. Tekale, C.J. Shishoo and K.S. Jain , “ Tuberculosis: New Drug Discovery Pipelines,” *Austin J Anal Pharm Chem. .*, vol. 1, no. 3, pp. 1014. –1023, Sept 2014.
3. B. I. Restrepo, “Convergence of the tuberculosis and diabetes epidemics: renewal of old acquaintances,” *Clin. Infect. Dis.*, vol. 45, no. 4, pp. 436–438, Aug. 2007.
4. K. B. Urdahl, S. Shafiani, and J. D. Ernst, “Initiation and regulation of T-cell responses in tuberculosis,” *Mucosal Immunol*, vol. 4, no. 3, pp. 288–293, May 2011.
5. A. J. Wolf *et al.*, “Initiation of the adaptive immune response to Mycobacterium tuberculosis depends on antigen production in the local lymph node, not the lungs,” *J. Exp. Med.*, vol. 205, no. 1, pp. 105–115, Jan. 2008.
6. V. Balasubramanian *et al.*, “Allelic exchange in Mycobacterium tuberculosis with long linear recombination substrates,” *J. Bacteriol.*, vol. 178, no. 1, pp. 273–279, Jan. 1996.
7. T. H. M. Ottenhoff and S. H. E. Kaufmann, “Vaccines against Tuberculosis: Where Are We and Where Do We Need to Go?,” *PLOS Pathogens*, vol. 8, no. 5, p. e1002607, mai 2012.
8. G. Delogu, M. Sali, and G. Fadda, “The Biology of Mycobacterium Tuberculosis Infection,” *Mediterr J Hematol Infect Dis*, vol. 5, no. 1, Nov. 2013.
9. W. R. Bishai, “Rekindling old controversy on elusive lair of latent tuberculosis,” *International Journal of STD and AIDS*, vol. 356, no. SUPPL., pp. 2113–2114, 2000.
10. W. H. Feldman and A. H. Baggenstoss, “The residual infectivity of the primary complex of tuberculosis,” *Am J Pathol*, vol. 14, no. 4, pp. 473-490.3, Jul. 1938.
11. Health and medicine infectious diseases. What is tuberculosis? (2015). Retrieved from
12. E. C. Böttger and B. Springer, “Tuberculosis: drug resistance, fitness, and strategies for global control,” *Eur. J. Pediatr.*, vol. 167, no. 2, pp. 141–148, Feb. 2008.
13. World Health Organization. Communicable Diseases Cluster, “Guidelines for drug susceptibility testing for second-line anti-tuberculosis drugs for dots-plus,” World Health Organization, Technical documents WHO/CDS/TB/ 2001.288, 2001.

14. “Extensively drug-resistant tuberculosis (XDR-TB): recommendations for prevention and control,” *Wkly. Epidemiol. Rec.*, vol. 81, no. 45, pp. 430–432, Nov. 2006.
15. “World Health Organization/International Union Against Tuberculosis and Lung Disease (WHO/IUATLD). 2008. Global report Global Project on Anti-Tuberculosis Drug Resistance Surveillance 2002–2007. Anti-tuberculosis drug resistance in the world: fourth global report .
16. H. M. Blumberg *et al.*, “American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America: treatment of tuberculosis,” *Am. J. Respir. Crit. Care Med.*, vol. 167, no. 4, pp. 603–662, Feb. 2003.
17. S. T. Cole *et al.*, “Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence,” *Nature*, vol. 393, no. 6685, pp. 537–544, Jun. 1998.
18. S. H. Gillespie, “Evolution of drug resistance in *Mycobacterium tuberculosis*: clinical and molecular perspective,” *Antimicrob. Agents Chemother.*, vol. 46, no. 2, pp. 267–274, Feb. 2002.
19. P. J. Brennan, D. B. Young and B. D. Robertson, “*Handbook of Anti-Tuberculosis Agents*,” *Tuberculosis*, vol. 88, no. 2, pp. 85–170, 2008.
20. B. Springer, Y. G. Kidan, T. Prammananan, K. Ellrott, E. C. Böttger, and P. Sander, “Mechanisms of streptomycin resistance: selection of mutations in the 16S rRNA gene conferring resistance,” *Antimicrob. Agents Chemother.*, vol. 45, no. 10, pp. 2877–2884, Oct. 2001.
21. A. G. Amin, R. Goude, L. Shi, J. Zhang, D. Chatterjee, and T. Parish, “EmbA is an essential arabinosyltransferase in *Mycobacterium tuberculosis*,” *Microbiology*, vol. 154, no. Pt 1, pp. 240–248, Jan. 2008.
22. T. Kocagöz, C. J. Hackbarth, I. Unsal, E. Y. Rosenberg, H. Nikaido, and H. F. Chambers, “Gyrase mutations in laboratory-selected, fluoroquinolone-resistant mutants of *Mycobacterium tuberculosis* H37Ra,” *Antimicrob. Agents Chemother.*, vol. 40, no. 8, pp. 1768–1774, Aug. 1996.
23. L. Nguyen and C. J. Thompson, “Foundations of antibiotic resistance in bacterial physiology: the mycobacterial paradigm,” *Trends Microbiol.*, vol. 14, no. 7, pp. 304–312, Jul. 2006.

24. A. F. Chalker and R. D. Lunsford, "Rational identification of new antibacterial drug targets that are essential for viability using a genomics-based approach," *Pharmacol. Ther.*, vol. 95, no. 1, pp. 1–20, Jul. 2002.
25. K. Lönnroth *et al.*, "Tuberculosis control and elimination 2010-50: cure, care, and social development," *Lancet*, vol. 375, no. 9728, pp. 1814–1829, May 2010.
26. R. G. Ducati, A. Ruffino-Netto, L. A. Basso, and D. S. Santos, "The resumption of consumption -- a review on tuberculosis," *Mem. Inst. Oswaldo Cruz*, vol. 101, no. 7, pp. 697–714, Nov. 2006.
27. Global Alliance for TB Drug Development, "Tuberculosis. Scientific blueprint for tuberculosis drug development," *Tuberculosis (Edinb)*, vol. 81 Suppl 1, pp. 1–52, 2001.
28. "[The etiology of tuberculosis by Dr. Robert Koch. From the Berliner Klinische Wochenschrift, Volume 19 (1882)]," *Zentralbl Bakteriol Mikrobiol Hyg A*, vol. 251, no. 3, pp. 287–296, Mar. 1982.
29. E. H. Runyon, "Anonymous mycobacteria in pulmonary disease," *Med. Clin. North Am.*, vol. 43, no. 1, pp. 273–290, Jan. 1959.
30. J. Eisenstadt and G. S. Hall, "Microbiology and classification of mycobacteria," *Clin. Dermatol.*, vol. 13, no. 3, pp. 197–206, Jun. 1995.
31. J. T. Belisle and P. J. Brennan, "Chemical basis of rough and smooth variation in mycobacteria.," *J Bacteriol*, vol. 171, no. 6, pp. 3465–3470, Jun. 1989.
32. A. A. Velayati *et al.*, "Totally drug-resistant tuberculosis strains: evidence of adaptation at the cellular level," *Eur. Respir. J.*, vol. 34, no. 5, pp. 1202–1203, Nov. 2009.
33. P. Farnia *et al.*, "Growth and cell-division in extensive (XDR) and extremely drug resistant (XXDR) tuberculosis strains: transmission and atomic force observation," *Int J Clin Exp Med*, vol. 3, no. 4, pp. 308–314, Sep. 2010.
34. I. Smith, "Mycobacterium tuberculosis Pathogenesis and Molecular Determinants of Virulence," *Clinical Microbiology Reviews*, vol. 16, no. 3, p. 463, Jul. 2003.
35. C. E. Barry *et al.*, "Mycolic acids: structure, biosynthesis and physiological functions," *Prog. Lipid Res.*, vol. 37, no. 2–3, pp. 143–179, Aug. 1998.
36. E. Dubnau *et al.*, "Oxygenated mycolic acids are necessary for virulence of Mycobacterium tuberculosis in mice," *Mol. Microbiol.*, vol. 36, no. 3, pp. 630–637, May 2000.

37. M. S. Glickman, J. S. Cox, and W. R. Jacobs, "A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of *Mycobacterium tuberculosis*," *Mol. Cell*, vol. 5, no. 4, pp. 717–727, Apr. 2000.
38. P. J. Brennan and H. Nikaido, "The envelope of mycobacteria," *Annu. Rev. Biochem.*, vol. 64, pp. 29–63, 1995.
39. H. Christensen, N. J. Garton, R. W. Horobin, D. E. Minnikin, and M. R. Barer, "Lipid domains of mycobacteria studied with fluorescent molecular probes," *Mol. Microbiol.*, vol. 31, no. 5, pp. 1561–1572, Mar. 1999.
40. M. Daffé and P. Draper, "The envelope layers of mycobacteria with reference to their pathogenicity," *Adv. Microb. Physiol.*, vol. 39, pp. 131–203, 1998.
41. M. Daffé, "The cell envelope of tubercle bacilli," *Tuberculosis (Edinb)*, vol. 95 Suppl 1, pp. S155-158, Jun. 2015.
42. V. Jarlier and H. Nikaido, "Mycobacterial cell wall: structure and role in natural resistance to antibiotics," *FEMS Microbiol. Lett.*, vol. 123, no. 1–2, pp. 11–18, Oct. 1994.
43. K. Takayama, C. Wang, and G. S. Besra, "Pathway to Synthesis and Processing of Mycolic Acids in *Mycobacterium tuberculosis*," *Clin Microbiol Rev*, vol. 18, no. 1, pp. 81–101, Jan. 2005.
44. J. B. Stock, A. J. Ninfa, and A. M. Stock, "Protein phosphorylation and regulation of adaptive responses in bacteria," *Microbiol. Rev.*, vol. 53, no. 4, pp. 450–490, Dec. 1989.
45. V. L. Robinson, D. R. Buckler, and A. M. Stock, "A tale of two components: a novel kinase and a regulatory switch," *Nat. Struct. Biol.*, vol. 7, no. 8, pp. 626–633, Aug. 2000.
46. J. Muñoz-Dorado, S. Inouye, and M. Inouye, "A gene encoding a protein serine/threonine kinase is required for normal development of *M. xanthus*, a gram-negative bacterium," *Cell*, vol. 67, no. 5, pp. 995–1006, Nov. 1991.
47. S. T. Cole *et al.*, "Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence," *Nature*, vol. 396, no. 6707, p. 190, Nov. 1998.
48. Y. Av-Gay and M. Everett, "The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*," *Trends Microbiol.*, vol. 8, no. 5, pp. 238–244, May 2000.
49. S. K. Hanks and T. Hunter, "Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification," *FASEB J.*, vol. 9, no. 8, pp. 576–596, May 1995.

50. S. Cowley *et al.*, “The Mycobacterium tuberculosis protein serine/threonine kinase PknG is linked to cellular glutamate/glutamine levels and is important for growth in vivo,” *Mol. Microbiol.*, vol. 52, no. 6, pp. 1691–1702, Jun. 2004.
51. P. Deol *et al.*, “Role of Mycobacterium tuberculosis Ser/Thr kinase PknF: implications in glucose transport and cell division,” *J. Bacteriol.*, vol. 187, no. 10, pp. 3415–3420, May 2005.
52. C. Ortega *et al.*, “Mycobacterium tuberculosis Ser/Thr Protein Kinase B Mediates an Oxygen-Dependent Replication Switch,” *PLoS Biol*, vol. 12, no. 1, Jan. 2014.
53. L. V. Sacks and R. E. Behrman, “Developing new drugs for the treatment of drug-resistant tuberculosis: a regulatory perspective,” *Tuberculosis*, vol. 88, pp. S93–S100, Aug. 2008.
54. M. Casenghi, S. T. Cole, and C. F. Nathan, “New approaches to filling the gap in tuberculosis drug discovery,” *PLoS Med.*, vol. 4, no. 11, p. e293, Nov. 2007.
55. J. van den Boogaard, G. S. Kibiki, E. R. Kisanga, M. J. Boeree, and R. E. Aarnoutse, “New drugs against tuberculosis: problems, progress, and evaluation of agents in clinical development,” *Antimicrob. Agents Chemother.*, vol. 53, no. 3, pp. 849–862, Mar. 2009.
56. Y. Zhang, “Persistent and dormant tubercle bacilli and latent tuberculosis,” *Front. Biosci.*, vol. 9, pp. 1136–1156, May 2004.
57. H. Tomioka and K. Namba, “[Development of antituberculous drugs: current status and future prospects],” *Kekkaku*, vol. 81, no. 12, pp. 753–774, Dec. 2006.
- A. M. Ginsberg and M. Spigelma
58. A. M. Ginsberg and M. Spigelman, “Challenges in tuberculosis drug research and development,” *Nat. Med.*, vol. 13, no. 3, pp. 290–294, Mar. 2007.

Chapter II:

Computational methods in drug discovery

1. Introduction

Drug discovery plays an important role in the growth of any pharmaceutical company and society, as newer and safer drugs are launched in the market with the sole objective of improving the therapeutic value and safety of drugs. The pharmaceutical industry has consistently shown that it can discover and develop innovative medicines for a wide range of diseases [1].

Drug research, as it is called today, began when chemistry had reached the peak of its career, allowing chemical principles and theories to be applied to problems outside the scope of chemistry, and when pharmacology became an independent scientific discipline on its own. By 1870, some of the important foundations of chemistry theory had been laid [2, 3]. In the twentieth century, biochemistry had remarkable influence on drug research in numerous ways. It was during this period that the concept of targeting enzymes and designing drugs as inhibitors came into existence [4]. However, the current drug-discovery process is very time consuming and expensive and can take up to 12–16 years of exhaustive research, huge financial investment, and clinical trials before a molecule can be recognized as a drug .

Despite the diverse research and development (R&D) approaches adopted by pharmaceutical companies, the attrition rate is inadmissibly high. One of the factors contributing to the high attrition rates is an active compound with unacceptable absorption, distribution, metabolism, excretion, and toxicity (ADMET) adverse effects that thus needs to be withdrawn from development. This factor represents approximately 50% of all costly failures in drug development [6], and it has become widely appreciated that these areas should be considered as early as possible in the drug-discovery process [7, 8]. It is evident that the pitfall in the current drug-discovery process urges an unconventional approach, which would not only truncate the R&D time but also reduce the cost involved [9].

2. Computer-aided drug design

2.1 Introduction

Computer aided drug design (CADD) is a multidisciplinary field attracting the researchers from information technology, medicine, pharmacology etc. Computational tools have become increasingly important in drug discovery and design processes (Figure II.1) Methods from computational chemistry are used routinely to study drug-receptor complexes in atomic detail and to calculate properties of small-molecule drug candidates. Tools from information sciences and statistics are increasingly essential to organize and manage the huge chemical and biological activity databases that all pharmaceutical companies now possess, and to make optimal use of these databases [10].

2.2 CADD applications in drug discovery and development

There are several key areas where CADD supports in designing an effective drug.

Virtual High Throughput Screening (vHTS). vHTS is a method for searching new lead molecules to develop into a promising drug for the selected disease target. In vHTS, small molecules of the drug like compounds stored in the database are screened against the protein targets to find which molecules can bind strongly to the target protein [11]. They are called lead molecules for the particular disease. These lead molecules are then extracted from the database for further testing. With the efficient CADD screening tools available nowadays time and the expenditure required for finding a promising lead molecule is considerably less than traditional methods.

Sequence Analysis. The insight knowledge about the amino acid sequence of protein molecules of various organisms is essential for a design of the successful drug. Many sequence analysis tools and algorithms developed by CADD researchers helps in finding out the similarity among the species based on the proteomic and genomic sequences. This sequence similarity information is useful in assuming the relationships among the various organisms involved in the study.

Homology Modeling. As most of the drug targets are proteins, scrupulous knowledge about the three dimensional structure of those protein molecules is essential during drug design. Very few 3-D structures of protein molecules are available in realism. However 3-D structures of protein molecules can be predicted using CADD techniques. As it is proved that many protein molecules have similar amino acid sequences. If a 3-D structure of a protein molecule is known, this structure is used to predict the 3-D structure

of the protein molecules which have high similarity scores with the matching protein molecules. This process is known as homology modeling. Many database like, SWISS MODEL [12] repository having predicted 3-D protein structures, are created using CADD homology modeling techniques.

Similarity Searches. A common activity during drug discovery is the search for drug analogues. Starting with an existing promising lead molecule of a drug, chemical compounds with similar protein structure (2D or 3D), common amino acid sequences or electrostatic properties etc can be searched using the CADD tools from existing proteomic and genomic databases. These drug analogues can be further tested to bring an improved drug candidate as an alternative for the existing drug.

Physicochemical Modeling: Drug-receptor interactions occur on atomic scales. The physicochemical properties such as hydrophobicity and polarity of the drug and target have an intense effect of how candidate drugs bind to protein targets. As the drug and the target interactions occur on atomic scales the study of bio chemical and biophysical properties of them provide an in depth understanding about these interactions.

Drug Optimization: When a promising candidate drug has been found during the drug discovery process, then the newly discovered drug has to be optimized to increase its affinity and binding towards target protein. This can be carried out by modifying the structure of the drug. Alternate templates or scaffolds like the newly discovered drug are evaluated in this stage to find out a promising drug for the disease target. The metabolic and toxic properties of the candidate drug are optimized to increase the potential of the drug.

ADMET properties of a drug: The key characteristics for drugs are Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET). These properties are called as the bioavailability and bioactivity of the drug. Most of the candidate drugs fail in clinical trials because of the problems of the toxicity and metabolism of the drugs in human beings which makes useless the billions of dollars and years of research spent up to this phase. Even though these properties should be measured in the lab, they can also be predicted in advance using CADD tools which save the years of research and the huge money spent for the experiments on the candidate drug [13,14].

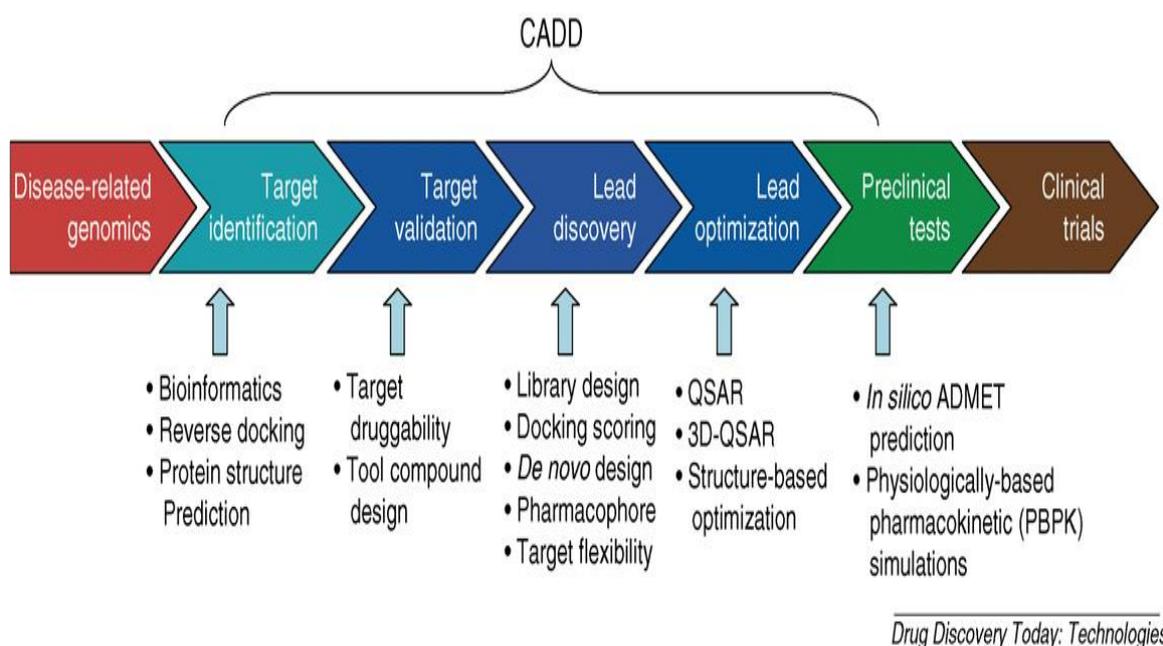


Figure II.1: CADD in drug discovery

2.3 Classification of CADD

CADD can be classified into two general categories: structure-based and ligand-based (Figure II.2).

2.3.1 Ligand-based drug design (LBDD)

Ligand-based drug design, or indirect drug design, relies on knowledge of other known active molecules with the potential against biological targets of interest [15]. Pharmacophore models are derived from these known molecules to define the necessary structural characteristics to enable binding to the biological target [16]. Alternatively, in quantitative structure-activity relationship (QSAR), we derive the correlation between the calculated molecular properties of a compound and their experimentally determined biological activity [17]. These predicated QSAR correlations may in turn be used to predict the activity of novel analogs [18].

2.3.2 Structure-based drug discovery (SBDD)

If the three-dimensional structure of a disease-related drug target is known, the most commonly used CADD techniques are structure-based. In SBDD the therapeutics are designed based on the knowledge of the target structure. Two commonly used methods in SBDD are molecular docking approaches and *de novo* ligand (antagonists, agonists, inhibitors, etc. of a target) design. Molecular dynamics (MD) simulations are frequently used in SBDD to give insights into not only how ligands bind with target proteins but also

the pathways of interaction and to account for target flexibility. This is especially important when drug targets are membrane proteins where membrane permeability is considered to be important for drugs to be useful [19, 20].

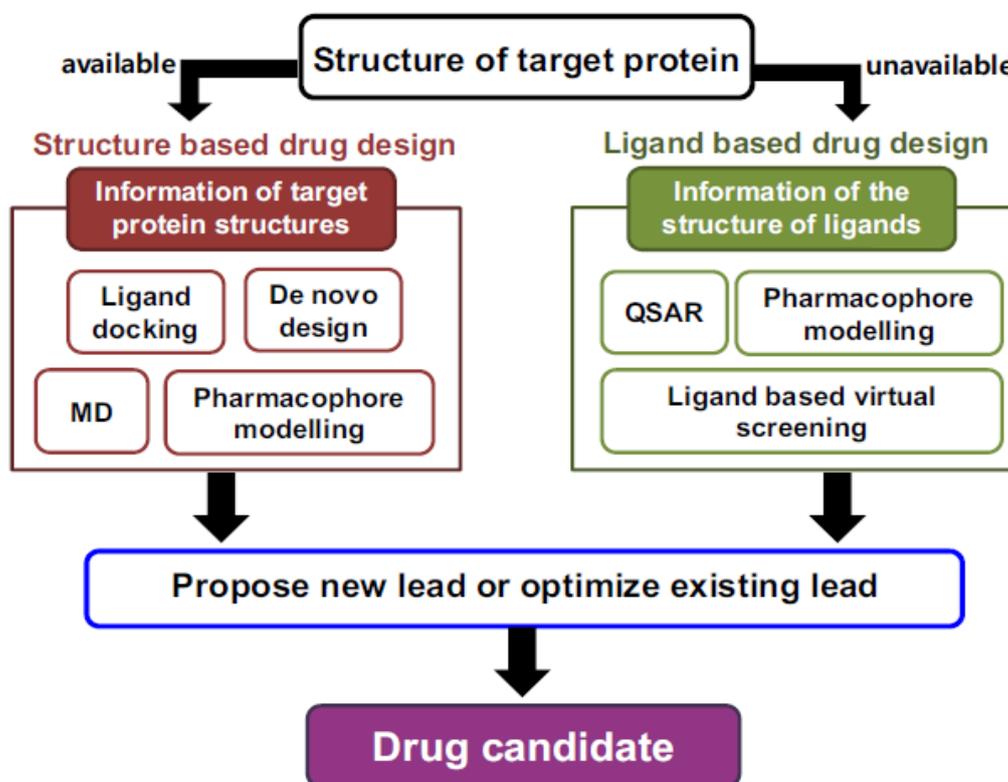


Figure II.2: Flow chart of CADD processes

A biomolecular system can be simulated using molecular mechanics (MM), QM, or a hybrid method (QM/MM), depending on the research problem to be answered.

2.4 Molecular mechanics

MM is commonly applied in large systems to calculate molecular structures and relative potential energies of a molecular conformation or atom arrangement [21-23]. The electrons in the studied system are not explicitly considered, but instead each atom – specifically, the atomic nucleus and the associated electrons – is treated as a single particle. The exclusion of electrons in MM is justified on the basis of Born–Oppenheimer approximation [24], which states that electronic and nuclear motions can be uncoupled from each other and considered separately. Energy differences between conformations are significant in such calculations, rather than absolute values of potential energies.

MM can simply be viewed as a ball-and-spring model of atoms and molecules with classical forces between them [25]. Such forces are accounted by potential energy

functions with respect to such structural features as bond length, bond angles, and torsional angles. Potential energy functions are equipped with parameters designed to reproduce experimental properties [22]. The MM or rather the total potential energy of a molecule is described as the sum of bond-stretching energy (E_{str}), bond angle-bending energy (E_{bend}), torsion energy (E_{tor}), and energy of interactions among unbound atoms (E_{nb}). Energy contributions of the latter constitute van der Waals and electrostatic interactions:

$$E_{tot} = E_{str} + E_{bend} + E_{tor} + E_{vdw} + E_{elec} \quad (1)$$

$$E_{tot} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_{\theta} (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2} (1 - \cos(n\phi - r)) + \sum_{i < j} \left[\frac{A_{ij}}{r_{ij}^{12}} + \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\epsilon r_{ij}} \right] \quad (2)$$

Where E_{tot} is total potential energy, stretch terms refer to E_{str} , bend terms refer to bond angle-bending energy E_{bend} , torsional terms refer to E_{tor} or twisting energy, and unbound interactions are van der Waals forces and electrostatic forces between atoms that are not chemically bonded. Energy contributions from special treatment of hydrogen bonding and stretch–bend coupling interactions may also be seen in MM.

2.5 Quantum mechanics

The QM method treats molecules as collections of nuclei and electrons without any reference to “chemical bonds”. QM is important in understanding the behavior of systems at the atomic level. QM methods apply the laws of QM to approximate the wave function and to solve the Schrödinger equation [21, 26]. The solution to the Schrödinger equation is in terms of the motions of electrons, which in turn lead directly to molecular structure and energy among other observables, as well as to information about bonding. However, the Schrödinger equation cannot actually be solved for any but a one-electron system (the hydrogen atom), and approximations need to be made. According to QM, an electron bound to an atom cannot possess any arbitrary energy or occupy any position in space. These characteristics can be determined by solving the time-independent Schrödinger equation: [27,28].

$$H = T + V \quad (3)$$

Where H is the Hamiltonian operator (sum of kinetic energy), T the potential energy, and V the operator. H can also be defined as:

$$H = \left[-\frac{\hbar^2}{8\pi^2} \sum_i \frac{1}{m_j} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \right] + \sum_i \sum_{<j} \left(\frac{e_i e_j}{r_{ij}} \right) \quad (4)$$

QM methods include ab initio [29] density functional theory (DFT) [30-32] and semi empirical calculations [33-35]. For more accurate QM calculations, electron correlation methods, namely, CCSDT and MP2, etc, are necessary [31]. They also help to identify the activated complex when applied to reacting chemical species and therefore in the identification of a reaction pathway. Since the Schrödinger equation cannot be solved for complex molecular systems, semi empirical ab initio DFT methods were developed to approximate the precise QM solution to the problem [21, 31, 36]. QM models are the most accurate, but also the most expensive methods in terms of time and computational resources, and are thus applied on small systems.

Density functional theory

Density functional theory (DFT) is presently the most successful quantum mechanical modeling method used in physics and chemistry to compute the electronic structure (principally the ground state) of many-body systems, in particular atoms, molecules, and the condensed phases. In chemistry, DFT is used to predict a variety of molecular properties, such as molecular structures, vibrational frequencies, atomization and ionization energies, electric and magnetic properties, reaction paths, etc. The modern DFT calculations are based on two Hohenberg and Kohn theorems, which proves that the electronic energy of a molecule in a ground state could be determined completely by electron density $\rho(\mathbf{r})$ [37]. The electron density $\rho(\mathbf{r})$ can be defined as in Equation (5), where r is spatial variable of electrons and s is the spin variable of electrons.

$$\rho(\mathbf{r}) = N \sum_{s_1} \dots \sum_{s_N} \int dr_2 \dots \int r_N |\Psi(r_1, s_1, r_2, s \dots r_N, s_N)|^2 \int \rho(\mathbf{r}) dr \quad (5)$$

The Kohn-Sham (KS) theories are the most common implementation of DFT, making it widely used. The KS equations are analogous to the Hartree-Fock equations. In the KS model, non-interacting electrons moving in an effective potential is introduced to solve the problem of interacting electrons of many-body moving in a static external potential. The most popular DFT method is the Becke3-Lee-Yand-Parr (B3LYP) hybrid functional, and was also used for the calculations in this thesis. Generally speaking, DFT is not a CADD method; however, it is involved in application in CADD to predict molecular properties.

3. Quantitative structure-activity relationships (QSAR)

3.1 Introduction

Quantitative structure – activity relationship (QSAR) modeling pertains to the construction of predictive models of biological activities as a function of structural and molecular information of a compound library. The concept of QSAR has typically been used for drug discovery and development and has gained wide application for correlating molecular information with not only biological activities but also with other physicochemical properties, which has therefore been termed quantitative structure – property relationship (QSPR). QSAR is widely accepted predictive and diagnostic process used for finding associations between chemical structures and biological activity. QSAR appeared and evolved in trying to respond to the need and desire of the specialist chemist to predict the biological response [38]. It found its way into the practice of agro chemistry, pharmaceutical chemistry, and eventually most facets of chemistry [39].

QSAR is the final result of computational processes that start with a suitable description of molecular structure and ends with some inference, hypothesis, and predictions on the behavior of molecules in environmental, physicochemical and biological system under analysis [40]. The final outputs of QSAR computations are set of mathematical equations relating chemical structure to biological activity [41-43]. Multivariate QSAR analysis employs all the molecular descriptors from various representations of a molecule (1D, 2D and 3D representation) to compute a model, in a search for the best descriptors valid for the property in analysis. QSAR's general mathematical form is represented by the following equation.

QSAR models are not only used for the prediction of properties but are also helpful in selection of alternative mechanism of action, determination of useful structural characteristics, projecting new design methodologies and help in proposing new hypotheses for future research work.

3.2 History of QSAR

Crum-Brown and Fraser expressed the suggestion that the physiological action of a substance was a function of its chemical composition. Later, in 1893, Richet showed that the cytotoxicities of a dissimilar set of uncomplicated organic compounds were inversely related to their corresponding water solubility. After that, Meyer and Overton independently recommended that the narcotic action of a group of organic molecules correlated with their olive oil/water partition coefficients. The extensive work of Albert, and Bell and Roblin established the importance of ionization of bases and weak acids in bacteriostatic activity.

In the physical organic border, great progress was being made in the clarification of substituent effects on organic reactions, led by the influential job of Hammett. Taft invented a way for separating polar, steric, and resonance effects and introducing the first steric parameter, ES.

The contributions of Hammett and Taft together laid the mechanistic source for the progress of the QSAR model by Hansch and Fujita. In 1962 Hansch et al [44] published their bright study on the structure-activity relationships of plant growth regulators and their dependency on Hammett constants and hydrophobicity. A Linear Free Energy Relationships (LFER) related model published by Fujita et al. and Hansch et al. [45], considered to be the official beginning for QSAR.

Their fragment and additive group contribution idea added two things: the use of calculated properties to correlate with biological activities and the detection that multiple properties may influence the biological activity. For this purpose, they implemented the use of the computer to fit QSAR equations.

The so-called Hansch equation [46] was developed to correlate physicochemical properties (descriptors) with biological activities is given in a general form by:

$$\text{Log}I/C = a(\text{log}P)^2 + b(\text{log}P) + c \sigma + \dots \dots \dots K \quad (6)$$

Where C is the molar concentration that produces the biological effect; P is the octanol/water partition coefficient and σ is the electronic Hammett constant.

Besides the Hansch approach, other methodologies were also developed to deal with structure- activity questions. The Free-Wilson approach [47] addresses structure-activity studies in a congeneric series in which the contribution of each structural feature was a

parameter of interest. These parameters, also called indicator variables, codify the presence or absence of particular structural feature. They are assigned the binary values of 1 and 0, accordingly.

3.3 Tools and Techniques of QSAR

3.3.1 Biological Parameters

In QSAR analysis, it is imperative that the biological data be both accurate and precise to develop a meaningful model. It must be realized that any resulting QSAR model that is developed is only as valid statistically as the data that led to its development. The equilibrium constants and rate constants that are used extensively in physical organic chemistry and medicinal chemistry are related to free energy values ΔG . Thus for use in QSAR, standard biological equilibrium constants such as K_i or K_m should be used in QSAR studies.

Likewise only standard rate constants should be deemed appropriate for a QSAR analysis. Percentage activities (e.g., % inhibition of growth at certain concentrations) are not appropriate biological endpoints because of the nonlinear characteristic of dose response relationships. These types of endpoints may be transformed to equi-effective molar doses.

Only equilibrium and rate constants pass muster in terms of the free-energy relationships or influence on QSAR studies. Biological data are usually expressed on a logarithmic scale because of the linear relationship between response and log dose in the mid-region of the log dose-response curve. Inverse logarithms for activity ($\log 1/C$) are used so that higher values are obtained for more effective analogs. Various types of biological data have been used in QSAR analysis. A few common endpoints are outlined in (Table II.1).

Biological data should pertain to an aspect of biological/biochemical function that can be measured. The events could be occurring in enzymes, isolated or bound receptors, in cellular systems, or whole animals [48].

Table II.1: Types of biological data utilized in QSAR analysis

Source of Activity	Biological Parameters
1. Isolated receptors	
Rate constants	Log k_{cat} ; Log K_{uncat} ; Log K
Michaelis-Menten Constants	Log $1/K_m$
Inhibition constants	Log $1/K_i$
Affinity data	pA_1 ; pA_2
2. Cellular systems	
Inhibition constants	Log $1/IC_{50}$
Cross resistance	Log CR
in vitro biological data	Log $1/C$
Mutagenicity states	Log TA_{98}
3. In vivo systems	
Biocentration factor	Log BCF
In vivo reaction rates	Log I (Induction)
Pharmacodynamic	rates Log T ((total clearance)

3.3.2 Molecular Descriptors

Molecular descriptors can be defined as the essential information of a molecule in terms of its physicochemical properties such as constitutional, electronic, hydrophobic, lipophilicity, steric, quantum chemical, and topological descriptors.

3.3.2.1 Constitutional descriptors:

Constitutional descriptors capture properties of the molecules that are related to elements constituting its structure. These descriptors are fast and easy to compute.

Examples of constitutional descriptors include molecular weight, total number of atoms in the molecule and number of atoms of different identity. Also, a number of properties relating to bonds are used including total numbers of single, double, triple or aromatic type bonds as well as number of aromatic rings [49].

3.3.2.2 Electrostatic and Quantum-Chemical Descriptors

Electrostatic descriptors capture information on electronic nature of the molecule. These include descriptors containing information on atomic net and partial charges [50]. Descriptors for highest negative and positive charges are also informative, as well as molecular polarizability. Partial negatively or positively charged solvent-accessible atomic surface areas have also been used as informative electrostatic descriptors for modeling intermolecular hydrogen bonding. Energies of highest occupied and lowest unoccupied molecular orbital from useful quantum chemical descriptors as do the derivative quantities such as absolute hardness [51].

3.3.2.3 Topological descriptors:

The topological descriptors treat the structure of the compound as a graph, with atoms as vertices and covalent bonds as edges. Based on this approach, many indices quantifying molecular connectivity were defined starting with Wiener index, which counts the total number of bonds in shortest paths between all pairs of non-hydrogen atoms. Other topological descriptors include Randic indices x , defined as sum of geometric averages of edge degrees of atoms within paths of given lengths, Balaban's J index and Shultz index. Information about valence electrons can be included in topological descriptors e.g. Kier and Hall indices xv or Galvez topological charge indices [49].

3.3.2.4 Steric descriptors:

Steric descriptors describe size and shape of a molecule. Molecular volume (a sum of the van der Waals volumes), molecular surface area or molar refractivity, as a measure of the size of a molecule, are commonly used in QSAR studies.

3.3.2.5 Hydrophobicity descriptors:

Hydrophobicity descriptor is an important group of descriptors that are widely used in drug design and discovery as they can be applied to modeling both pharmacodynamic (receptor binding) and pharmacokinetic properties (e.g. the uptake and distribution of a xenobiotic relying on partitioning through biological membranes). Partition coefficient ($\log P$), distribution coefficient ($\log D$) and aqueous solubility ($\log S$) are important hydrophobicity/hydrophilicity descriptors [52].

3.3.3 Statistical methods

Statistical methods used in QSAR analysis Statistical methods are an essential component of QSAR work. They help to build models, estimate a model's predictive abilities, and find relationships and correlations among variables and activities. A suitable statistical method coupled with a variable selection method allows analysis of this data in order to establish a QSAR model with the subset of descriptors that are most statistically significant in determining the biological activity. The statistical method can be broadly divided in to two: linear and non-linear method. In statistics a correlation is established between dependent variables (biological activity) and independent variables (physiochemical properties or molecular descriptor). The liner method fits a line between the selected descriptor and activity as compared to non-linear method which fit a curved between the selected descriptor and activity.

The statistical method to build QSAR model is decided based on the type of biological activity data. Following are commonly used statistical methods 1. Principal component analysis (PCA), Cluster analysis (CA), Simple liner regression (SLR) , Multiple liner regression (MLR). Stepwise multiple liner regression (MLR), Principle component regression (PCR) ,Continuum Regression (CR),Partial least squares (PLS) ,Genetic function approximation (GFA) ,Genetic partial least squares (GPLS) ,Logistic regression (LR) , K-Nearest Neighbors classification (KNN) ,Neural Network(NN) , Discriminant analysis (DA) , Decision Trees (DT), Canonical Correlation (CC) [53].

3.3.4 Multiple linear regressions (MLR)

MLR is a mathematical method used to find a linear relationship between the observed response and a number of independent variables (descriptors) as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i \quad i = 1, 2, \dots, n \quad (7)$$

Where y_i is the observed response, $x_{i1}, x_{i2}, \dots, x_{ip}$ are the independent variables for the i th sample, p is the number of variables, n is the number of samples and ε_i is the error of prediction. By estimating the parameters $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ the equation of the

Linear model is:

$$\hat{Y}_1 = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_p x_{ip} \quad (8)$$

Where $b_0, b_1, b_2, \dots, b_p$ are the estimates of the previous parameters and \hat{Y}_1 is the predicted value of the model.

MLR is based on the Orthogonal Least Square (OLS) algorithm that minimizes the sum of squares of the error between the predicted and the observed values $\Sigma(y - \hat{Y})^2$.

The vector of predicted values \hat{Y} is obtained as following:

$$\hat{Y} = \mathbf{b} X \quad (9)$$

Where \mathbf{b} is the vector of estimated parameters $b_0, b_1, b_2, \dots, b_p$ calculated as:

$$\mathbf{b} = (X'X)^{-1}X'Y \quad (10)$$

Where \mathbf{X} and \mathbf{Y} are the matrix of descriptors and the vector of experimental responses, respectively.

MLR modeling is based on the assumption that the errors are a normally distributed random variable with constant variance. The obtained model is optimal when the regression estimators are unbiased, efficient, and consistent with a bias and variance approaching zero when the number of samples tends to the infinity.

The disadvantage of this method is that collinearity between the descriptors highly affects the reliability of the regression coefficient estimates. Thus, reducing the number of included variables by removing those with insignificant coefficients can reduce the risk of multi-collinearity and contribute to enhance the reliability of predictions [54].

3.4 Validation of QSAR Models

After the model equation is obtained, moreover the stability and the goodness of fit of the model, it is also significant to estimate the power and the validity of the model before using it to predict the biological activity. Validity is to establish the reliability and significance of the method for a particular use. Therefore, validation of a QSAR model must be done. There are two validation methods used for a QSAR model: internal and external validation techniques to establish the confidence and strength of the model. In general, QSAR modeling involves a systematic process with multiple steps. These include dataset preparation, molecular descriptors selection and generation, mathematical or statistical models derivation, model training and validation using a training dataset and model testing on a testing dataset.

3.4.1 Internal Validation:

3.4.1.1 Least Squares Fit

The most common internal method of validating the model is least squares fitting. This method of validation is similar to linear regression and is the R^2 (squared correlation coefficient) for the comparison between the predicted and experimental activities. An improved method of determining R^2 is the robust straight line fit, where data points are away from the central data points (essentially data points a specified standard deviation away from the model) are given less weight when calculating the R^2 . An alternative to this method is the removal of outliers (compounds from the training set) from the dataset in an attempt to optimize the QSAR model and is only valid if strict statistical rules are followed. The difference between the R^2 and R^2_{adj} value is less than 0.3 indicates that the number of descriptors involved in the QSAR model is acceptable. The number of descriptors is not acceptable if the difference is more than 0.3 [56].

$$R^2 = \left[\frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[(N\sum X^2 - (\sum X)^2)][N\sum Y^2 - (\sum Y)^2]}} \right]^2 \quad (11)$$

3.4.1.2 Fit of the Model

Fit of the QSAR models can be determined by the methods of chi-squared χ^2 and root-mean squared error (RMSE). These methods are used to decide if the model possesses the predictive quality reflected in the R^2 . The use of RMSE shows the error between the mean of the experimental values and predicted activities. The chi squared value exhibits the difference between the experimental and predicted bioactivities:

$$\chi^2 = \sum_{i=1}^n \left(\frac{Y_i - \hat{Y}_i}{\hat{Y}_i} \right)^2 \quad (12)$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^n \frac{(Y_i - \hat{Y}_m)^2}{n-1}} \quad (13)$$

Where, y and \hat{y} are the experimental and predicted bioactivity for an individual compound in the training set, y_m is the mean of the experimental bioactivities, and n is the number of molecules in the set of data being examined. Large chi-square or RMSE values reflect the model's poor ability to accurately predict the bioactivities even the model is having large R^2 value (≥ 0.7). For good predictive model the chi and RMSE values should be low (< 0.5 and < 0.3 , respectively).

However, excellent values of R^2 , κ^2 and RMSE are not sufficient indicators of model validity. Thus, alternative parameters must be provided to indicate the predictive ability of models. In principle, two reasonable approaches of validation can be envisaged one based on prediction and the other based on the fit of the predictor variables to rearranged response variables.

3.4.1.3 Cross-validation

Cross-validation is one of the most extensively used methods for internal validation.

This involves a leave-one-out (calculation of q^2 LOO) cross-validation [57]. For calculating $R_{CV}^2(Q2)$, each sample in the training set was eliminated once and the activity of the eliminated sample was predicted by using the model developed by the remaining samples. LOO- $R_{CV}^2(Q2)$ calculated according to the below formula:

$$R_{CV}^2(Q2) = 1 - \frac{\sum(Y_{pred}-Y_{obs})^2}{\sum(Y_{pred}-Y_{mean})^2} > 0.5 \quad (14)$$

In Eq. (14), Y_{pred} and Y_{obs} indicate predicted and observed activity values accordingly and Y_{mean} signify mean activity value. A model is considered acceptable when the value of $R_{CV}^2(Q2)$ exceeds 0.5.

The following statistical parameters were calculated to test the validation of developed models; PRESS, SSY, SPRESS and R_{adj}^2 . The following equations are used to calculate above parameters.

$$PRESS = \sum (Y_{obs} - Y_{pred})^2 \quad (15)$$

$$SSY = \sum (Y_{obs} - Y_{mean})^2 \quad (16)$$

$$SPRESS = \sqrt{PRESS/n} \quad (17)$$

$$R_{adj}^2 = 1 - (r^2) [n-1/n-p-1] \quad (18)$$

$$PE = 0.6745 (1-r^2) / \sqrt{n} \quad (19)$$

PRESS is used for predicting sum of squares. To validate a regression model with respect to predictability, PRESS is utilized. The deviation of Y_{obs} from the actual Y value is called the prediction error. The sum of the squared prediction errors is known as PRESS value. The lesser PRESS value shows higher predictability of the model. If PRESS value is lesser than SSY then predictability of model is better than chance and it is statistically significant.

3.4.2 External validation

According to Tropsha et al. [58], for considering the validity of the developed models, in addition to the internal validation, the models should be externally validated using the test set compounds. According to their study, the following criteria could be considered as of acceptable predictability:

$$R_{\text{pred}}^2 = 1 - \frac{\sum(Y_{\text{pred}} - Y_{\text{obs}})^2}{\sum(Y_{\text{pred}} - Y_{\text{mean}(\text{train})})^2} > 0.6 \quad (20)$$

Where R_{pred}^2 is squared correlation coefficient between the experimental and predicted biological activities of dataset compounds (training and test sets).

Y_{pred} and Y_{obs} indicate predicted and observed activity values for the test set and $Y_{\text{mean}(\text{train})}$ indicates mean activity value of the training set.

$$\frac{(R^2 - R_0^2)}{R^2} < 0.1 \text{ or } \frac{(R^2 - R_0'^2)}{R^2} < 0.1 \quad (21)$$

$$|R_0^2 - R_0'^2| < 0.3 \quad (22)$$

Where R^2 is determination coefficient between the experimental and predicted pIC50 values of test set compounds with intercept. R_0^2 and $R_0'^2$ are determination coefficients of predicted versus experimental and experimental versus predicted biological activity values in which their regression line must pass through the origin, respectively.

$$0.85 \leq K \leq 1.15 \quad \text{or} \quad 0.85 \leq K' \leq 1.15 \quad (23)$$

The slope of regression line through the origin which mentioned in the above was indicated by K and K' respectively.

4. Molecular Docking

4.1 Introduction

Docking is the computational determination of binding affinity between a protein structure and a ligand. This method involves proficient sampling of all possible poses of the ligand in the binding pocket of the target protein to ease optimal binding geometry, as measured by the defined scoring functions [59, 60].

Molecular docking protocols can also be defined as a blend of a search algorithm and a scoring function. [61-63]. Many scoring functions and algorithms are currently available. The search algorithm is supposed to provide support and freedom to the protein-ligand coordination to enable accurately and sufficient sampling, including the binding modes. Logically, the search algorithm is supposed to have good speed and effectiveness, while the scoring function must be able to analyze physicochemical properties of molecules and thermodynamics of interaction.

A reliable docking algorithm should exhaustively search all possible binding modes between the ligand and target; however, this is impractical because of the large size of the search space. Therefore, constraints, restraints, and approximations are applied to reduce the dimensionality of the problem in an attempt to locate the global minima as efficiently as possible. Since large conformational space is available to protein structures, partial flexibility (side chain) has recently been incorporated into some docking algorithms, e.g., GLIDE, GOLD, AUTODOCK, FlexX, etc. Genetic algorithms (AUTODOCK, GOLD) and Monte Carlo simulated annealing algorithms (GLIDE) are widely used.

The genetic algorithm is an iterative process that sustains a population of individuals that are candidates of the solutions to the problem being elucidated. However, simulated annealing is an iterative procedure that constantly appraises one candidate solution until it reaches a termination condition [64].

4.2 Search Algorithms

Docking applications can be classified by their search algorithms, which are defined by a set of rules and parameters applied to predict the conformations. When we consider the flexibility of the ligand and/or the receptor docking algorithms can be classified in two large groups: rigid-body and flexible docking. Rigid-body docking method does not take into account the flexibility of neither ligand nor receptor, limiting the specificity and accuracy of results, considering essentially geometrical complementarities

between two molecules. Rigid-body docking simulation even not taking into account flexibility is capable of identifying ligand binding sites for several different proteins. Illustrate structures where rigid-body docking simulations were able to predict the correct position of ligand, when compared to crystallographic structures [65-70]. All rigid-body docking simulations were carried out with ZDOCK [71], and superposition of the best results, evaluated using empirical scoring functions, against crystallographic structures generated RMSD lower than 1.0 Å. This RMSD is calculated between two sets of atomic coordinates, in this case, one for the crystallographic structure (x_c, y_c, z_c) and another for the atomic coordinates obtained from the docking simulations (x_d, y_d, z_d) , the summation is taken over all N atoms being compared, the equation is as follows:

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_{ci} - x_{di})^2 + (y_{ci} - y_{di})^2 + (z_{ci} - z_{di})^2} \quad (24)$$

In docking simulations we expect that the best results generate RMSD values below 1.5 Å, when compared to crystallographic structures.

Rigid-body docking simulation has been employed for virtual-screening initiatives, this method has been used as the fastest way to perform an initial screening of a small molecule database. It has a relatively high accuracy, when compared against crystallographic structures. This accuracy is even higher if we introduced an analysis of the best results using an empirical scoring function for the best results obtained using rigid-body docking simulations. Usually, flexible docking or/and scoring functions have been used for applying a more specific refinement and lead optimization after initial rigid body docking procedure, since these methods demand for computational power and CPU time.

Flexible docking methods can consider several possible conformations of ligand or receptor, as well as for both molecules at the same time, at a higher computational time cost. Docking applications usually make use of one or more of the following methods: fast shape matching (SM), incremental construction (IC), Monte Carlo simulations (MC), distance geometry (DG), evolutionary programming (EP), genetic algorithms (GA), tabu search (TS) and simulated annealing (SA) [72].

4.3 Scoring functions

Scoring function is the most important component in structure based drug design for evaluating the efficacy of ligands binding to their target proteins [73]. The docked poses are ranked and evaluated using scoring functions that approximate the binding free energy of a ligand to a receptor, which is a crucial step to differentiate correct poses from incorrect ones. The scoring functions make various assumptions and simplification in the evaluation of binding free energy for modeled complexes and do not fully consider some of the physical phenomena that are important for molecular recognition, i.e, entropic effects, as the scoring functions must be calculated rapidly during the docking run. Normally the scoring functions are expressed as a sum of separate terms that describe the various contributions to ligand binding. A large number of scoring functions are available, such as force field-based, empirical and knowledge-based scoring functions, which differ in which terms that are included in the expression of the binding free energy. Terms expressing no bonded interactions, including Van der Waals interactions and electrostatic interactions, and solvation effects are commonly included [74].

Reference

1. E. A. Ratti and D. G. Trist, "The continuing evolution of the drug discovery process in the pharmaceutical industry.," *Farmaco*, vol. 56, no. 1–2, pp. 13–19, 2001.
2. J. Drews, "Drug discovery: a historical perspective," *Science*, vol. 287, no. 5460, pp. 1960–1964, Mar. 2000.
3. J. Drews, *In Quest of Tomorrow's Medicines*. New York: Springer-Verlag, 1999.
4. N. U. Meldrum and F. J. W. Roughton, "Carbonic anhydrase. Its preparation and properties," *J Physiol*, vol. 80, no. 2, pp. 113–142, Dec. 1933.
5. J. Kuhlmann, "Drug research: from the idea to the product.," *Int J Clin Pharmacol Ther*, vol. 35, no. 12, pp. 541–552, Dec. 1997.
6. E. N. Bharath, S. N. Manjula, and A. Vijaychand, "In silico drug design: tool for overcoming the innovation deficit in the drug discovery process," *Int J Pharm Pharm Sci*. vol. 3, no. 2, pp. 5, Janry.2011.
7. M. J. Alomar, "Factors affecting the development of adverse drug reactions (Review article)," *Saudi Pharm J*, vol. 22, no. 2, pp. 83–94, Apr. 2014.
8. C. Jochum, and J. Gasteiger Canonical, " numbering and constitutional symmetry," *J Chem Inf Comput Sci.*, vol. 17, no. 12, pp. 113–117, May 1977.
9. A. Baldi , "Computational approaches for drug design and discovery: An overview." *Sys. Rev. Pharm .*, vol. 1, no. 1, pp. 99–105, 2010.
10. J. J. Irwin *et al.*, "Automated docking screens: a feasibility study." *J. Med. Chem.*, vol. 52, no. 18, pp. 5712–5720, Sep. 2009.
11. C. McInnes, "Virtual screening strategies in drug discovery," *Curr Opin Chem Biol*, vol. 11, no. 5, pp. 494–502, Oct. 2007.
12. A. Sundaram, "strategies in computer aided drug design." *Int. J. Pharm. Sci. Res. Rev. ,* vol.43, no. 1, pp. 84-86, March - April 2017.
13. G. Sliwoski, S. Kothiwale, J. Meiler, and E. W. Lowe, "Computational methods in drug discovery," *Pharmacol. Rev.*, vol. 66, no. 1, pp. 334–395, 2014.
14. H. J. Huang *et al.*, "Structure-based and ligand-based drug design for HER 2 receptor," *J. Biomol. Struct. Dyn.*, vol. 28, no. 1, pp. 23–37, Aug. 2010.
15. S. Wold and W. J. Dunn, "Multivariate quantitative structure-activity relationships (QSAR): conditions for their applicability," *J. Chem. Inf. Comput. Sci.*, vol. 23, no. 1, pp. 6–13, Feb. 1983.

16. G. Klebe, U. Abraham, and T. Mietzner, "Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity," *J. Med. Chem.*, vol. 37, no. 24, pp. 4130–4146, Nov. 1994.
17. V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, "Random forest: a classification and regression tool for compound classification and QSAR modeling," *J Chem Inf Comput Sci*, vol. 43, no. 6, pp. 1947–1958, Dec. 2003.
18. Y. Wang, S. A. Shaikh, and E. Tajkhorshid, "Exploring transmembrane diffusion pathways with molecular dynamics," *Physiology (Bethesda)*, vol. 25, no. 3, pp. 142–154, Jun. 2010.
19. S. M. Hanson, S. Newstead, K. J. Swartz, and M. S. P. Sansom, "Capsaicin interaction with TRPV1 channels in a lipid bilayer: molecular dynamics simulation," *Biophys. J.*, vol. 108, no. 6, pp. 1425–1434, Mar. 2015.
20. *Atkins Physical Chemistry" 8TH EDITION*, 8th edition. WH Freeman, 2006.
21. D.W. Rogers. *Computational Chemistry Using the PC*. 3rd ed. Hoboken (NJ): Wiley; 2003.
22. E. G. Lewars, *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*, 2nd ed. Springer Netherlands, 2011.
23. J. C. Tully, "Perspective on 'Zur Quantentheorie der Molekeln,'" *Theor Chem Acc*, vol. 103, no. 3, pp. 173–176, Feb. 2000.
24. M. Hotokka, *Calculation of vibrational frequencies by molecular mechanics*. In: Chalmers JM, Griffiths PR, editors. *Handbook of Vibrational Spectroscopy*. Hoboken (NJ): Wiley; 2006.
25. D. J. Tannor, *Introduction to Quantum Mechanics*, 2007 edition. Sausalito, Calif: University Science Books, 2007.
26. R. Mortimer, *Physical Chemistry*. 3rd ed. Cambridge (MA): Academic Press; 2008.
27. E.G. Lewars, *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. Heidelberg: Springer; 2011.
28. F. Jensen, *Introduction to Computational Chemistry*. 2nd ed. Hoboken (NJ): Wiley; 1999.
29. J. P. Perdew and A. Ruzsinszky, "Fourteen easy lessons in density functional theory," *International Journal of Quantum Chemistry*, vol. 110, no. 15, pp. 2801–2807, Dec. 2010.

30. K. Burke, "Perspective on density functional theory," *The Journal of Chemical Physics*, vol. 136, no. 15, p. 150901, Apr. 2012.
31. S. Grimme, J. Antony, T. Schwabe, and C. Mück-Lichtenfeld, "Density functional theory with dispersion corrections for supramolecular structures, aggregates, and complexes of (bio)organic molecules," *Org. Biomol. Chem.*, vol. 5, no. 5, pp. 741–758, Feb. 2007
32. W. J. Hehre, *A Guide to Molecular Mechanics and Quantum Chemical Calculations*. Irvine, CA: Wavefunction, 2003.
33. M.J.S Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, "AM1: a new general purpose quantum mechanical molecular model," *J. Am. Chem.*, vol. 107, pp. 3902–3909, 1985.
34. M. Zerner, *Reviews of Computational Chemistry*. Weinheim, Germany: Wiley-VCH; 1991.
35. P. Comba, T. W. Hambley, and B. Martin, *Molecular Modeling of Inorganic Compounds*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2009.
36. W. Koch, M. C. Holthausen and M. Kaupp, *Angewandte Chemie*, German Edition, 2001, vol. 113, pp.989-989.
37. G. L. Flynn, "Substituent constants for correlation analysis in chemistry and biology. By Corwin Hansch and Albert Leo. Wiley, 605 Third Ave., New York, NY 10016. 1979. 339 pp. 21 × 28 cm. Price \$24.95," *Journal of Pharmaceutical Sciences*, vol. 69, no. 9, pp. 1109–1109, Sep. 1980.
38. A. F. A. Cros, Université de Strasbourg (1538?-1969), and Faculté de médecine, "Action de l'alcool amylique sur l'organisme," [s.n.], Strasbourg, 1863.
39. L. Eriksson, J. Jaworska, A. P. Worth, M. T. D. Cronin, R. M. McDowell, and P. Gramatica, "Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs," *Environ Health Perspect*, vol. 111, no. 10, pp. 1361–1375, Aug. 2003.
40. A. Golbraikh, M. Shen, Z. Xiao, Y.-D. Xiao, K.-H. Lee, and A. Tropsha, "Rational selection of training and test sets for the development of validated QSAR models," *J. Comput. Aided Mol. Des.*, vol. 17, no. 2–4, pp. 241–253, Apr. 2003.
41. C. Hansch, J. F. Sinclair, and P. R. Sinclair, "Induction of Cytochrome P450 by Barbiturates in Chick Embryo Hepatocytes: A Quantitative Structure-Activity Analysis," *Quantitative Structure-Activity Relationships*, vol. 9, no. 3, pp. 223–226, Jan. 1990.

42. E. B. Wedebye, M. Dybdahl, N. G. Nikolov, S. Ó. Jónsdóttir, and J. R. Niemelä, “QSAR screening of 70,983 REACH substances for genotoxic carcinogenicity, mutagenicity and developmental toxicity in the ChemScreen project,” *Reprod. Toxicol.*, vol. 55, pp. 64–72, Aug. 2015.
43. C. Hansch, P. P. Maloney, T. Fujita, and R. M. Muir, “Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients,” *Nature*, vol. 194, no. 4824, pp. 178–180, Apr. 1962.
44. T. Fujita, J. Iwasa, and C. Hansch, “A New Substituent Constant, π , Derived from Partition Coefficients,” *J. Am. Chem. Soc.*, vol. 86, no. 23, pp. 5175–5180, Dec. 1964.
45. C. Hansch, “Quantitative approach to biochemical structure-activity relationships,” *Acc. Chem. Res.*, vol. 2, no. 8, pp. 232–239, Aug. 1969.
46. S. M. Free and J. W. Wilson, “A mathematical contribution to structure-activity studies,” *J. Med. Chem.*, vol. 7, pp. 395–399, Jul. 1964.
47. Burger’s Medicinal Chemistry & Drug Discovery Vol 5 (6th Edition) By Donald Abraham. .
48. N. S. Sethi, “on Computational Methods in Developing Quantitative Structure-Activity Relationship (QSAR),” *I.J.D.D.D.*, vol. 3, no. 3, pp. 815–838, Sept. 2012.
49. R. S. Mulliken, “Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I,” *J. Chem. Phys.*, vol. 23, no. 10, pp. 1833–1840, Oct. 1955.
50. G. Klopman, “Chemical reactivity and the concept of charge- and frontier-controlled reactions,” *J. Am. Chem. Soc.*, vol. 90, no. 2, pp. 223–234, Jan. 1968.
51. P. Piechota “Development of in silico models for the prediction of toxicity incorporating ADME information ,” Liverpool John moores University, Faculty of Science, 2015
52. N. K. Mahobia, R. D. Patel, N. W. Sheikh, S. K. Singh, A. Mishra, and R. Dhardubey, “Validation Method Used In Quantitative Structure Activity Relationship,” *Der Pharma Chemica*, vol. 2, no. 5, pp. 260–271, 2010.
53. O. Skřehota, “Quantitative structure-property relationship modeling algorithms, challenges and IT solutions,” Rigorózní práce, Masaryk University, Faculty of Informatics, 2010.
54. K. Z. Myint “QSAR methods development, virtual and experimental screening for cannabinoid ligand discovery ,” University of Pittsburgh, Faculty of Medicine , 2012
55. S. Kalyaanamoorthy and Y.-P. P. Chen, “Structure-based drug design to augment hit discovery,” *Drug Discov. Today*, vol. 16, no. 17–18, pp. 831–839, Sep. 2011.

56. A. Golbraikh and A. Tropsha, "Beware of q^2 !," *J. Mol. Graph. Model.*, vol. 20, no. 4, pp. 269–276, Jan. 2002.
57. A. Tropsha, P. Gramatica, V. K. Gombar, "The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models," *QSAR & Combinatorial Science*, vol.22, no. 1, pp. 69–77, Dec.2003.
58. M. K. Gilson and H.-X. Zhou, "Calculation of protein-ligand binding affinities," *Annu Rev Biophys Biomol Struct*, vol. 36, pp. 21–42, 2007.
59. V. Mohan, A. C. Gibbs, M. D. Cummings, E. P. Jaeger, and R. L. DesJarlais, "Docking: successes and challenges," *Curr. Pharm. Des.*, vol. 11, no. 3, pp. 323–333, 2005.
60. D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, "Docking and scoring in virtual screening for drug discovery: methods and applications," *Nat Rev Drug Discov*, vol. 3, no. 11, pp. 935–949, Nov. 2004.
61. C. Bissantz, G. Folkers, and D. Rognan, "Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations," *J. Med. Chem.*, vol. 43, no. 25, pp. 4759–4767, Dec. 2000.
62. B. Islam, C. Sharma, A. Adem, E. Aburawi, and S. Ojha, "Insight into the mechanism of polyphenols on the activity of HMGR by molecular docking," *Drug Des Devel Ther*, vol. 9, pp. 4943–4951, 2015.
63. M. H. Baig *et al.*, "Computer Aided Drug Design: Success and Limitations," *Curr. Pharm. Des.*, vol. 22, no. 5, pp. 572–581, 2016.
64. R. Perozzo *et al.*, "Structural Elucidation of the Specificity of the Antibacterial Agent Triclosan for Malarial Enoyl Acyl Carrier Protein Reductase," *J. Biol. Chem.*, vol. 277, no. 15, pp. 13106–13114, Dec. 2002.
65. J. H. Pereira *et al.*, "Structure of shikimate kinase from *Mycobacterium tuberculosis* reveals the binding of shikimic acid," *Acta Crystallogr. D Biol. Crystallogr.*, vol. 60, no. Pt 12 Pt 2, pp. 2310–2319, Dec. 2004.
66. J. S. Oliveira *et al.*, "Crystallographic and pre-steady-state kinetics studies on binding of NADH to wild-type and isoniazid-resistant enoyl-ACP(CoA) reductase enzymes from *Mycobacterium tuberculosis*," *J. Mol. Biol.*, vol. 359, no. 3, pp. 646–666, Jun. 2006.
67. U. Schulze-Gahmen, H. L. De Bondt, and S. H. Kim, "High-resolution crystal structures of human cyclin-dependent kinase 2 with and without ATP: bound waters and

natural ligand as guides for inhibitor design,” *J. Med. Chem.*, vol. 39, no. 23, pp. 4540–4546, Nov. 1996.

68. J. E. Chrencik *et al.*, “Mechanisms of camptothecin resistance by human topoisomerase I mutations,” *J. Mol. Biol.*, vol. 339, no. 4, pp. 773–784, Jun. 2004.

69. J. E. Chrencik *et al.*, “Mechanisms of camptothecin resistance by human topoisomerase I mutations,” *J. Mol. Biol.*, vol. 339, no. 4, pp. 773–784, Jun. 2004.

70. L. F. S. M. Timmers *et al.*, “Structural studies of human purine nucleoside phosphorylase: towards a new specific empirical scoring function,” *Arch. Biochem. Biophys.*, vol. 479, no. 1, pp. 28–38, Nov. 2008.

71. R. Chen, L. Li, and Z. Weng, “ZDOCK: an initial-stage protein-docking algorithm,” *Proteins*, vol. 52, no. 1, pp. 80–87, Jul. 2003.

72. R. Dias. and W. Filgueira de Azevedo Jr , “Molecular Docking Algorithms,” *Current Drug Targ.*, vol. 9, no. 12, pp. 28–38, 1040-1047, 2008.

73. R. Wang, Y. Lu, and S. Wang, “Comparative evaluation of 11 scoring functions for molecular docking,” *J. Med. Chem.*, vol. 46, no. 12, pp. 2287–2303, Jun. 2003.

74. D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, “Docking and scoring in virtual screening for drug discovery: methods and applications,” *Nat Rev Drug Discov*, vol. 3, no. 11, pp. 935–949, Nov. 2004.

Chapter III:

**Qualitative Structure-Activity
Relationships and 2D-QSAR**

**Modeling of PKnB Inhibition by
Amino-pyrimidine Derivatives**

1. Introduction:

Amino-pyrimidine is one of the most important and well-known heterocycles and constitute a common structural motif in a large number of naturally occurring and biologically active compounds. In fact amino-pyrimidines with substituents either at or at 4th-position are present in many drug-like scaffolds with great chemotherapeutic potential. They are associated with antifungal, pesticide and enzyme inhibitory activities. They inhibit several kinases, such as Bcr-Abl kinase, rho-associated protein kinase and serine/threonine protein kinase PKnB [1,2].

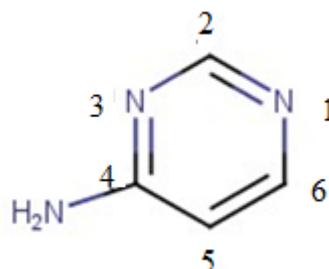


Figure III.1: 2D Structure of amino-pyrimidine

The discovery and development of new anti-TB therapeutics is widely recognized as one of the major global health emergencies, yet it is also a major pharmaceutical challenge. A theoretical technique such as quantum chemical descriptors have been extensively used in Quantitative Structure-Activity Relationship (QSAR) [3,4] to predict the physiological and biological properties of compounds under study. Whether these compounds amino-pyrimidine derivatives provide structural precedence and may lead to the generation of novel anti-TB therapeutics.

The process of drug development is time-consuming and cost-intensive. Several years are required for lead identification, optimization, *in vitro* and *in vivo* testing before starting the first clinical trials. Drug discovery activities are producing ever-larger volumes of complex data that carry significant levels of uncertainty; multi-parameter optimization methods enable this data to be better utilized to quickly target compounds with a good balance of properties, but they all have their strengths and weaknesses [5]. Therefore, we can use the MPO methods to predict the best balance of properties, among these methods we carry out rules of thumb and calculated metrics.

Rules of thumb are the most common approach used to consider the quality of compounds relative to criteria beyond potency that provides guidelines regarding desirable compound characteristics. Several rules have been proposed; the most commonly used are Lipinski and Veber rules [6, 7]. On the other hand, calculated metrics aim to combine the potency with other parameters into a single metric which may be monitored during optimization. The earliest and most commonly applied metrics are the Ligand Efficiency (LE) and the Lipophilic Efficiency (LipE) [5].

2. Materials and methods

2.1 Data Set

In this study, we selected 29 amino-pyrimidine derivatives as Mycobacterium tuberculosis PknB inhibitors were taken from the literature [2, 8, 9]. The activities are reported as pIC₅₀ (pIC₅₀=-log (IC₅₀)), where IC₅₀ is the molar concentration of the compound required for 50% inhibition of inhibitory activity. The compounds and their corresponding biological activity pIC₅₀ values are shown in (Figure III.2) and (Table III.1).

The data set was divided into the two subsets, training set of 22 compounds and test set of 7 compounds. The test set compounds were selected manually considering the distribution of biological data and structural diversity. Training set was used to build a regression model, and the test set was used to evaluate the predictive ability of the model obtained.

2.2 Molecular Descriptors

In the first part of this work, the twenty-nine molecules were pre-optimized using the Molecular Mechanics Force Field (MM+) method included in HyperChem version 8.08 package [10]. The resultant minimized structures were further refined using the semi empirical PM3 Hamiltonian as implemented also in HyperChem. The gradient norm limit of 0.01 kcal/Å was chosen for the geometry optimization.

In the second part, the series of the amino-pyrimidine derivatives was re-optimized by the density functional theory DFT/B3LYP [11-13] at the 6-311G(d,p) basis set of level, by using Gaussian 09 software [14]. Further, the regression analysis was performed using the SPSS version 21 for Windows [15].

The QSAR properties module from HyperChem (8.08) was used to calculate: molar polarizability(Pol), the molar refractivity (MR), partition coefficient octanol/water(log P), molar volume (MV), hydration energy(HE), Surface area grid (SAG) and molar weight (MW). Molinspiration[16] web based software was used to obtain parameter such as PSA, nrotb. The calculated results have been reported in the present work.

The Quantum Chemical descriptors: dipole moment (DM), HOMO-LUMO energy gap (ΔE), total energy (ET) and atomic net charges (Natural bond orbital charges NBO) (qN1, qC2, qN3, qC5 and qN6) were computed using Gaussian 09 software. The 2D structures of amino- pyrimidine derivatives were drawn using MarvinSketch 17.1.2 software[17] (Figure III.1).

2.3 Model validation

The multiple linear regression statistic technique is used to study the relation between one dependent variable and several independent variables. It is a mathematic technique that minimizes differences between actual and predicted values.

Statistical properties of the proposed equation such as the squared correlation coefficient ($R^2 > 0.6$) which is a relative measure of quality of fit. Standard error of estimate (SEE < 0.3) representing absolute measure of quality of fit, Fischer's value (F), is the Fisher ratio, reflects the ratio of the variance explained by the model and the variance due to the error in the regression. High values of the F -test indicate that the model is statistically significant by Dearden *et al.*[18].

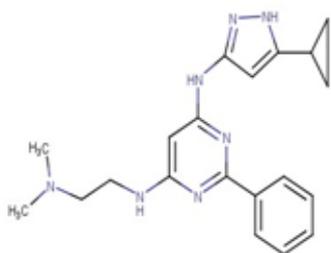
The main objective of a QSAR study is to obtain a model with the highest predictive and generalization abilities. In order to evaluate the predictive power of the QSAR models developed, two principals (internal validation and external validation) were performed (see chapter II). For the internal validation the leave-one-out cross-validation ($R_{CV}^2(Q^2)$) was used to evaluate the stability and the internal capability of the models in the present paper. A high $R_{CV}^2(Q^2)$ value means a high internal predictive power of a QSAR model and a good robustness. Nevertheless, the study of Globarikh [19] indicated that there is no correlation between the value of $R_{CV}^2(Q^2)$ for the training set and predictive ability of the test set, revealing that the $R_{CV}^2(Q^2)$ is still inadequate for a reliable estimate of model's predictive power for all new chemicals.

So, the external validation remains the only way to determine both the generalizability and the true predictive power of QSAR models for new chemicals. For this reason, the statistical external validation was applied to the models as described by Globarikh and Tropsha [19, 20].

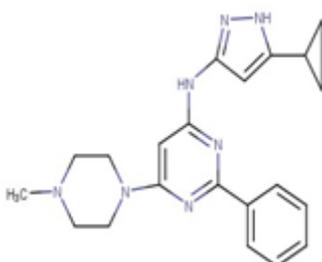
3. Results and discussion

3.1 Structure Activity Relationship (SAR) and drug likeness

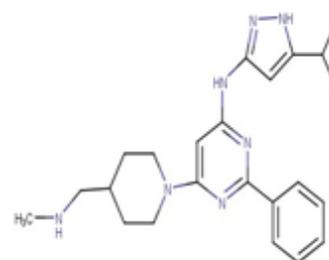
In the first step of our studies, we have studied some physico-chemical proprieties of twenty nine amino- pyrimidine derivatives by HyperChem software, Molecular weight (MW), Molecular volume (MV), Molecular surface (MS), the octanol/water partition coefficient (LogP), hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), polarizability(Pol), refractivity(MR)and hydratation energy(HE). Polar surface area (PSA), number of rotatable bond (NRB) calculated by Molinspiration, ligand efficiency (LE) and Lipophilic efficiency (LipE) are the properties studied in the present work. Results are listed respectively in (Table III.2) and (Table III.1).



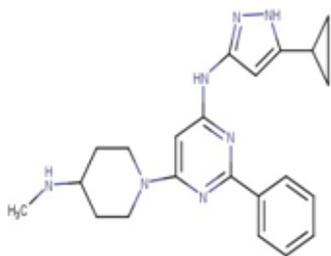
1



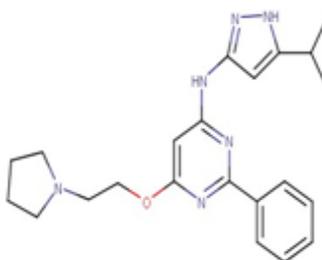
2



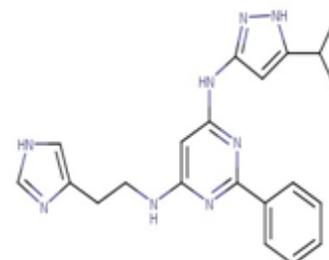
3



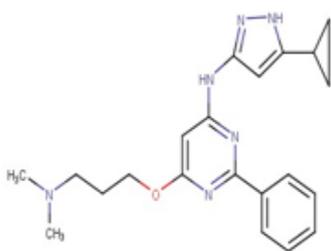
4



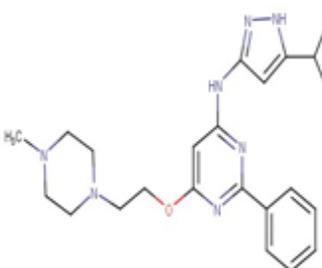
5



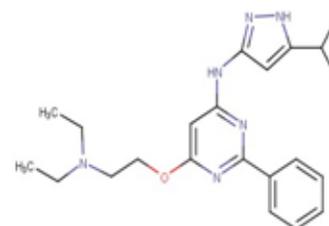
6



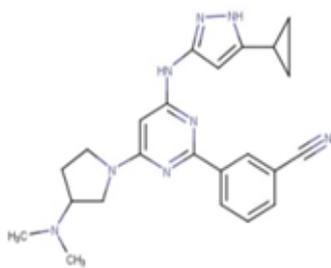
7



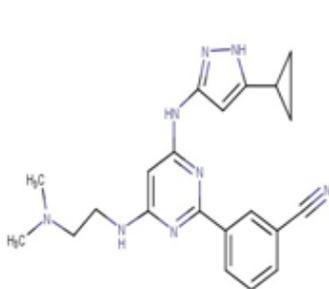
8



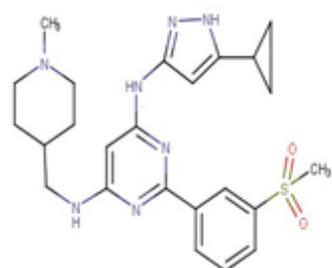
9



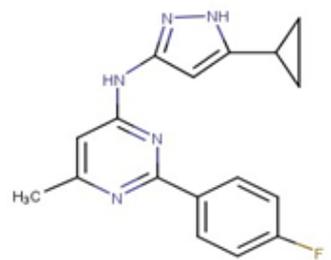
10



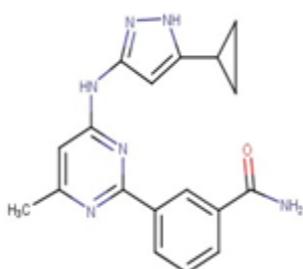
11



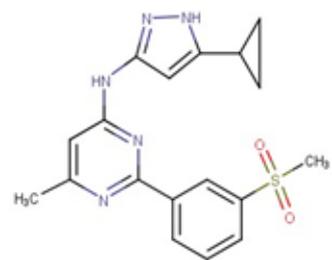
12



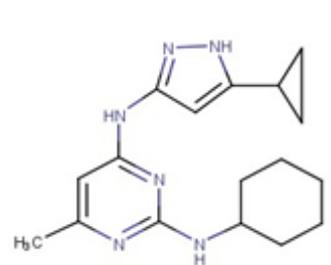
13



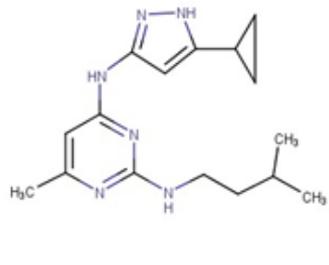
14



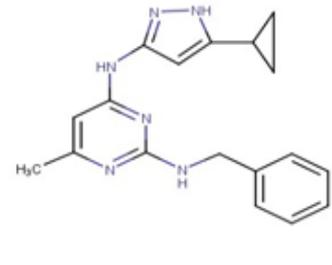
15



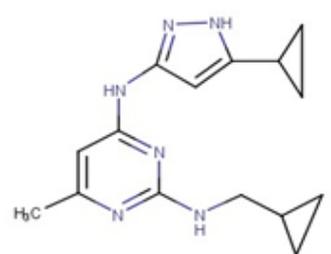
16



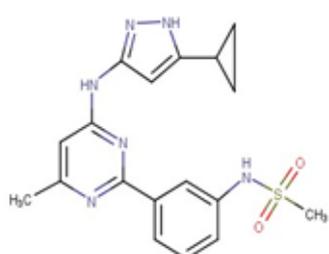
17



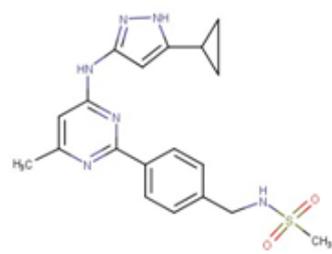
18



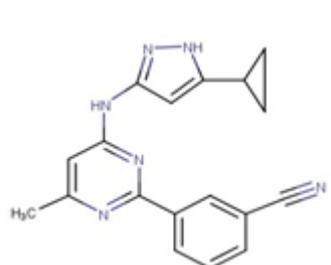
19



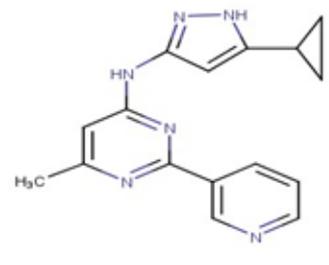
20



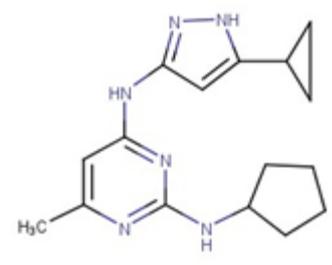
21



22



23



24

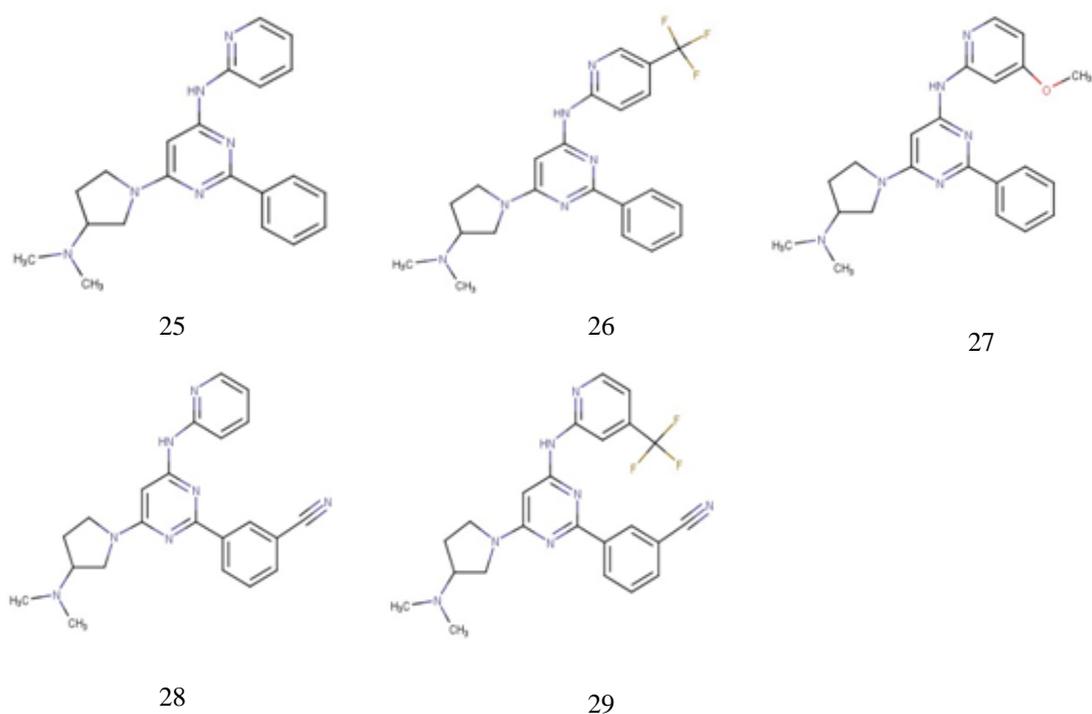


Figure III.2: Structure of amino- pyrimidine derivatives.

These parameters allow ascertaining oral absorption or membrane permeability that occurs when the evaluated molecule follows Lipinski's rule of five.

- (1) There are less than 5 H-bond donors (expressed as the sum of OHs and NHs).
- (2) The molecular weight is under 500 DA.
- (3) The log P is under 5.
- (4) There are less than 10 H-bond acceptors (expressed as the sum of Ns and Os) for an ideal oral bioavailability, there are two other descriptors identified by Veber et al :
 - (1) Rotatable bonds are under 10.
 - (2) Polar surface area is under 140 \AA^2 .

Veber rules suggest that molecular flexibility and polar surface area (PSA) are important determinants of oral bioavailability [7].

Computation of molar refractivity was made via the same method as logP, Ghose and Crippen presented atomic contributions to the refractivity [21].

The solvent-accessible surface bounded molecular volume and van der Waals-surface-bounded molecular volume calculations are based on a grid method derived by Bodor et al [22]. using the atomic radii of Gavezzotti [23].

The hydration energy is a key factor determining the stability of different molecular conformation [24].

Lipophilicity is a property that has a major effect on solubility, absorption, distribution, metabolism, and excretion properties as well as pharmacological activity. Lipophilicity has been studied and applied as an important drug property for decades. It can be quickly measured or calculated. Lipophilicity has been correlated to many other properties, such as bioavailability, storage in tissues, permeability, volume of distribution, toxicity, plasma protein binding and enzyme receptor binding [25,26].

The polarizability of a molecule depends only on its volume; the thermal agitation of the non-polar molecules does not have any influence on the appearance of dipole moments in these molecules. Furthermore, the polar molecules, the polarizability of the molecule does not depend solely on volume but also depends on other factors such as the temperature, because of the presence of the permanent dipole [27]. The polarizability values (Table III.2) are generally proportional to surfaces and of volumes. The decreasing order of polarizability for these studied amino-pyrimidines is: 12,8,10,3, 29,9,4,27,5,28,11,26,6,7,2,25,1,21,22,19,18,14,15,16,22,17,24,13,23.

The order of polarizability is approximately the same one for volume and surface. This is explained by the relation between polarizability and volume, for the relatively non polar molecules.

Molecular weight (MW) is related to the size of the molecule. As molecular size increases, a larger cavity must be formed in water in order to solubilize the compound. Increasing MW reduces the compound concentration at the surface of the intestinal epithelium, thus reducing absorption. Increasing size also impedes passive diffusion through the tightly packed aliphatic side chains of the bilayer membrane. All the studied compounds of amino-pyrimidine derivatives have a molecular weights less than 500 Da (rule number 2), so they are likely soluble and easily pass through cell membranes.

Molecular volume determines transport characteristics of molecules, such as intestinal absorption or blood-brain barrier penetration. Volume is therefore often used in QSAR studies to model molecular properties and biological activity [28].

These compounds have a great variation of distribution volume, in particular compound 12, which has the volume: 1369.02 \AA^3 (Table III.2).

The most important hydration energy in the absolute value, is that of the compound 6 (17.6 kcal/mol) and the weakest is that of compound 26 (4.58 kcal/mol) (Table III.2). Indeed in the biological environments the polar molecules are surrounded by water molecules.

The first correspond to the complex with the strongest hydrogen bond. These hydrated molecules are dehydrated at least partially before and at the time of their interaction. These interactions of weak energy, which we observe in particular between messengers and receivers, are generally reversible.

Compound 12 has nine acceptor sites of proton, and three donor sites. Moreover, compounds 25, 26 have six acceptor sites, and one donor site of proton. The first having higher value, it has three more acceptor sites of proton, and two donor sites (Table III.1). This property supports the first compound, not only by fixing on the receiver, but also by activating. It is thus about an agonist. It has as a consequence, a better distribution in fabrics.

For good oral bioavailability, the LogP must be greater than zero and less than 3 ($0 < \text{Log}P < 3$). For LogP too high, the drug has low solubility and a LogP too low, the drug has difficulty penetrating the lipid membranes [29].

Compounds 20, 15 present the lowest coefficient of division (-0.16, -0.12). When the coefficient of division is rather low, it has as a consequence for a better gastric tolerance and good solubility. In the other hand compounds 16 present the highest value of 3.41, which means it has the capacity to the fixation on plasmatic proteins and a good passage through the biological membrane. The majority of ($\text{Log}P$) of studied molecules have optimal values.

The number of rotatable bonds (NRB) was defined as any single bond, not in a ring, bound to a non-terminal heavy (i.e., non-hydrogen) atom. Excluded from the count were amide C–N bonds because of their high rotational energy barrier [23]. The low number of rotatable bonds (reduced flexibility) in the studied series indicates that these ligands upon binding to a protein change their conformation only slightly. Rotatable bonds are under 10 so all the screened compounds were flexible.

Polar surface area (PSA) is a very useful parameter for prediction of drug transport properties. Polar surface area is defined as a sum of surfaces of polar atoms (usually oxygens, nitrogens and attached hydrogens) in a molecule. This parameter has been shown to correlate very well with the human intestinal absorption, Caco-2 monolayer's permeability and blood-brain barrier penetration [30]. PSA of amino-pyrimidine derivatives were found in the range of 57.18-112.66 and are well below the 140 \AA^2 belong to the compounds with reduced absorption (Table III.1).

In our case, the Lipinski and Veber rules are validated. Therefore, theoretically, all chosen compounds will not have a problem with oral bioavailability.

Table III.1: Drug-likeness parameters of amino-Pyrimidine derivatives.

Compound	Lipinski rules				Veber rules			Ligand efficiency and Lipophilicity efficiency		
	Molecular Mass(amu)	LogP	HBA	HBD	Rules of Five violation	NRB	PSA (Å ²)	pIC50	LE	LipE
1	363.47	1.81	7	3	0	8	81.76	7.06	0.366	5.25
2	375.48	1.99	7	2	0	5	72.97	6.92	0.346	4.93
3	403.53	2.23	7	3	0	7	81.76	7.14	0.333	4.91
4	389.50	1.81	7	3	0	6	81.76	7.28	0.351	5.47
5	390.49	2.45	7	2	0	8	78.97	7.25	0.350	4.80
6	386.46	0.42	8	4	0	8	10.20	6.46	0.311	6.04
7	378.48	2.18	7	2	0	9	78.97	7.26	0.363	5.08
8	419.53	1.93	8	2	0	8	82.20	7.02	0.317	5.09
9	392.50	2.81	7	2	0	10	78.97	7.04	0.339	4.23
10	414.51	1.84	8	2	0	6	96.76	7.25	0.327	5.41
11	388.47	1.53	8	3	0	8	105.55	7.41	0.334	5.88
12	481.62	0.64	9	3	0	8	115.90	7.40	0.304	6.76
13	309.35	2.06	8	2	0	4	66.49	6.43	0.391	4.67
14	334.38	1.18	5	4	0	5	109.59	6.42	0.359	5.24
15	369.44	-0.12	7	2	0	5	100.64	7.07	0.380	7.19
16	312.42	3.41	7	3	0	5	78.52	7.08	0.430	3.64
17	300.41	3.36	6	3	0	7	78.52	6.45	0.410	3.09
18	320.40	2.84	6	3	0	6	78.52	6.15	0.358	3.31
19	330.39	2.72	6	3	0	6	78.52	6.66	0.444	3.94
20	384.46	-0.16	8	3	0	6	112.66	6.41	0.332	6.57
21	398.48	1.54	8	3	0	7	112.66	6.62	0.331	5.08
22	316.37	2.39	6	2	0	4	90.29	7.06	0.411	4.78
23	292.34	1.32	6	2	0	4	79.39	6.69	0.425	5.37
24	298.39	1.83	6	3	0	5	78.52	6.94	0.441	5.11
25	360.46	2.30	6	1	0	5	57.18	5.21	0.270	2.91
26	428.46	3.25	6	1	0	6	57.18	6.28	0.283	3.03
27	390.49	1.31	7	1	0	6	66.41	5.95	0.287	4.64
28	385.47	3.15	7	1	0	5	80.97	5.70	0.275	2.55
29	453.47	2.59	7	1	0	6	80.97	7.40	0.323	4.81

Besides that, we have studied the Ligand efficiency (LE) is a simple metric for assessing whether a ligand derives its potency from optimal fit with the target protein or simply by virtue of making many contacts. It shows generally a dependency on ligand size. Ligand efficiency drops dramatically when the size of the ligand increases.

Ligand efficiency is used in drug discovery programs to assist in narrowing focus to lead compounds with optimal combinations of physicochemical properties and pharmacological properties. It is frequently used to evaluate fragment compounds in fragment-based drug discovery [31].

Lipophilicity efficiency (LipE) provides as straight forward and meaningful way to evaluate the quality of research compounds, linking potency and lipophilicity in an attempt to estimate drug-likeness, attempts to maximize the minimally acceptable lipophilicity per unit of *in vitro* potency or more simply, to improve potency, while maintaining low lipophilicity [32].

Ligand efficiency (LE) and Lipophilicity efficiency (LipE) are defined as follows:

$$LE = 1,4pIC_{50}/NH \quad (1)$$

Where: NH is the number of heavy atoms. So LE decreases with increasing number of heavy atoms [33].

$$LipE = pIC_{50} - \log P \quad (2)$$

High LE prefers compounds that gain to escape the affinity-biased selection and optimization towards larger ligands. The focus should be directed towards the generation of compounds that use their atoms most efficiently. As regards LipE, it prefers compounds that gain a lot of their affinity through directed interactions, thus making the interaction with the receptor more specific. While one can say that LipE describes how efficient a Ligand exploits its lipophilicity, no explicit measure of molecular size is used from the results obtained in (Table III.1) we can say for ligands containing $pIC_{50} > 7$ we are able to penalize compound 8 with lowest LE value (0.317), for ligands containing $7 > pIC_{50} > 6$ we can penalize the compound 26 with lowest LE value (0.283) and for compounds containing $pIC_{50} < 6$ we can penalize compounds 25 and 28 with lowest LE values 0.270 and 0.275 respectively on the other side, Compounds 1,4,6,7,8,10,11,12,14,20,21,23 and 24 reach a LipE of 5.25, 5.47, 6.04, 5.08, 5.09, 5.41, 5.88, 6.76, 5.24, 6.57, 5.08, 5.37 and 5.11 respectively, which are situated in the suggested range of 5–7, this indicates that these compounds were successfully optimized.

We can see through the results in (Table III.1) that compound 12 had the highest LipE value of the data set and was deemed to be the most optimal compound.

3.2 Quantitative Structure-Activity Relationships Studies

In the second step, the several physico- chemical descriptors were used as independent variables (Table III.2) and were correlated with biological activities of amino-pyrimidine derivatives for the generation of QSAR model by multiple linear regressions (MLR) analyzes. The structures of studied compounds are shown in (Figure.III 1).

Table III.2: Values of molecular descriptors used in the regression analysis.

Compound	MW (amu)	SAG (Å ²)	MV (Å ³)	Pol (Å ³)	MR (Å ³)	HE (Kcal/mol)	LogP
1	363.47	676.89	1137.04	41.62	113.02	-8.87	1.81
2	375.48	666.97	1130.58	42.69	116.35	-7.28	1.99
3 ^T	403.53	727.93	1234.55	46.36	125.27	-7.95	2.23
4 ^T	389.50	685.27	1175.67	44.52	120.46	-7.93	1.81
5	390.49	692.41	1182.78	43.81	118.62	-7.51	2.45
6	386.46	648.19	1128.08	43.43	118.95	-17.6	0.42
7	378.48	700.00	1178.93	42.75	115.94	-7.95	2.18
8	419.53	720.61	1250.68	46.99	127.13	-7.72	1.93
9	392.50	721.55	1221.92	44.58	120.57	-7.45	2.81
10	414.51	715.62	1240.43	46.37	125.86	-9.83	1.84
11	388.47	690.78	1175.96	43.48	118.00	-9.81	1.53
12	481.62	791.24	1369.02	49.33	138.94	-12.26	0.64
13	309.35	561.65	917.18	33.33	91.97	-7.37	2.06
14	334.38	580.68	982.60	36.69	99.66	-11.52	1.18
15	369.44	604.52	1035.66	36.39	106.69	-9.22	-0.12
16 ^T	312.42	604.63	996.87	35.34	92.11	-6.44	3.41
17	300.41	593.65	985.85	34.28	89.52	-6.57	3.36
18 ^T	320.40	603.62	999.47	36.60	99.85	-9.38	2.84
19	330.39	617.61	1115.49	37.10	83.04	-7.44	2.72
20	384.46	645.49	1080.42	37.74	109.99	-12.29	-0.16
21	398.48	675.55	1141.48	39.58	112.97	-11.81	1.54
22	316.37	588.60	962.93	35.27	96.82	-13.21	2.39
23 ^T	292.34	552.48	895.67	32.71	88.34	-9.69	1.32
24 ^T	298.39	571.88	943.99	33.51	88.70	-6.85	1.83
25 ^T	360.46	637.86	1092.89	41.92	114.92	-9.00	2.30
26	428.46	682.82	1173.21	43.48	120.33	-4.58	3.25
27	390.49	688.88	1173.45	44.39	121.30	-7.30	1.31
28	385.47	669.46	1153.85	43.77	116.52	-10.49	3.15
29	453.47	694.49	1209.04	45.33	125.12	-9.85	2.59

Table III.2: Continued.

Compound	MD D	ET a.u	ΔE a.u	qN1	qC2	qN3	qC4	qC5	qC6
1	4.74	-1160.695	0.234	-0.581	0.469	-0.558	0.460	-0.406	0.455
2	5.71	-1198.813	0.162	-0.593	0.466	-0.585	0.453	-0.399	0.469
3 ^T	2.58	-1277.461	0.234	-0.588	0.474	-0.592	0.460	-0.396	0.463
4 ^T	1.74	-1238.144	0.241	-0.593	0.475	-0.592	0.449	-0.394	0.463
5	2.50	-1258.010	0.239	-0.604	0.476	-0.577	0.454	-0.375	0.591
6	2.00	-1251.789	0.242	-0.597	0.474	-0.592	0.455	-0.388	0.455
7	3.72	-1219.888	0.237	-0.597	0.472	-0.542	0.451	-0.378	0.593
8	2.07	-1352.677	0.240	-0.608	0.475	-0.578	0.453	-0.374	0.591
9	4.16	-1259.211	0.237	-0.595	0.473	-0.542	0.451	-0.378	0.594
10	6.11	-1330.400	0.242	-0.592	0.467	-0.593	0.459	-0.391	0.460
11	5.90	-1252.970	0.247	-0.583	0.465	-0.592	0.464	-0.399	0.456
12	9.65	-1865.431	0.251	-0.601	0.466	-0.579	0.455	-0.398	0.464
13	3.94	-1031.279	0.244	-0.557	0.468	-0.570	0.458	-0.336	0.282
14	1.45	-1100.770	0.052	-0.557	0.464	-0.564	0.458	-0.332	0.279
15	4.28	-1519.949	0.244	-0.556	0.461	-0.566	0.458	-0.324	0.276
16 ^T	2.68	-991.024	0.251	-0.613	0.648	-0.611	0.464	-0.382	0.288
17	2.83	-952.905	0.252	-0.612	0.645	-0.613	0.464	-0.380	0.290
18 ^T	2.84	-1026.712	0.255	-0.612	0.645	-0.611	0.464	-0.380	0.290
19	2.94	-912.333	0.250	-0.617	0.645	-0.612	0.464	-0.383	0.289
20	2.30	-1575.339	0.253	-0.547	0.460	-0.536	0.460	-0.338	0.279
21	9.70	-1614.662	0.244	-0.550	0.462	-0.532	0.456	-0.337	0.280
22	8.99	-1024.280	0.052	-0.556	0.459	-0.557	0.460	-0.349	0.292
23 ^T	8.39	-948.055	0.245	-0.548	0.453	-0.562	0.454	-0.341	0.291
24 ^T	2.95	-951.687	0.247	-0.620	0.644	-0.612	0.464	-0.382	0.288
25 ^T	3.60	-1143.499	0.229	-0.428	0.362	-0.424	0.462	-0.237	0.464
26	7.75	-1480.623	0.226	-0.576	0.471	-0.549	0.451	-0.387	0.463
27	5.34	-1258.055	0.230	-0.585	0.466	-0.588	0.447	-0.390	0.465
28	5.61	-1235.762	0.241	-0.579	0.470	-0.598	0.454	-0.384	0.463
29	8.48	-1572.897	0.233	-0.580	0.467	-0.578	0.448	-0.400	0.472

3.2.1 Internal validation

The MLR model was developed by applying the selected molecular descriptors (Table III.2) to the “active compounds” of the training data (22 compounds).

Pearson’s correlation matrix has been performed on all descriptors by using SPSS statistics 21 Software. The analysis of the matrix revealed ten descriptors for the development of MLR model.

The correlation between the biological activities and descriptors expressed by the following relation:

$$pIC50 = -50.585 - 0.002SAG + 0.065HE - 0.260 \text{ Log } P + 0.128MD + 0.001ET \\ -49.366qN1 - 11.471qC2 + 14.051qN3 + 107.556qC4 + 11.553qC5 \quad (3)$$

$$N= 22 \quad R=0.986 \quad SEE= 0.110 \quad F = 39.341 \quad RMSE= 0.078 \quad Q = 8.963$$

Where N is the number of compounds (training set).

The model shows a good correlation coefficient (*R*) of 0.986 between descriptors (Log *P*, SAG, HE, MD, ET, qN₁, qC₂, qN₃, qC₄ and qC₅).

In this model the values of qC₄, qN₃, qC₅, ET, MD and HE suggest that the activity increases with the increase of these descriptors values, on the other hand the activity decreases with increasing the values of qN₁, qC₂, log *P* and SAG. Since they have a negative value in this equation (3).

Low value of standard error of estimate (<0.3) indicates the accuracy of the statistical fit. All the values of the *t*-statistic are significant which confirms the significance of each descriptor.

The calculated *F* value for the generated QSAR model exceeds the tabulated *F* value by large margin as desired for a meaningful regression. Furthermore, the calculated *F* Value also determines a confidence limit superior to 95% for this model. The positive value of quality factor (*Q*) for QSAR model suggests its high predictive power and lack of over fitting, low standard deviation of the model demonstrates accuracy of the model.

Table III.3: Experimental and predicted activities pIC₅₀ of the molecules under study.

Compound	pIC ₅₀ Exp.	pIC ₅₀ pred.	Resid.
1	7.06	7.118	-0.058
2	6.92	6.873	0.046
3 ^T	7.14	6.865	0.274
4 ^T	7.28	7.240	0.040
5	7.25	7.165	0.084
6	6.46	6.491	-0.031
7	7.26	7.204	0.055
8	7.02	7.216	-0.196
9	7.04	6.951	0.088
10	7.25	7.182	0.067
11	7.41	7.372	0.037
12	7.40	7.402	-0.002
13	6.43	6.618	-0.188
14	6.42	6.354	0.064
15	7.07	6.951	0.118
16 ^T	7.08	6.327	0.752
17	6.45	6.389	0.061
18 ^T	6.15	6.304	-0.154
19	6.66	6.712	-0.053
20	6.41	6.432	-0.022
21	6.62	6.644	-0.025
22	7.06	7.047	0.013
23 ^T	6.69	6.649	0.041
24 ^T	6.94	7.225	-0.285
25 ^T	5.21	5.147	0.062
26	6.28	6.313	-0.033
27	5.95	5.941	0.009
28	5.70	5.683	0.016
29	7.40	7.451	-0.051

T: Test Set

The QSAR model expressed by (Eq.3) were cross validated by the high value of $R_{CV}^2 = 0.973$ obtained by leave one out (LOO). The value of $R_{CV}^2 (Q^2)$ is indeed greater than 0.5 which is the essential condition to qualify a QSAR model as valid [19], and RMSE of cross-validation was 0.078, which indicates reliability of the proposed model.

The developed model were validated by calculation of the following statistical parameters: predicted residual sum of squares (PRESS), total sum of squares deviation (SSY), the predictive error of the coefficient of correlation (PE) and cross-validated correlation coefficients (R_{adj}^2 and R_{cv}^2) (Table III.4).

PRESS is an important cross-validation parameter as it is a good approximation of the real predictive error of the models. Its value being less than SSY points out that the model has a good predictive power and can be considered statically significant. The smaller PRESS value means the better model predictability.

Table III.4: Cross-validation parameters.

Model	PRESS	SSY	PRESS/SSY	S_{PRESS}	R_{cv}^2	R_{adj}^2	6PE
	0.135	4.963	0.027	0.077	0.973	0.948	0.023

Furthermore, for reasonable QSAR model, the PRESS/SSY ratio should be lower than 0.4. [35]. The data presented in (Table III.4) indicate that for the developed model this ratio is 0.027. The high value of R_{CV}^2 and R_{adj}^2 are essential criteria for the best qualification of the QSAR model.

The predictive error of the coefficient of correlation (PE) is yet another parameter used to evaluate the predictive power of the proposed models. We have calculated the PE value of the proposed model and they are reported in (Table III.4). For the developed model the condition $R = 0.986 > 6PE$ is confirmed and hence, they could be considered as good predictive power.

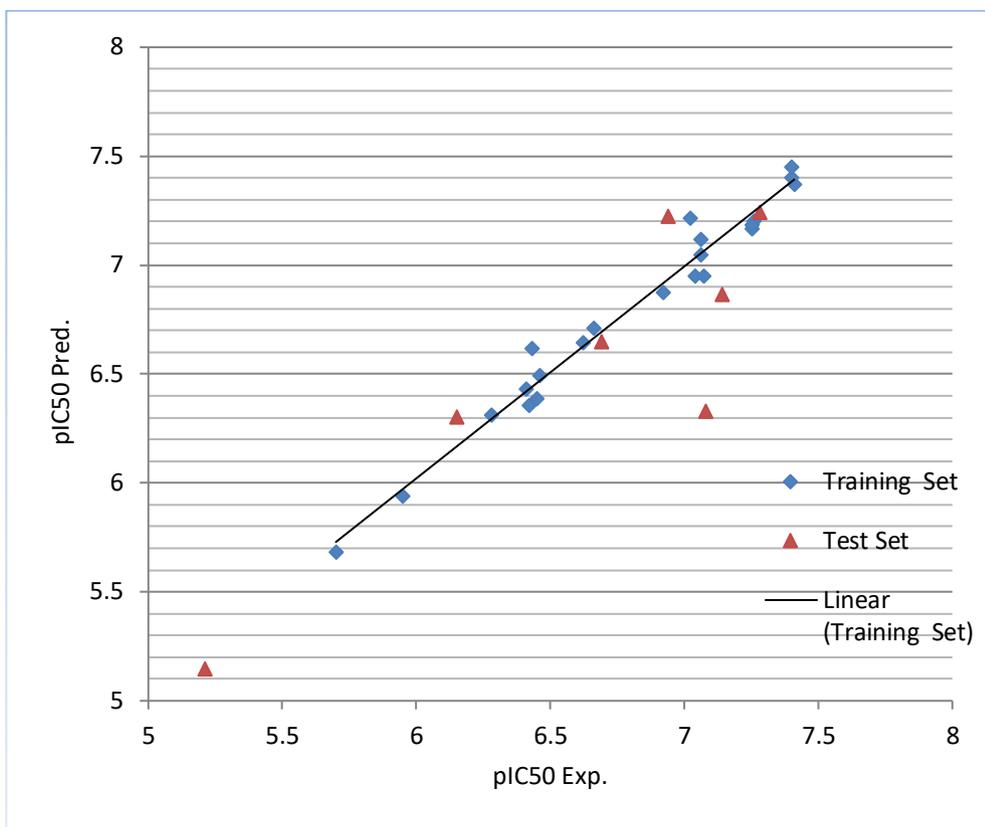


Figure III.3: Predicted plot versus experimentally observed anti-tubercular activity

The plots of the predicted pIC50 versus the experimental pIC50, obtained by the MLR modeling, are demonstrated in (Figure III.3). The plots for this model show to be more convenient with $R^2 = 0.973$.

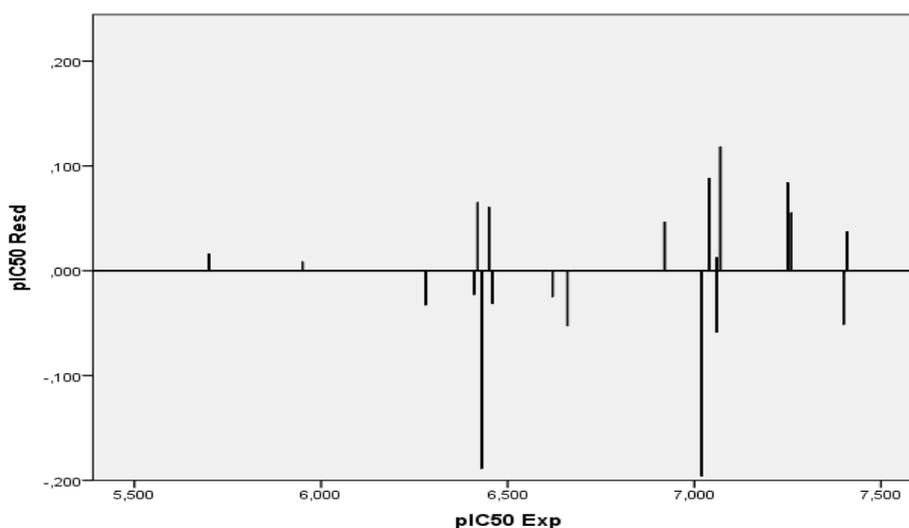


Figure III.4: Plots of the residual against the experimental values of pIC50.

To investigate the presence of a systematic error in developing the QSAR models, the residuals of predicted values of the biological activity pIC50 was plotted against the experimental values (Table III.3), as shown in (Figure III.4). The propagation of the residuals on both sides of zero indicates that no systematic error exists, as suggested by Jalali-Heravi and Kyani [36]. It indicates that this model can be successfully applied to predict the anti-tubercular activity of this class of molecules.

3.2.2 External Validation

The model (Eq. 3) also passed successfully Tropsha's [20] recommended tests for predictive ability:

$$R_{pred}^2 = 0.778 > 0.6$$

$$\frac{(R^2 - R_0^2)}{R^2} = -0.225 < 0.1$$

$$\frac{(R^2 - R_0'^2)}{R^2} = -0.218 < 0.1$$

$$k = 1.014 \quad k' = 0.983 \quad (0.85 \leq k \leq 1.15)$$

$$|R_0^2 - R_0'^2| = 0.006 < 0.3$$

The predicted values for the test set are given in (Table III.3) (Figure III.3). These results might be considered as an indicator of good external predictability.

The built model produced good results for the training set and the test set. It is noteworthy, that the MLR equation has acceptable quality and can predict the activity of training and test set with $R_{CV}^2 = 0.973$ and $R_{pred}^2 = 0.778$, respectively. The plots of the predicted pIC50 versus the experimental pIC50, obtained by the MLR modeling, are demonstrated in (Figure III.3). Therefore, we conclude that the antitubercular activity is related to the physicochemical molecular descriptors and Quantum descriptors.

4. Conclusion

The qualitative study of the relation structure-properties was applied on the series of amino-pyrimidine has pharmacological activities.

Compound 12 having higher value of acceptor sites of proton and LipE of the data set and was deemed to be the most optimal compound. The application of Lipinski and Veber rules on the studied amino-pyrimidine derivatives shows that all these compounds, theoretically, will not have problems with oral bioavailability.

The QSAR analysis was conducted with a series of 29 amino-pyrimidine as mycobacterium tuberculosis (PknB) inhibitors, the model depending on the (eq.3). MLR is the best produced model with very good statistical fit as evident, $R = 0.986$ and $F=39.341$. The physicochemical molecular and Quantum descriptors were found to have a key role in governing the change in biological activity. The model was validated using LOO cross-validation, and external test set. It indicates the model can be successfully applied to predict the anti-tubercular activity of these classes of molecules.

References

1. N. Singh *et al.*, “Synthesis, molecular modeling and bio-evaluation of cycloalkyl fused 2-aminopyrimidines as antitubercular and antidiabetic agents,” *Bioorg Med Chem Lett*, vol. 21, no. 15, pp. 4404–4408, 2011.
2. T. M. Chapman, N. Bouloc, R. S. Buxton, J. Chugh, K.E.A. Lougheed, S.A. Osborne, B. Saxty, S. J. Smerdon, D. L. Taylor and D. Whalley, “Substituted aminopyrimidine protein kinase B (PknB) inhibitors show activity against Mycobacterium tuberculosis,” *Bioorg Med Chem Lett*, vol. 22, no. 9, pp. 3349–3353, May 2012.
3. M. C. Raviglione and I. M. Smith, “XDR tuberculosis--implications for global public health,” *N. Engl. J. Med.*, vol. 356, no. 7, pp. 656–659, Feb. 2007.
4. S. Belaidi, T. Salah, N. Melkemi, L. Sinha, O. Prasad, ” Structure Activity Relationship and Quantitative Structure-Activity Relationships Modeling of Antitrypanosomal Activities of AlkyldiamineCryptolepine Derivatives, ” *J. Comput. Theor Nanosci.*, vol. 12, pp. 1–7, 2015.
5. M. D. Segall, “Multi-parameter optimization: identifying high quality compounds with a balance of properties,” *Curr. Pharm. Des.*, vol. 18, no. 9, pp. 1292–1310, 2012.
6. C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, *J. Adv. Drug Deliv. Rev.* 64, 4 (2012).
7. D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward, and K. D. Kopple, *J. Med. Chem.* 45, 2615 (2002).
8. K. E.A. Lougheed, S. A. Osborne, B. Saxty, D. Whalley, T. M. Chapman, N. Bouloc, J. Chugh, T. J. Nott, D. Patel, V. L. Spivey, C. A. Kettleborough, J. S. Bryans, D. L. Taylor, S. J. Smerdon, and R. S. Buxton, “Effective inhibitors of the essential kinase PknB and their potential as anti-mycobacterial agents,” *Tuberculosis (Edinb)*, vol. 91, no. 4, pp. 277–286, Jul. 2011.
9. M. V. Damre, R. P. Gangwal, G. V. Dhoke, M. Lalit, D. Sharma, K. Khandelwal, and A. T. Sangamwar, “3D-QSAR and molecular docking studies of amino-pyrimidine derivatives as PknB inhibitors,” *J. Taiwan Inst. Chem. Eng.*, vol. 45, no. 2, pp. 354–364, 2014.
10. HyperChem (Molecular Modeling System) Hypercube (2008) Inc., 1115 NW, 4th Street, Gainesville, FL 32601, USA.

11. S. Belaidi, Z. Almi, D. Bouzidi, "Electronic Structure and Physical-Chemistry Properties Relationship for Phenothiazine Derivatives by Quantum Chemical Calculations," *J Comput Theor Nanosci.*, vol. 11, no. 12, pp. 2481-2488, Dec. 2014.
12. S. Khamouli, S. Belaidi, Z. Almi, S. Medjahed, H. Belaidi, "Property/Activity Relationships and Drug Likeness for Pyrimidine Derivatives as Serine/Threonine Protein Kinase B Inhibitors," *J. Bionanosci.*, vol. 11, no. 4, pp. 301-309, Aug. 2017.
13. K. Dermeche, N. Tchouar, S. Belaidi, T. Salah, "Qualitative Structure-Activity Relationships and 2D-QSAR Modeling of TNF- Inhibition by Thalidomide Derivatives," *J. Bionanosci.*, vol. 9, no. 5, pp. 395-400, Oct. 2015.
14. Gaussian 09, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, 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, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian 09, Revision A.02, Gaussian, Inc., Wallingford CT (2009).
15. SPSS software packages, SPSS Inc., 444 North Michigan Avenue, Suite 3000, Chicago, Illinois, 60611, USA.
16. Database, [<http://www.molinspiration.com>].
17. MarvinSketch 17.1.2, Chemaxon (2017), (<http://www.chemaxon.com>).
18. J. C. Dearden, M. T. D. Cronin, and K. L. E. Kaiser, "How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR)," *SAR. QSAR. Environ.*, vol. 20, no. 3-4, pp. 241-266, Apr. 2009.
19. A. Golbraikh and A. Tropsha, "Beware of q^2 !," *J. Mol. Graph. Model.*, vol. 20, no. 4, pp. 269-276, Jan. 2002.
20. P. Gramatica, "Principles of QSAR models validation: internal and external," *QSAR. Comb. Sci.*, vol. 26, no. 5, pp. 694-701, 2007.

21. A. K. Ghose and G. M. Crippen, "Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions," *J. Chem. Inf. Comput. Sci.*, vol. 27, no. 1, pp. 21–35, Feb. 1987.
22. N. Bodor, Z. Gabanyi, and C. K. Wong, "A new method for the estimation of partition coefficient," *J. Am. Chem. Soc.*, vol. 111, no. 11, pp. 3783–3786, May 1989.
23. A. Gavezzotti, "The calculation of molecular volumes and the use of volume analysis in the investigation of structured media and of solid-state organic reactivity," *J. Am. Chem. Soc.*, vol. 105, no. 16, pp. 5220–5225, Aug. 1983.
24. T. Ooi, M. Oobatake, G. Némethy, and H. A. Scheraga, "Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides.," *Proc Natl Acad Sci U S A*, vol. 84, no. 10, pp. 3086–3090, May 1987.
25. L. Di and E. H. Kerns, *Drug-like Properties: Concepts, Structure Design and Methods: from ADME to Toxicity Optimization*, 1 edition. Amsterdam ; Boston: Academic Press, 2008.
26. V. Pliska, B. Testa, H. van de Waterbeemd, R. Mannhold, H. Kubinyi, and H. Timmerman, *Lipophilicity in Drug Action and Toxicology*, Wiley-VCH, Weinheim, Germany (1996).
27. B. M. IÅvorskii and A. A. Detlaf, *Handbook of physics*, 3rd ed. Moscow : Mir ; [London] : Distributed by Central, 1980.
28. L. B. Kier, *Molecular Orbital Theory in Drug Research*, Academic Press, New York (1981).
29. S. Fahn, "Systemic therapy of dystonia," *Can J Neurol Sci*, vol. 14, no. 3 Suppl, pp. 528–532, Aug. 1987.
30. P. Ertl, B. Rohde, and P. Selzer, "Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties," *J. Med. Chem.*, vol. 43, no. 20, pp. 3714–3717, Oct. 2000.
31. J. Dundas, Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz, and J. Liang, "CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues," *Nucleic Acids Res.*, vol. 34, no. Web Server issue, pp. W116-118, Jul. 2006
32. P. D. Leeson and B. Springthorpe, "The influence of drug-like concepts on decision-making in medicinal chemistry," *Nat Rev Drug Discov*, vol. 6, no. 11, pp. 881–890, 2007.

33. C. H. Reynolds, S. D. Bembenek, and B. A. Tounge, "The role of molecular size in ligand efficiency," *Bioorg Med Chem Lett.*, vol. 17, no. 15, pp. 4258–4261, Aug. 2007.
34. K. J. Sanmati, J. Rahul, S. Lokesh, K. Y. Arvind, "QSAR Analysis on 3,5-disubstituted-4,5-dihydropyrazole-1-carbothioamides as epidermal growth factor receptor (EGFR) kinase inhibitors," *J Chem Pharm Res*, vol. 4, no. 6, pp. 3215-3223, Jan. 2012.
35. S. O. Podunavac-Kuzmanović, D. D. Cvetković, and D. J. Barna, "QSAR Analysis of 2-Amino or 2-Methyl-1-Substituted Benzimidazoles Against *Pseudomonas aeruginosa*," *Int J Mol Sci*, vol. 10, no. 4, pp. 1670–1682, Apr. 2009.
36. M. Jalali-Heravi, A. Kyani, "Use of computer-assisted methods for the modeling of the retention time of a variety of volatile organic compounds: a PCA-MLR-ANN approach," *J. Chem. Inform. Comput. Sci.*, vol. 44, no. 4, pp. 1328-1335, Jun. 2004.

Chapter IV:

**Molecular docking and in silico
ADMET study of amino-pyrimidine
Derivatives as mycobacterium
tuberculosis PKnB inhibitors**

1 Introduction

PknB is one of the most important serine/threonine protein kinases for *Mycobacterium tuberculosis* (TB) [1,2]. The intracellular domain of PknB has the main activity of the holoenzyme, and is able to autophosphorylate and combine with ATP and its analogues [3-6]. PknB plays an important part in the growth of TB, and is necessary for survival of TB [7-9]. Change of expression index or phosphorylation of PknB can give to alteration of growth rate and morphology of TB, due to the defects in cell wall synthesis and cell division [10-12]. Because of the differences between PknB and the human protein kinases, it is widely accepted as the drug target for anti-TB. To date, several PknB inhibitors have been reported, and some of them have shown certain degree of anti-TB capability. Most of these compounds are aminopyrimidines, aminoguanidines and anthraquinones [13-19].

Computer aided method is a first approach to screening novel therapeutic agents and the discipline is an emerging strategy as it reduces many complexities of drug design process.

The study of receptor-ligand interaction is a fundamental concept of rational drug design and the prediction of such interactions by molecular docking has increasing importance in the field of structure based drug discovery [20]. The screening of lead molecule with good therapeutic properties and drug likeness is a tedious task in drug discovery process. Computer aided method is an easy platform to search such kinds of biologically active compounds with favorable ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity).

Molecular modeling of the amino-pyrimidine compounds based on their interactions with *Mycobacterium tuberculosis* PknB, using Molecular docking to determine the best ligands conformation when bound to the active site [21].

Finally in silico, ADMET studies were performed on the best ligands to compare the computed ADMET descriptor values with the accepted ranges.

2 Material and methods

2.1 Protein Preparation Structure

The downloading of serine/threonine protein kinases for Mycobacterium tuberculosis PknB was made from the data base Brookhaven Protein Data Bank (www.rcsb.org/pdb) [22] (access code 2FUM)[23] .

It is co-crystallized with inhibitors Mitoxantrone, 1,4-Dihydroxy-5,8-bis({2-[(2-Hydroxyethyl)Amino]Ethyl}Amino)Anthra-9,10-Quinone(three-letter code: MIX). All the heteroatom's and coordinates are removed from the PDB file.

The three-dimensional structure of 2FUM was obtained by X-ray diffraction in complex with a selective inhibitor Mitoxantrone with EC Number: 2.7.11.1 chains (A,B,C,D)), resolution 2.89 Å, and R-value 0.218. In this study we have taken a chain A, 263 residues and 1994 atoms. This allowed us to obtain the model shown in (Figure IV.1).

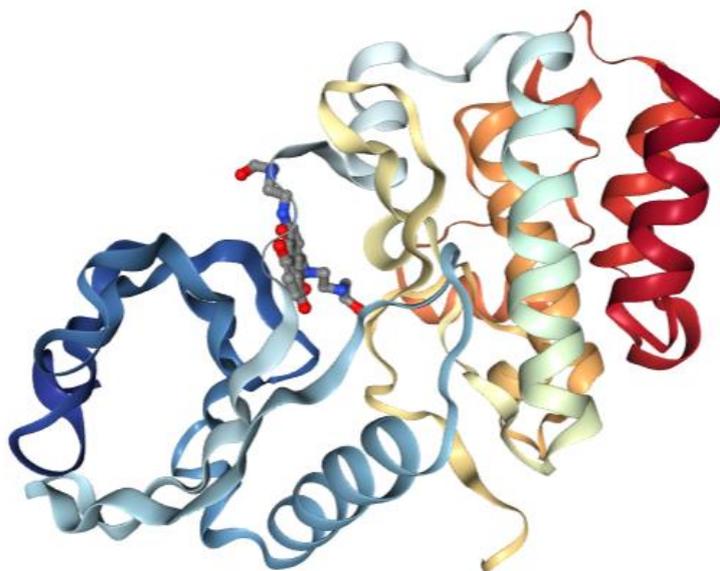


Figure IV.1: Three-dimensional crystal structure of the target protein

Ramachandran Plot

In this study, the stereo chemical quality of the predicted best model was validated using Rampage [24].

2.2 Ligands structure optimization

Screened of 29 amino-pyrimidine derivatives as Mycobacterium tuberculosis PknB inhibitors was selected from the literature [13,14,25], were optimized before docking using MM+ force-field (rms= 0.01 Kcal/Å) and the semi-empirical PM3 method, both of which are implemented in HyperChem 8.08 software [26]. The resulting structures were saved in “.mol” file formats for molecular docking studies. The ligand structures of 29 amino-pyrimidine derivatives (table 1) were made by Marvin Sketch17.1.2 software [27].

2.3 Molecular Docking

Molegro Virtual Docker (MVD2012) software [28] is advanced docking analysis software used to predict protein-ligand interactions. The potential binding site of the target protein and lead candidates are identified by a molecular docking algorithm called Mol Dock.

MVD works on the basis of MolDock SE search algorithm. The docking algorithm was set at a maximum iteration of 1500 with a simple evolution size of 50 and minimum of 5 runs. The population size was set at 50 with energy threshold of 100 at each step. The least minute was set as 10 minutes, the torsions/translations/rotations of the ligand-protein interaction were tested and the one giving lower energies is chosen for further studies. The bond flexibility of the ligands was fixed, and the side chain flexibility of the amino acids in the binding cavity was set with a tolerance of 1.10 and strength of 0.90 for docking simulations. Root-mean-square deviation threshold for multiple cluster poses was set at <2.00 [29].

Different docking programs available and they differ in the nature of the sampling algorithms they employ, in their manner of handling ligand and protein flexibility, in the scoring functions and in the CPU time they required. In the studies reported here, MVD was used, because it showed higher docking accuracy when benchmarked against other available docking programs. (MD: 87%, Glide: 82%, Surflex: 75%, FlexX: 58%) have been shown to be successful in several recent studies, but also for reasons of cost and user friendliness.

Binding affinities were estimated using Molegro data Modeler The scoring function used by MolDock is derived from the piecewise linear potential scoring functions which further improves this score with a new H-bonding term and new charge schemes ,being flexible,i.e. all non-ring torsions were allowed [30].

This molecular docking protocol generate five best predicted poses for each amino-pyrimidine compound with Moldock score, Rerank score, and Hydrogen bond score. The docked conformations or pose with the minimum MolDock score values is the optimal pose.

The docked conformation further analyzed on the basis of the Re-Rank score function. The Re-Ranking score function is generally more reliable than the MolDock scorefunction at selecting the best solution among multiple solutions derived from the same ligand [31].

Ligplot plus 1.4.5 [32], a program to generate schematic diagrams of protein ligand interactions.

2.4 Drug Likeness and Pharmacokinetic Properties of selected ligands

Physicochemical properties of interest included predicted lipophilicity ($\log P$), predicted aqueous solubility ($\log S$), topological polar surface area (TPSA), molecular weight (MW), hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), value predicted by Marvin Sketch. The calculation of partition coefficient was conducted by applying both consensus and ChemAxon methods as implemented in MarvinSketch.

The ligands were predicted for drug likeliness, ADMET (adsorption, distribution, metabolism, excretion and toxicity studies by PreADMET [33].The main filters used for prediction of drug likeliness were Lipinski's rule of five [34], CMC [35], MDDR [34], Lead-like Rule [36], and WDI (World drug index)-like rule[37].These rules were scrutinized and were subjected to ADME prediction.

The pharmacokinetic parameters, absorption and distribution, were considered for selection of compounds as drug candidates. In this study, the PreADMET program was used to predict ADMET of amino-pyrimidine derivatives. The aspect prediction of absorption properties included percentage human intestinal absorption (% HIA) and Caco2 (heterogeneous human epithelial colorectaladeno carcinoma), and MDCK, and MDCK (Madin-Darby canine kidney) cell permeability [38]. And blood brain barrier prediction[39].Those molecules qualified the above rules based on specific statistical cut-off available for each model was selected for toxicity prediction. However, virtual screening was also performed to evaluate toxicological properties including carcinogenicity and risk of inhibition of human ether-a-go-go-related (hERG) gene [33].

These works have been performed by using Windows 7 64-bit operating system having Intel core 2duo processor.

3 Results and discussion

3.1 Validation of Model structures

The stereo chemical quality of the predicted model was evaluated after the refinement process using Ramachandran Map calculations computed with the Procheck. The phi and psi distribution of the Ramachandran Map generated by non glycine and non proline residues are depicted in (FigureV.2) Ramachandran plot indicated that no residues have phi/psi angles in the disallowed regions and hence the quality of the model is acceptable. The percentage of residues in the “core” region of our modeled protein was found to be satisfactory [40].

The red regions in the graph indicate the most allowed regions whereas the yellow regions represent allowed regions. In this protein model, 86.6% of the residues were in the most favored region, 10.9% in allowed region, 1.8 % in generously allowed region and 0.8% of the residues lying in the disallowed regions.

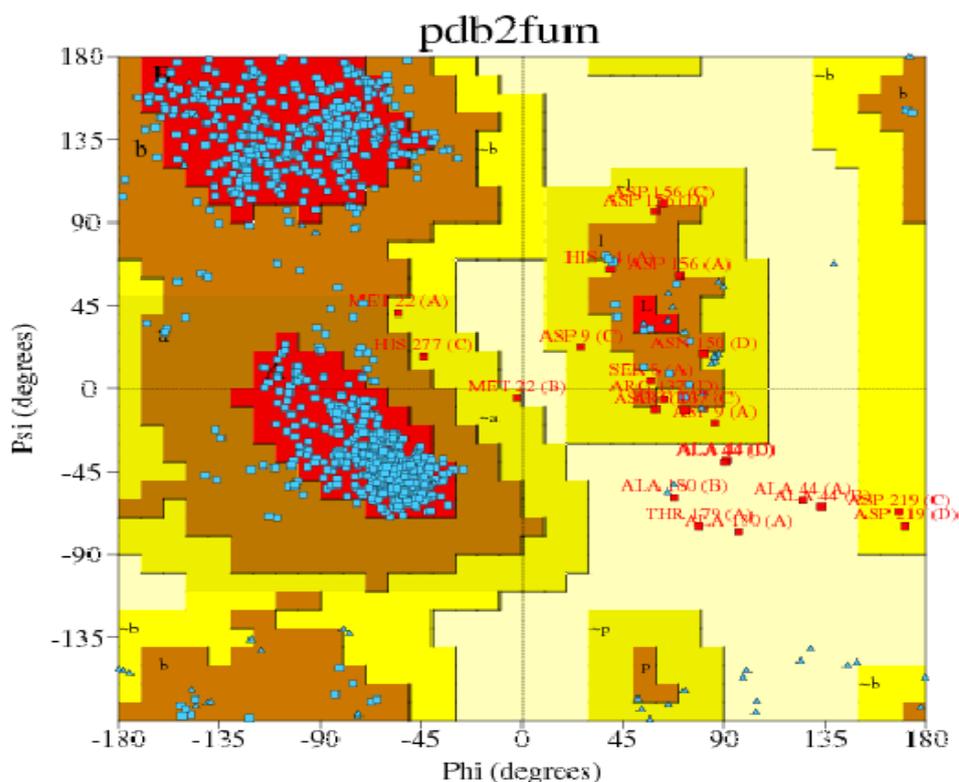


Figure IV.2: Ramachandran plot for 2FUM protein generated by Procheck

3.2 Prediction of binding sites

MVD automatically identifies potential binding sites (also referred as cavities or active sites) by using its cavity detection algorithm. The cavities within a 30 x 30 x 30 Å³ cube centered at the experimentally known ligand position were used.

In the case of the crystal structures for serine/threonine protein kinase B (2FUM) complexes, the program generally identified five different binding sites (Figure IV.3). From these five predicted cavities (Table IV.1) the one with the highest volume (107.52 Å³) and surface (295.69 Å²), was selected for consideration, as it includes the bound ligand.

Table IV.1: Cavity information of enzyme (2FUM.)

Cavity Name	Volume(Å ³)	Surface Area(Å ²)
Cavity 1	107.52	295.69
Cavity 2	59.392	170.24
Cavity 3	23.040	96
Cavity 4	16.896	80.64
Cavity 5	16.384	70.4

The binding site was defined as a spherical region which encompasses all protein atoms within 6.0 Å of bound crystallographic ligand atom dimensions X (61.79 Å), Y (2.44 Å), Z (-25.10 Å) axes, respectively). Default settings were used for all the calculations.

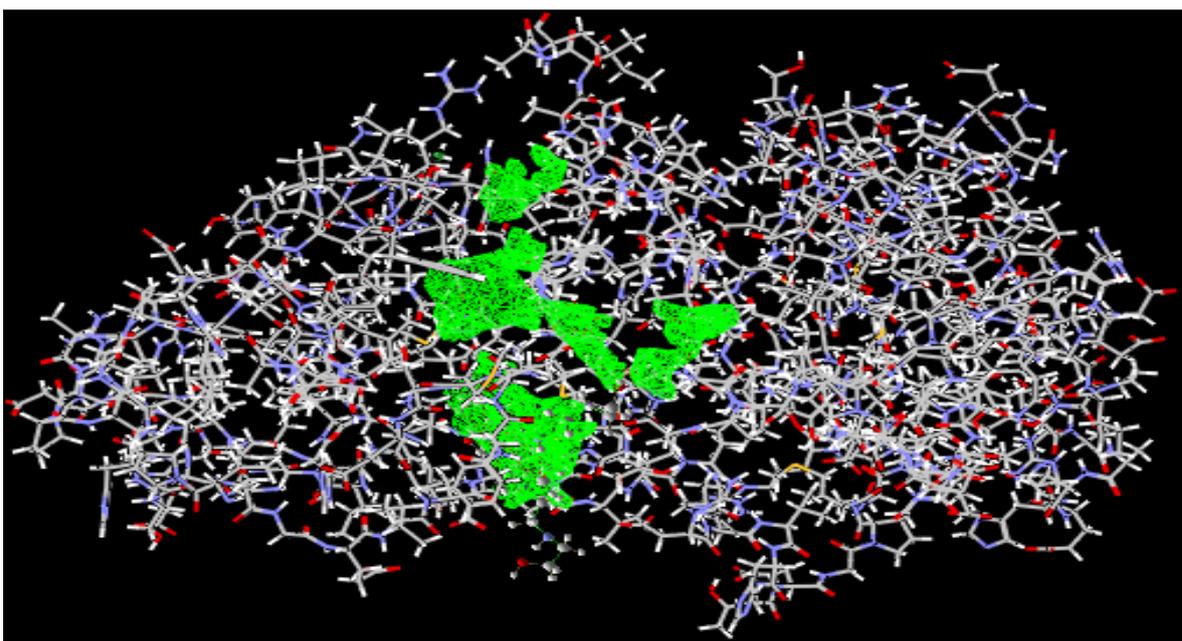


Figure IV.3: The five cavities MVD- detected cavities in PknB, PDB code; 2FUM .

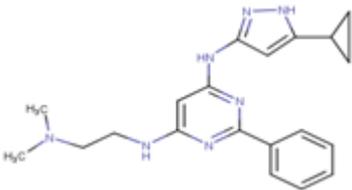
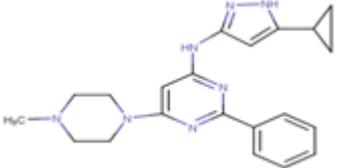
3.3 Protein-Ligand Docking:

One application of molecular docking is to design pharmaceutical in *silico* by optimizing targeted lead candidates against protein. The lead candidates can be found using a docking algorithm that aims to identify the optimal binding mode of a small molecule (ligand) to the active site of macromolecular target. Ligands have been designed to obtain more potent compounds as inhibitors of PknB.

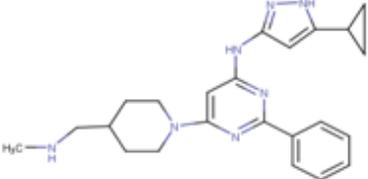
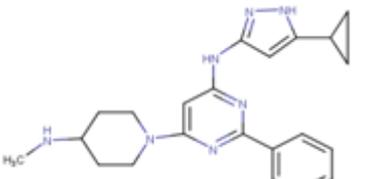
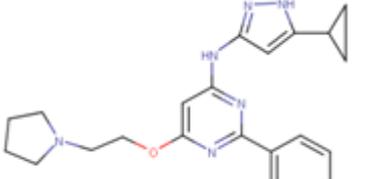
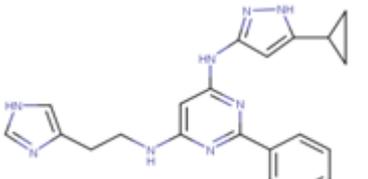
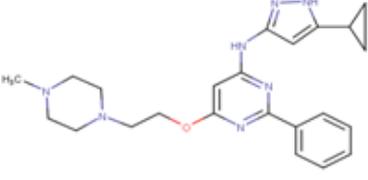
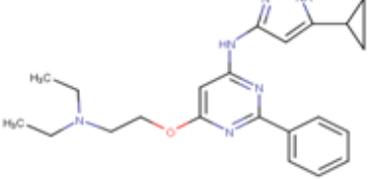
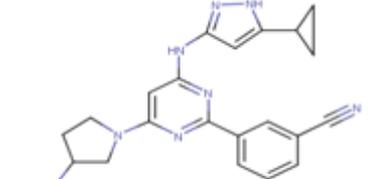
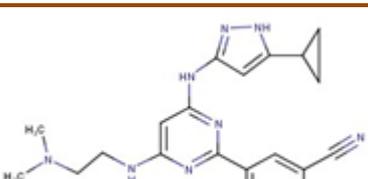
Computational analysis was carried out on chain A of the enzyme 2FUM. The 29 ligands L1-29 were selected for the study of the protein-ligand interactions, five top poses for each ligand were returned in the simulation, out of which one best pose for each ligand was selected on the basis of their MolDock score. The MVD score and the re-rank score and H-Bound score (KJ/mol), for each docking studies of amino-pyrimidine ligands with 2FUM are summarized in (Table IV.2).

The docking of ligands outcome produced the three best ranked ligands, namely, L9, L12 and L21, which showed lower Moldock score, re-rank score and a higher number of hydrogen bonding interaction than the other compounds. The binding energy values of compounds L9, L12 and L21 are -161.475, -152.003 and -143.359 kcal/mol, respectively, which are better than MIX with binding energy value of -75.683 kcal/mol. These results show that, compared to MIX as selective PknB, those three top-ranked ligand will form more stable complex and selective with, as well as, be better able to inhibit and reduce the activity of PknB.

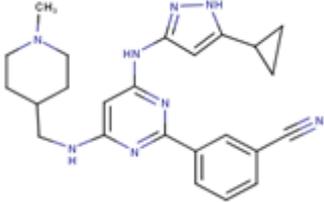
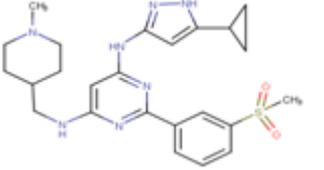
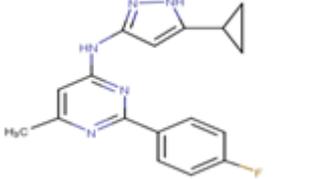
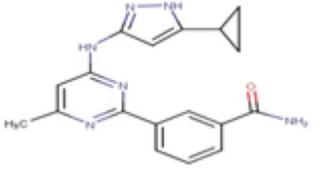
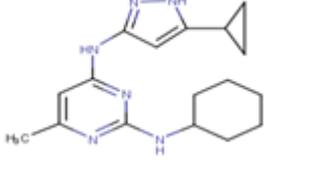
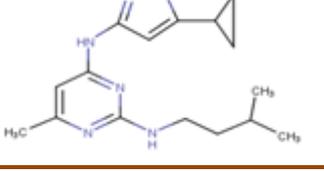
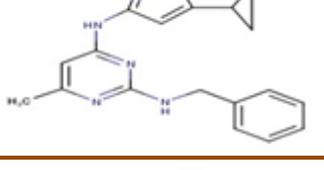
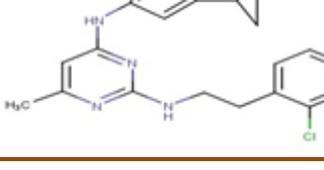
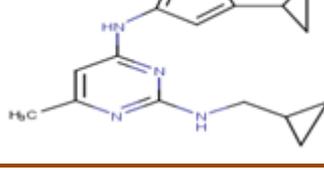
Table IV.2: The structure of amino-pyrimidine derivatives with docking scores.

No	Ligand	Moldock score (KJ/mol)	Rerank score (KJ/mol)	H-Bound score (KJ/mol)
1		-147.646	-69.264	-2.276
2		-144.615	-95.918	-2.014

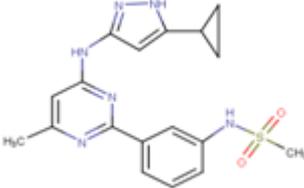
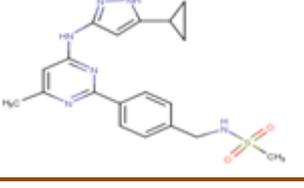
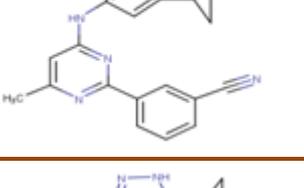
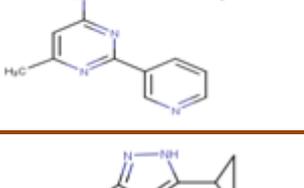
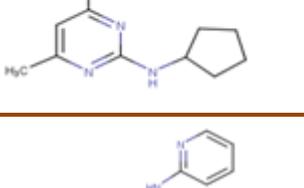
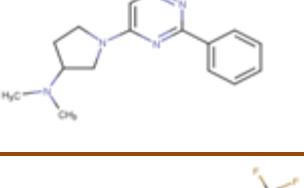
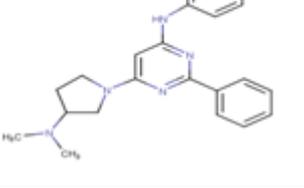
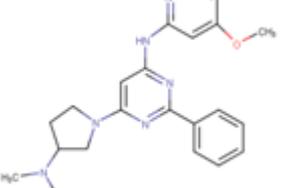
Chapter IV Molecular docking and in silico ADMET study of amino- pyrimidine Derivatives as mycobacterium tuberculosis PKnB inhibitors

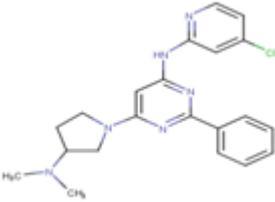
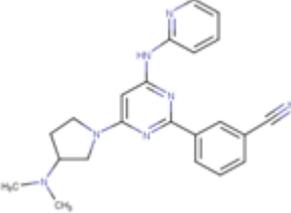
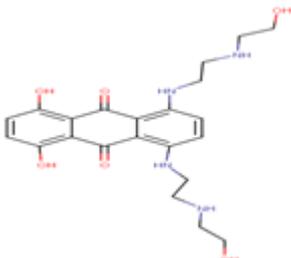
3		-139.421	-116.664	-6.416
4		-139.71	-29.120	-5.586
5		-155.848	-96.502	-7.761
6		-147.347	-109.939	0
7		-155.794	-120.147	-2.618
8		-142.356	-90.290	-2.009
9		-161.475	-128.826	-7.237
10		-153.264	-116.047	-0.639

Chapter IV Molecular docking and in silico ADMET study of amino- pyrimidine Derivatives as mycobacterium tuberculosis PKnB inhibitors

11		-151.906	-119.35	-2.961
12		-152.003	-126.941	-8.380
13		-123.111	-99.842	-4.821
14		-140.882	-112.329	-5.180
15		-133.398	-105.768	-4.163
16		-127.542	-94.644	-2.693
17		-134.871	-96.723	-2.636
18		-135.756	-100.839	-7.316
19		-125.914	-95.786	-7.169

Chapter IV Molecular docking and in silico ADMET study of amino- pyrimidine Derivatives as mycobacterium tuberculosis PKnB inhibithors

20		-143.71	-112.077	-5.767
21		-143.359	-120.853	-9.685
22		-133.614	-107.115	-4.613
23		-120.633	-99.877	-5.483
24		-135.237	-106.568	-5.233
25		-134.254	-62.821	-1.500
26		-156.908	-93359	-1.364
27		-170.109	-94.358	-3.122

28		-138.238	-111.629	0
29		-136.307	-108.308	-2.035
MIX		-75.683	-17.158	-6.501

Hydrogen bonds are key players and directly responsible for affinity and specificity in the protein-ligand complexes. Thus, to evaluate the binding affinity of the ligands with the target protein, therefore number of hydrogen bonds between ligands and amino acid of the target protein indicate its stability to inhibit the protein target.

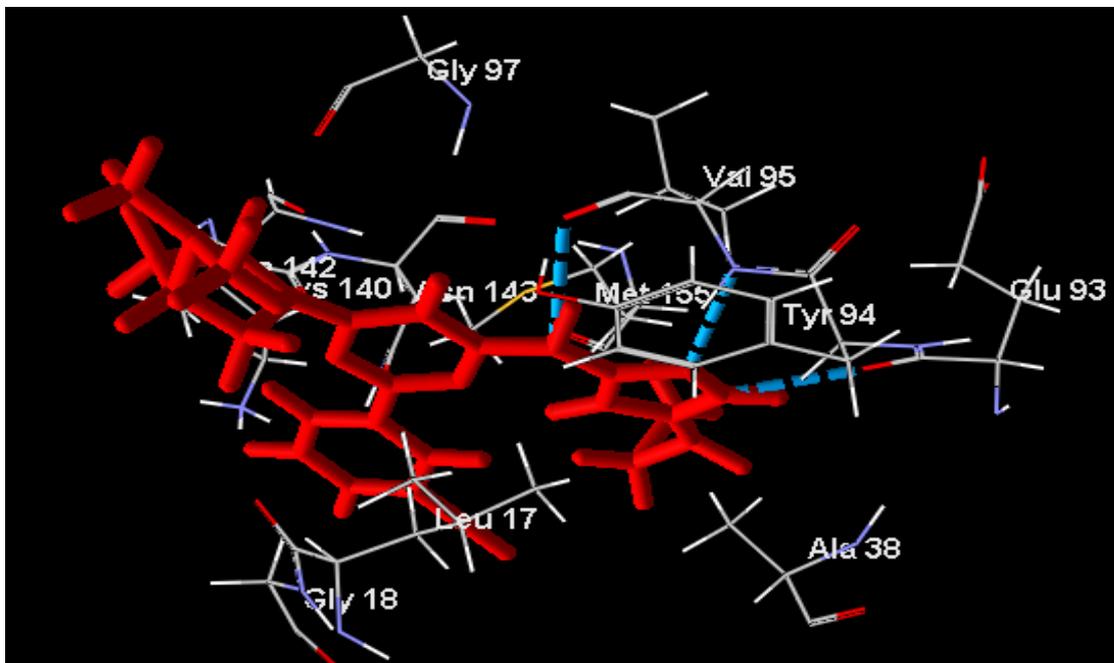
* Ligand 21, the most active ligand, formed six hydrogen bonds, and interacted with the Gly93, Val95, Lys140 and Asn143 active amino acid residues. Gly93 formed H-bonds with NH group of Pyrazole ring and Val 95 formed two hydrogen bonds with the N atom and the -NH group of Pyrazole ring, Lys140 forms two H-bonds with the two O (oxygen diatom) of the sulfonylmethane group and Asn143 formed H-bonds with the hydrogen atom of O (oxygen diatom) of the sulfonyl group.

* Ligand 12 is the second most active ligand; it forms five hydrogen bonds, and interacted with the Gly93, Val95, Leu17 and Tyr 94 active amino acid residues. Gly93 formed H-bonds with NH group of pyrazole ring, whereas Val 95 formed two hydrogen bonds with the N atom and the -NH group of the pyrazole ring. Leu 17 and Tyr 94 formed respectively one H-bond with NH group and one H-bond with the H-bond atom of O (oxygen diatom) of the sulfonyl group.

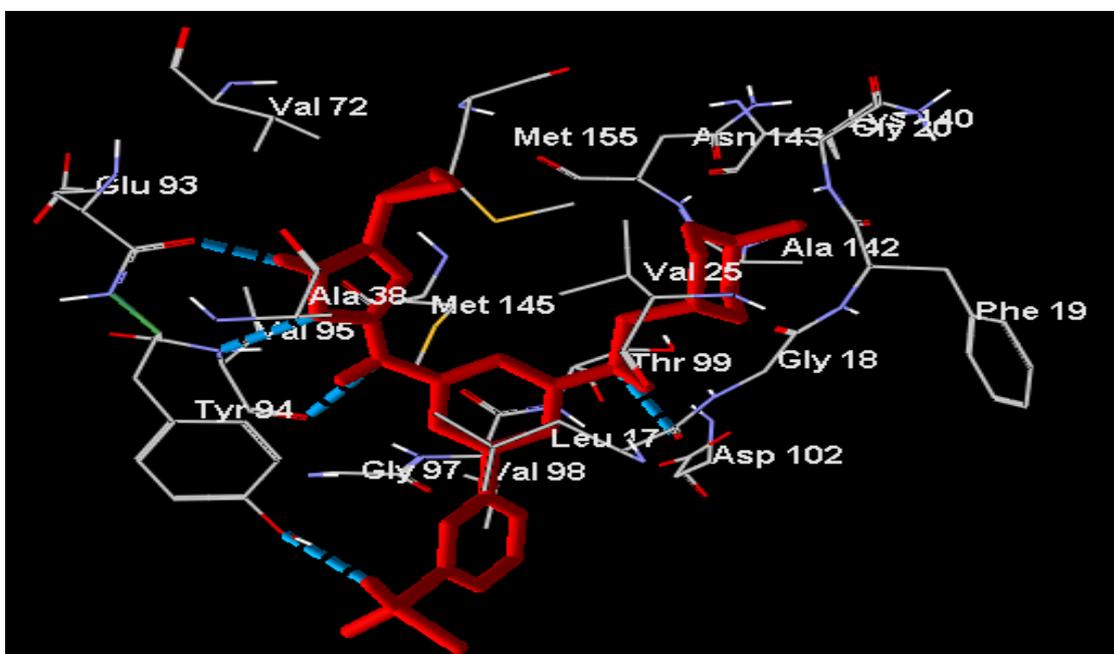
* Ligand 9 is the less most active ligand; it forms three hydrogen bonds, and interacted with the Gly93 and Val95 active amino acid residues, respectively. Gly93 formed one H-

bond with the NH group of pyrazole ring whereas Val 95 formed two hydrogen bonds with the N atom the –NH group of the pyrazole ring.

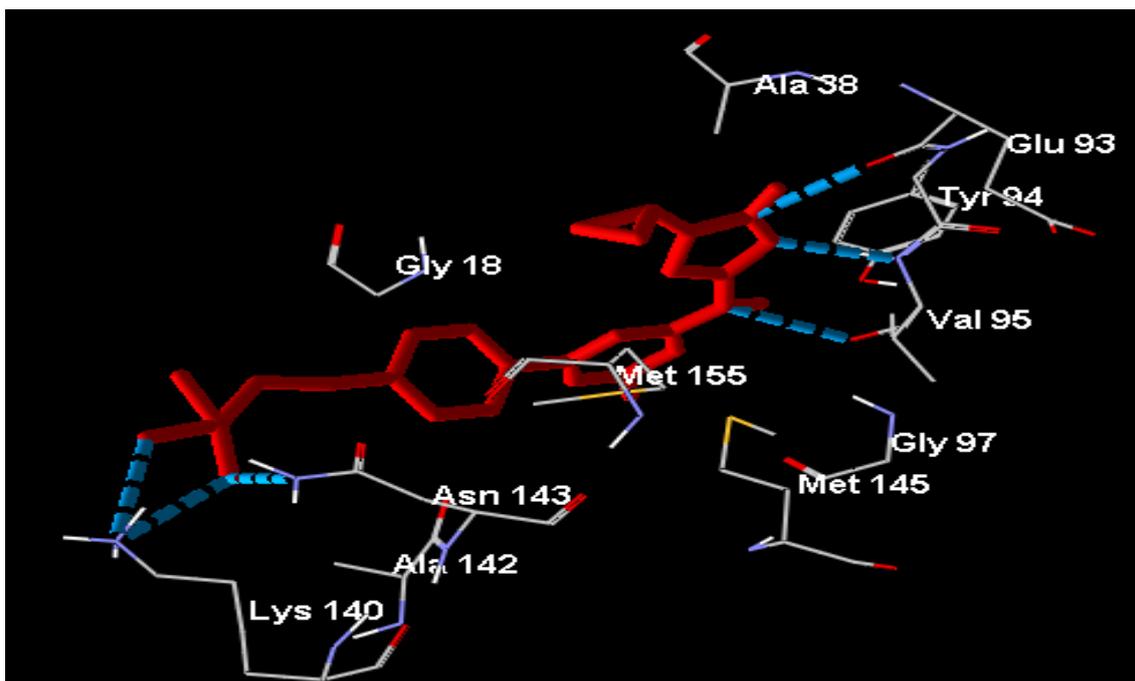
(A)



(B)



(C)



(D)

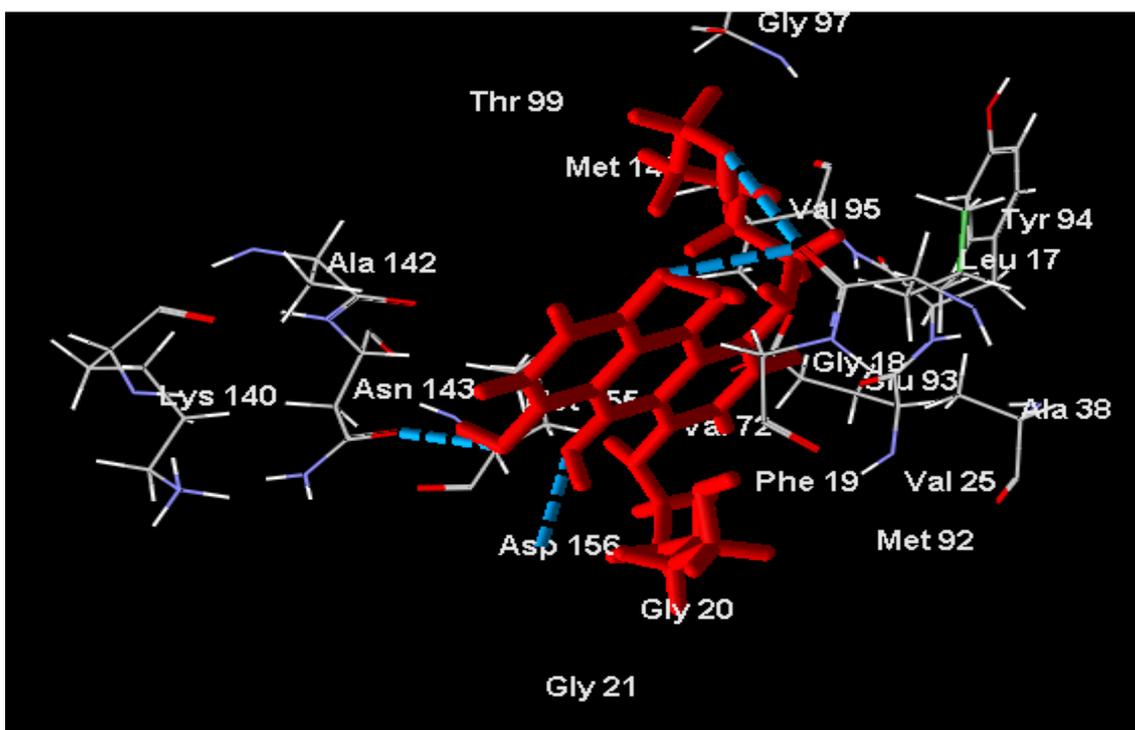


Figure IV.4: Hydrogen bonds interaction between ligands L9 (A), L12 (B), L21(C) and MIX (D) and residues of active site of 2FUM.

Table IV.3: Prediction interactions

Ligand		Protein					
Compound	Atom ID	Element	Annotated Distance	Internal Residue	Atom ID	Residue	Element
L 9	0	H	2.60	90	691	Gly93	O
	1	N	2.63	92	709	Val95	H
	9	H	3.06	92	712	Val95	O
L 12	13	H	2.63	90	691	Gly93	O
	14	N	2.61	92	709	Val95	H
	9	H	2.69	92	712	Val95	O
	22	H	2.83	14	116	Leu17	O
	60	O	2.60	91	708	Tyr94	H
L 21	13	H	2.84	90	691	Gly93	O
	12	N	2.73	92	709	Val95	H
	9	H	3.12	92	712	Val95	O
	45	O	3.15	137	1058	Lys140	H
	43	O	3.23	137	1058	Lys140	H
	43	O	3.16	140	1078	Asn143	H

The lengths of the hydrogen bonds in the interval: $2.5\text{\AA} \leq x \leq 3.1\text{\AA}$ are considered strong interactions; however values in the interval $3.1\text{\AA} \leq x \leq 3.55\text{\AA}$: the averages interactions, whereas values $> 3.55\text{\AA}$ are considered weak. It is noticed that the values obtained for distances of hydrogen bonds between the ligands L9, L12 and the residues of active site belong to the interval $2.5\text{\AA} \leq x \leq 3.1\text{\AA}$. These results indicate that the strong affinity of L9 and L12 on 2FUM could lead to the potent inhibition of the catalytic activity of the enzyme [41].

LigPlot is known as the comprehensive tool for expressing the hydrogen bonding and hydrophobic interactions involving the ligand molecule and active site residues (Figure IV.5) gives a more detailed insight of the interactions with particular amino acids in enzyme binding pocket. Based on the presented results, it can be concluded that hydrophobic interactions between ligands L9, L12, L21 and binding pocket play an important role. However, number, bond length and bond energy of hydrogen bonds formed between ligand and enzyme has an important role in ligand effect on investigated activity.

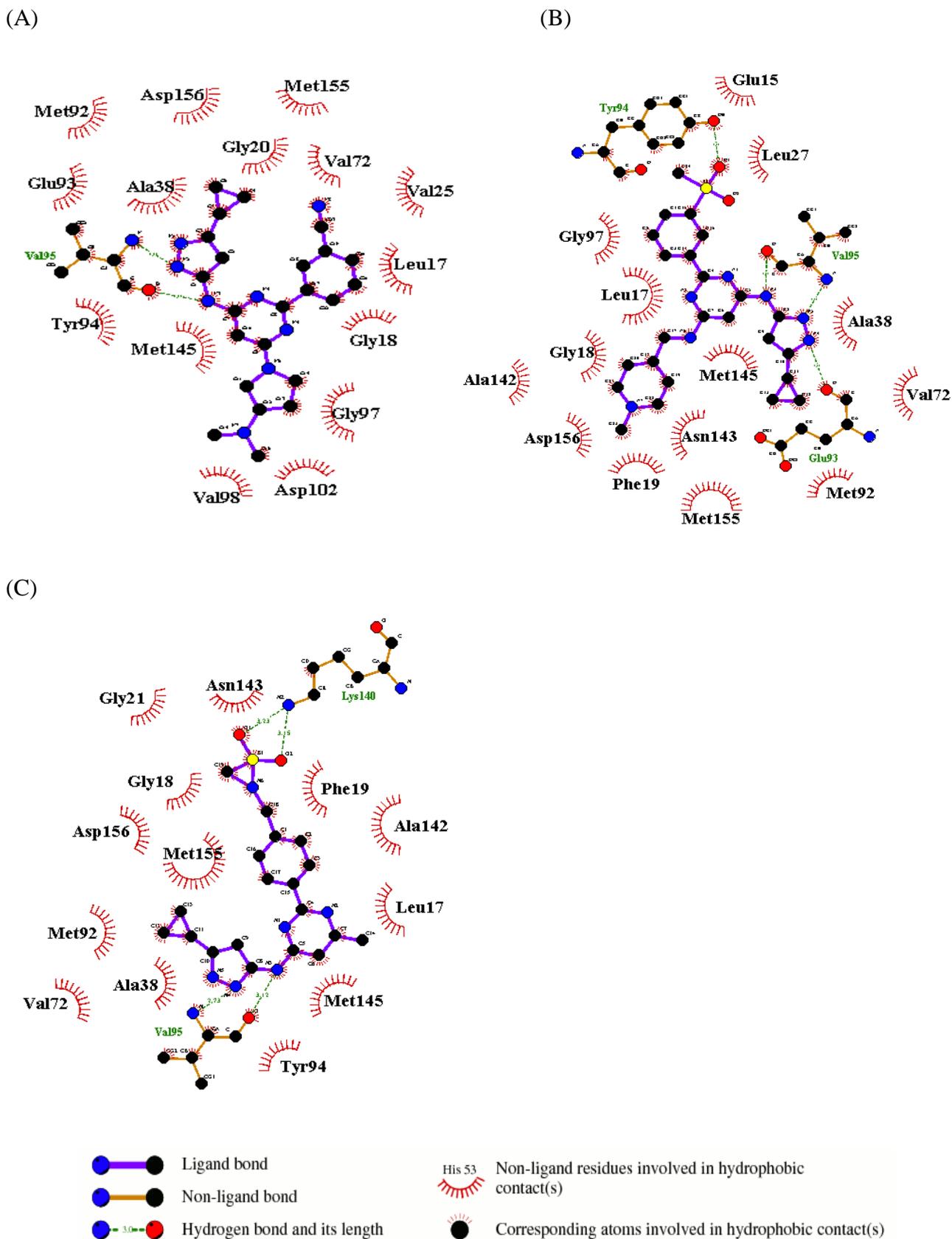


Figure IV.5: LigPlot generated for the best poses obtained with ligands L9 (A), L12 (B), L21 (C) and receptor 2FUM.

3.4 Drug-likeness and ADMET Properties of selected ligands

3.4.1 Drug-likeness of selected ligands

The selected ligands used in this study were evaluated as selective inhibitor 2FUM protein target by comparison. The oral bioavailability of the compounds projected as potential drugs were evaluated by determining the molecular weight, number of rotatable bonds, H bond donor and acceptor, and drug's polarsurface (TPSA). Since the individual molecular weights of all the compounds were less than 500, the numbers of the rotatable bond were <10, the number of hydrogen bond donors and acceptors were < 10, an octanol-water partition coefficient logP not greater than 5. TPSA values being <140, they qualified to be an ideal oral drug. Ligands tested in this study were also predicted to have good oral bioavailability and all the ligand qualify of the Lipinski's Rule of 5.

According to (Table IV.4), all the ligands possess limited aqueous solubility which ranged from very slightly soluble to practically insoluble [42]. This result is also supported by the calculation of partition coefficient which indicates that all the tested compounds are lipophilic and have the best affinity to reside in *n*-octanol than in water i.e. all the ligands will pass the plasma membrane easily (Table IV.4).

Table IV.4: Physicochemical property of anti-tubercular ligands.

Ligand	Molecular weight	Intrinsic aqueous solubility logS	Partition coefficient (logP)	H bond donor	H bond acceptor	TPSA
L9	414.517	-5.95	4.81	2	8	96.76
L12	481.62	-4.87	3.55	3	9	115.90
L21	398.49	-4.65	2.88	3	8	112.66

In drug-likeness prediction, a molecule can be considered to have drug-like features Only if it satisfies most of the rules including rule of five, CMC-like rule, MDDR-like rule, Lead-like rule, WDI-like rule. All of the selected ligands qualified Lipinski's rule, WDI rule as it was in the cut-off range (90%) and CMC-like rule expected L12 and MIX, such as MIXMDDR rule, whereas L9, L12, L21 showed mid-structure to MDDR-like rule. However, all ligands Violated Lead-like rule (Table IV.5).

Table IV.5: Drug likeness prediction using PreADMET Server

Ligands	CMC like rule	Lead like rule	Lipinski's "Rule of five"	WDI like rule	MDDR Like
L9	Qualified	Violated	Suitable	Out of 90 %	Mid-structure
L12	Not Qualified	Violated	Suitable	Out of 90 %	Mid-structure
L21	Qualified	Violated	Suitable	Out of 90 %	Mid-structure
MIX	Not Qualified	Violated	Suitable	Out of 90 %	Drug-like

3.4.2 ADMET study of selected ligands

The pharmacokinetic studies such as absorption, distribution of L9, L12, L21 and MIX were performed using online server PreADMET (<https://preadmet.bmdrc.kr/>). The calculated absorption, distribution parameters are presented in (Table IV.6). The calculated human intestinal absorption (HIA) was ranged from 85.48% to 90.926 % which suggest that all the tested compounds are well absorbed through the intestinal cell [43].

Table IV.6: ADME properties of ligands

Ligands	Absorption			Distribution		
	Human intestinal Absorption (HIA, %) in vitro Caco-2 cell permeability (nm/s)	Skin permeability	plasma Protein binding (%)	blood brain barrier penetration (c.brain/c.blood	P-Glycoprotein (Inhibition)	
L9	90.926	15.148	-3.721	82.193	1.141	Inhibitor
L12	89.806	1.605	-2.642	87.243	0.292	Inhibitor
L21	85.480	6.312	-2.704	93.097	0.118	Non inhibitor
MIX	24.710	18.258	-5.094	22.314	0.029	Non inhibitor

In addition, all compounds exhibited high permeability [44] as absorption values through Caco-2 cell (PCaco-2) were within 15.148 - 6.312 nm/s excepted 19 (1.605 nm/s). The skin permeability (PSkin) is a vital parameter for the assessment of drugs and chemical that might require transdermal administration. All the compounds were found to be impermeable through skin since the calculated PSkin value was negative.

The distribution properties were assessed by evaluating the brain to blood partition coefficient ($C_{\text{brain}}/C_{\text{blood}}$), plasma binding (PPB) and interaction with the P-glycoprotein (Pgp). The calculated values of PPB were 82.193 to 93.097 %. Generally compounds with more than 90% of PPB are classified as strongly bound chemicals whereas less than 90% are weakly bound chemicals (<https://preadmet.bmdrc.kr/adme-prediction/>). Therefore, among L21 bound strongly with plasma protein whereas L9 and L12 are weakly bound chemicals. The $C_{\text{brain}}/C_{\text{blood}}$ values were 0.118 to 1.141. Based on $C_{\text{brain}}/C_{\text{blood}}$ ratio all chemicals fall under three categories namely high absorption to CNS ($C_{\text{brain}}/C_{\text{blood}}$ value more than 2.0), middle absorption to CNS ($C_{\text{brain}}/C_{\text{blood}}$ value within 2.0 - 0.1) and low absorption to CNS ($C_{\text{brain}}/C_{\text{blood}}$ value less than 0.1). The ratio of $C_{\text{brain}}/C_{\text{blood}}$ suggests middle absorption of these agents to CNS indicating moderate to higher ability to cross blood brain barrier (BBB). P-glycoprotein (Pgp), produced from the multi drug resistance (MDR) gene and an ATP dependent efflux transporter that affects the absorption, distribution and excretion of clinically important drugs [45].

The ligands L9 and L12 are inhibitors for Pgp and L21 is not inhibitors for Pgp. Ligand L21 showed PPB=93.097%, HIA= 85.48 and cell permeability =6.312nm/s, this ligand will be with greater availability to bind with the target receptor when compared to other compounds. (Table IV.7) shows the results of mutagenic (Ames test) and carcinogenic (using mouse and rat model) properties of L9, L12, L21 and MIX. Toxicological investigation of drug candidates is one of the key steps for drug discovery. This means that the toxicity study is very important for new compounds.

The Ames test is widely used and an accepted test to evaluate the mutagenicity of a chemical agent. In this test, all the compounds exhibited negative prediction mutagenic compound excepted 19.

Table IV.7: Toxicological properties of anti-tubercular agents.

Ligands	Mutagenicity (Ames test)	Carcinogenicity		in vitro hERG inhibition
		Mouse	Rat	
L9	mutagen	Negative	Negative	Medium risk
L12	non-mutagen	Negative	Negative	Ambiguous
L21	non-mutagen	Negative	Negative	Ambiguous
MIX	non-mutagen	Negative	Negative	Ambiguous

In carcinogenicity study, the PreADMET server was utilized to predict the carcinogenicity of chemical agent. In the prediction of carcinogenicity, negative prediction indicates there is evidence of carcinogenic activity whereas positive means the tested compound does not exhibit carcinogenic activity. Among all ligands demonstrated carcinogenicity in both mouse and rat model. The risk of inhibition of human ether-a-go-go-related (hERG) gene was varied from medium and ambiguous.

4 Conclusion

Virtual screening methods are widely used for reducing cost and time of drug discovery process. From this study of *in silico* drug designing and molecular docking of the 29 amino-pyrimidine derivatives, we conclude that those derivatives have the ability to inhibit the Mycobacterium tuberculosis protein kinase B. The scoring results reveal the higher negative mol dock score, rerank score and hydrogen bond interaction of the title compounds in comparison to MIX. It was found that the three ligands L9, L12, L21 showed better results from 29 docked ligands. Furthermore, Pharmacokinetic effects of the ligand L21 observed comparatively better bioavailability, distribution, absorption, than MIX and other ligands. Hence, it could be concluded that the ligand L21 could be considered as potent drug candidate of M. tuberculosis.

References

1. S. Lapenna and A. Giordano, "Cell cycle kinases as therapeutic targets for cancer," *Nat Rev Drug Discov*, vol. 8, no. 7, pp. 547–566, Jul. 2009.
2. P. Barthe, G. V. Mukamolova, C. Roumestand, and M. Cohen-Gonsaud, "The Structure of PknB Extracellular PASTA Domain from Mycobacterium tuberculosis Suggests a Ligand-Dependent Kinase Activation," *Structure*, vol. 18, no. 5, pp. 606–615, May 2010.
3. T. A. Young, B. Delagoutte, J. A. Endrizzi, A. M. Falick, and T. Alber, "Structure of Mycobacterium tuberculosis PknB supports a universal activation mechanism for Ser/Thr protein kinases," *Nat. Struct. Biol.*, vol. 10, no. 3, pp. 168–174, Mar. 2003.
4. Y. Av-Gay, S. Jamil, and S. J. Drews, "Expression and characterization of the Mycobacterium tuberculosis serine/threonine protein kinase PknB," *Infect. Immun.*, vol. 67, no. 11, pp. 5676–5682, Nov. 1999.
5. P. Fernandez *et al.*, "The Ser/Thr Protein Kinase PknB Is Essential for Sustaining Mycobacterial Growth," *J. Bacteriol.*, vol. 188, no. 22, pp. 7778–7784, Nov. 2006.
6. C. Mieczkowski, A. T. Iavarone, and T. Alber, "Auto-activation mechanism of the Mycobacterium tuberculosis PknB receptor Ser/Thr kinase," *EMBO J*, vol. 27, no. 23, pp. 3186–3197, Dec. 2008.
7. K. Sharma, M. Gupta, A. Krupa, N. Srinivasan, and Y. Singh, "EmbR, a regulatory protein with ATPase activity, is a substrate of multiple serine/threonine kinases and phosphatase in Mycobacterium , *FEBS J.*, vol. 273, no. 12, pp. 2711–2721, Jun. 2006.
8. A. Dasgupta, P. Datta, M. Kundu, and J. Basu, "The serine/threonine kinase PknB of Mycobacterium tuberculosis phosphorylates PBPA, a penicillin-binding protein required for cell division," *Microbiology (Reading, Engl.)*, vol. 152, no. Pt 2, pp. 493–504, Feb. 2006.
9. C.M. Kang, D. W. Abbott, S. T. Park, C. C. Dascher, L. C. Cantley, and R. N. Husson, "The Mycobacterium tuberculosis serine/threonine kinases PknA and PknB: substrate identification and regulation of cell shape," *Genes Dev*, vol. 19, no. 14, pp. 1692–1704, Jul. 2005.
10. M. Ortiz-Lombardía, F. Pompeo, B. Boitel, and P. M. Alzari, "Crystal structure of the catalytic domain of the PknB serine/threonine kinase from Mycobacterium tuberculosis," *J. Biol. Chem.*, vol. 278, no. 15, pp. 13094–13100, Apr. 2003.

11. A. Narayan, P. Sachdeva, K. Sharma, A. K. Saini, A. K. Tyagi, and Y. Singh, "Serine threonine protein kinases of mycobacterial genus: phylogeny to function," *Physiol. Genomics*, vol. 29, no. 1, pp. 66–75, Mar. 2007.
12. 4A. Villarino *et al.*, "Proteomic identification of M. tuberculosis protein kinase substrates: PknB recruits GarA, a FHA domain-containing protein, through activation loop-mediated interactions," *J. Mol. Biol.*, vol. 350, no. 5, pp. 953–963, Jul. 2005.
13. K. E.A. Loughheed, S. A. Osborne, B. Saxty, D. Whalley, T. M.Chapman,N. Bouloc, J. Chugh, T. J. Nott, D. Patel, V. L. Spivey, C. A.Kettleborough, J. S. Bryans, D. L. Taylor, S. J. Smerdon, and R. S.Buxton, "Effective inhibitors of the essential kinase PknB and their potential as anti-mycobacterial agents," *Tuberculosis (Edinb)*, vol. 91, no. 4, pp. 277–286, Jul. 2011.
14. T. M. Chapman, N. Bouloc, R. S. Buxton, J. Chugh, K.E.A. Loughheed, S.A. Osborne, B. Saxty, S. J. Smerdon, D. L. Taylor and D. Whalley, "Substituted aminopyrimidine protein kinase B (PknB) inhibitors show activity against Mycobacterium tuberculosis," *Bioorg .Med .Chem. Lett*, vol. 22, no. 9, pp. 3349–3353, May 2012.
15. A. Seal, P. Yogeewari, D. Sriram, O. Consortium, and D. J. Wild, "Enhanced ranking of PknB Inhibitors using data fusion methods," *Journal of Cheminformatics*, vol. 5, no. 1, p. 1-11, Jan. 2013.
16. R. Székely *et al.*, "A novel drug discovery concept for tuberculosis: inhibition of bacterial and host cell signalling," *Immunol. Lett.*, vol. 116, no. 2, pp. 225–231, Mar. 2008.
17. B. Hegymegi-Barakonyi *et al.*, "Signalling inhibitors against Mycobacterium tuberculosis--early days of a new therapeutic concept in tuberculosis," *Curr. Med. Chem.*, vol. 15, no. 26, pp. 2760–2770, 2008.
18. A. Wehenkel *et al.*, "The structure of PknB in complex with mitoxantrone, an ATP-competitive inhibitor, suggests a mode of protein kinase regulation in mycobacteria," *FEBS Lett.*, vol. 580, no. 13, pp. 3018–3022, May 2006.
19. S. Khamouli, S. Belaidi, Z. Almi, S. Medjahed, H. Belaidi, "Property/Activity Relationships and Drug Likeness for Pyrimidine Derivatives as Serine/Threonine Protein Kinase B Inhibitors," *J. Bionosci.* ,vol. 11,no.4 pp.301–309,Aug. 2017.
20. S. Lyskov and J. J. Gray, "The RosettaDock server for local protein–protein docking," *Nucleic Acids Res*, vol. 36, no. suppl_2, pp. W233–W238, Jul. 2008.
21. D.Vidal, , R. Garcia-Sern, J. Mestres,"Ligand-based approaches to in silico pharmacology," *Methods Mol.Biol.* ,vol. 672,pp.489–502, 2011.

22. 2FUM: 3-dimensional structure downloaded from <http://www.rcsb.org/pdb>.
23. A. Wehenkel *et al.*, “The structure of PknB in complex with mitoxantrone, an ATP-competitive inhibitor, suggests a mode of protein kinase regulation in mycobacteria,” *FEBS Lett.*, vol. 580, no. 13, pp. 3018–3022, May 2006.
24. G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, “Stereochemistry of polypeptide chain configurations,” *J. Mol. Biol.*, vol. 7, pp. 95–99, Jul. 1963.
25. M. V. Damre, R. P. Gangwal, G. V. Dhoke, M. Lalit, D. Sharma, K. Khandelwal, and A. T. Sangamwar, “3D-QSAR and molecular docking studies of amino-pyrimidine derivatives as PknB inhibitors,” *J. Taiwan Inst. Chem. Eng.*, vol. 45, no. 2, pp. 354–364, 2014.
26. Hyper Chem (Molecular Modeling System) Hypercube, Inc., 1115 NW, 4th Street, Gainesville, FL 32601, USA.2008.
27. MarvinSketch 17.1.2, Chemaxon .2017, (<http://www.chemaxon.com>).
28. <http://molegro.com/mvd-product.php>
29. Bursulaya B D, Totrov M, Abagyan R, and Brooks C L .Comparative study of several algorithms for flexible ligand docking. *J. Comput. Aided Mol. Des.*, 2003, 17(11), 755–763
30. R. Thomsen and M. H. Christensen, “MolDock: A New Technique for High-Accuracy Molecular Docking,” *J. Med. Chem.*, vol. 49, no. 11, pp. 3315–3321, 2006.
31. A. R. Anuf, K Ramaraj, “Inhibition of FGFR2 signal transduction in Acne Vulgaris using Bioactive Flavonoids: An Insilico approach,”*J. Med. Pharm. Sci.*, vol. 2, no. 2, pp. 136–142, 2014.
32. A. C. Wallace, R. A. Laskowski, and J. M. Thornton, “LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions,” *Protein Eng.*, vol. 8, no. 2, pp. 127–134, Feb. 1995.
33. D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward, and K. D. Kopple, *J. Med. Chem.* 45, 2615 (2002).
34. C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, *J. Adv. Drug Deliv. Rev.* 64, 4 (2012).
35. Ajay, W. P. Walters, and M. A. Mureko, “Can We Learn To Distinguish between ‘Drug-like’ and ‘Nondrug-like’ Molecules?,” *J. Med. Chem.*, vol. 41, no. 18, pp. 3314–3324, Aug. 1998.

36. T. I. Oprea, T. K. Allu, D. C. Fara, R. F. Rad, L. Ostopovici, and C. G. Bologa, "Lead-like, Drug-like or 'Pub-like': How different are they?," *J Comput Aided Mol Des*, vol. 21, no. 1–3, pp. 113–119, 2007.
37. M. Wagener and V. J. van Geerestein, "Potential Drugs and Nondrugs: Prediction and Identification of Important Structural Features," *J. Chem. Inf. Comput. Sci.*, vol. 40, no. 2, pp. 280–292, Mar. 2000.
38. A. Kulkarni, Y. Han, and A. J. Hopfinger, "Predicting Caco-2 Cell Permeation Coefficients of Organic Molecules Using Membrane-Interaction QSAR Analysis," *J. Chem. Inf. Comput. Sci.*, vol. 42, no. 2, pp. 331–342, Mar. 2002.
39. J. D. Irvine *et al.*, "MDCK (Madin-Darby canine kidney) cells: A tool for membrane permeability screening," *J Pharm Sci*, vol. 88, no. 1, pp. 28–33, Jan. 1999.
40. S. C. Lovell *et al.*, "Structure validation by Calpha geometry: phi,psi and Cbeta deviation," *Proteins*, vol. 50, no. 3, pp. 437–450, Feb. 2003.
41. A. Imberty, K. D. Hardman, J. P. Carver, and S. Pérez, "Molecular modelling of protein-carbohydrate interactions. Docking of monosaccharides in the binding site of concanavalin A," *Glycobiology*, vol. 1, no. 6, pp. 631–642, Dec. 1991.
42. P. J. Sinko, *Martin's Physical Pharmacy and Pharmaceutical Sciences: Physical Chemical and Biopharmaceutical Principles in the Pharmaceutical Sciences, 50th Anniversary Edition*, Sixth, North American Edition. Baltimore, MD: Lippincott Williams and Wilkins, 2010.
43. S. Yee, "In vitro permeability across Caco-2 cells (colonic) can predict in vivo (small intestinal) absorption in man--fact or myth," *Pharm. Res.*, vol. 14, no. 6, pp. 763–766, Jun. 1997.
44. M. Yazdanian, S. L. Glynn, J. L. Wright, and A. Hawi, "Correlating partitioning and caco-2 cell permeability of structurally diverse small molecular weight compounds," *Pharm. Res.*, vol. 15, no. 9, pp. 1490–1494, Sep. 1998.
45. A.H. Schinkel, "P-Glycoprotein, a gatekeeper in the blood-brain barrier," *Adv. Drug Deliv. Rev.*, vol. 36, no. 2–3, pp. 179–194, Apr. 1999.

General Conclusion

In this work, we have used several computational techniques that have been designed to aid early drug discovery in a setting with limited prior information. They have also been applied to improve our understanding of the requirements for binding of a target, and to identify small molecule inhibitors of the potential anti-TB.

The first part, we have studied the structure-activity relationship (SAR) and drug likeness proprieties of a bioactive series of amino-pyrimidine derivatives have shown that structural units involved in the biological activity. In our case, the Lipinski and Veber rules are validated. Therefore, theoretically, all compounds will not have a problem with oral bioavailability.

Next to, we have applied Quantitative structure activity relationship (QSAR) analysis to a series of 29 amino-pyrimidine as PknB inhibitors using a combination of various physicochemical and quantum descriptors. A multiple linear regression (MLR) procedure was used to model the relationships between molecular descriptors and the chemotherapeutic activity of the amino-pyrimidine derivatives. Good agreement between experimental and predicted activity values, obtained in the validation procedure, indicated the good quality of the derived QSAR model. The best QSAR model developed show a good predictive correlation coefficient $R^2= 0.973$ and external predictive ability of prediction $R_{\text{test}}^2=0.778$ was developed by MLR. The proposed model has good robustness and predictability when verified by internal and external validation.

In the second part, the docking study using Molegro Virtual Docker has been performed on these predicted compounds to find their interactions with the receptor PDB (2FUM). We have selected three best ranked ligands L9, L12 and L21, which showed lower Moldock Score, re-rank score and a higher number of hydrogen bonding interaction than the other compounds.

In silico ADMET predictions in PreADMET server revealed that except L21 compound had good absorption as well as solubility characteristics and had minimal toxic effects. It could be concluded that the ligand L21 could be considered as potent drug candidate of *M. tuberculosis*.

General Conclusion

In this research we conclude that computational chemistry is of great importance in the diagnosis of large and small molecules. There are several approaches in which to diagnose these compounds. Computational chemistry has an importance in medicine, diagnosis and discovery of drugs and the importance of the docking and QSAR approaches in diagnosing of these pharmaceutical substances.

Abstract:

The discovery of anti-tuberculosis agents is crucial for effective tuberculosis therapy. The present strategy for new drug development is directed towards identifying essential enzyme systems in the bacteria and developing potent molecules to inhibit them. We have study the inhibition of 29 amino-pyrimidine derivatives of Mycobacterium tuberculosis PKnB using different CADD approaches to search for potential drug candidates.

The multi-parameter optimization (MPO) process to predict the best balance of properties of these compounds was expressed by various approaches such as Lipophilic Efficiency (LipE), Lipinski, Veber rules. The Multiple linear regression method (MLR) was applied to derive QSAR model which was further validated for statistical significance by internal and external validation. In Silico molecular docking and ADMET properties used to predict novel PKnB inhibitors and guide the discovery of new potential analogs.

Keywords: anti-tuberculosis, CADD, MPO, MLR, QSAR, docking, ADMET.

Résumé :

La découverte d'agents antituberculeux est cruciale pour un traitement efficace de la tuberculose. La stratégie actuelle de développement de nouveaux médicaments vise à identifier les systèmes enzymatiques essentiels dans les bactéries et à développer des molécules puissantes pour les inhiber. Nous avons étudié l'inhibition de 29 dérivés d'aminopyrimidine de Mycobactérie tuberculoses PKnB en utilisant différentes approches CADD pour la recherche de candidats médicaments.

Le processus d'optimisation multi-paramètres (MPO) permettant de prédire le meilleur équilibre des propriétés de ces composés a été exprimé par des diverses approches telles que les règles Efficacité lipophylique de ligand (LipE), Lipinski, Veber. La méthode de régression linéaire multiple (MLR) a été appliquée pour dériver de modèle QSAR qui a ensuite été validés pour la signification statistique par validation interne et externe. In Silico, les propriétés de doking moléculaire et ADMET sont utilisées pour prédire de nouveaux inhibiteurs de la PKnB et guider la découverte de nouveaux analogues potentiels.

Mots clés : antituberculeux, CADD, MPO, MLR, QSAR, docking, ADMET.

ملخص

اكتشاف العوامل المضادة للسل أمر بالغ الأهمية لعلاج السل الفعال. تهدف الإستراتيجية الحالية لتطوير أدوية جديدة وتحديد أنظمة الإنزيمات الأساسية في البكتيريا وتطوير جزيئات قوية لتثبيطها. لقد قمنا بدراسة 29 مشتق امينو بريميدين (amino- Pyrimidine) للمتفطرة السل باستخدام مختلف مناهج (CADD للبحث عن أحسن مرشح كدواء.

تم التعبير عن عملية التحسين المتعددة العوامل (MPO) للتنبؤ بأفضل توازن لخصائص هذه المركبات من خلال طرق مختلفة مثل كفاءة الدهون (LipE), Lipinski, Veber. تم تطبيق طريقة الانحدار الخطي المتعدد لاشتقاق نموذج QSAR والذي تم التحقق من صحته باستعمال طرق إحصائية طبقت للتحقق الداخلي والخارجي. تُستخدم خصائص الالتحام الجزيئي docking و ADMET للتنبؤ بمثبطات جديدة لـ PKnB وتوجيه اكتشاف مركبات جديدة مماثلة.

الكلمات الدالة: المضادة للسل , CADD , MPO, QSAR, MLR, الالتحام الجزيئي , ADMET,