# QUANTUM MODELING OF SOME POTENT NON-TOXIC ANTI-TUBERCULOSIS COMPOUNDS

**BY**

**GOWAL MUDIMA ERIC**

**DEPARTMENT OF CHEMISTRY,**

**AHMADU BELLO UNIVERSITY,**

**ZARIA, NIGERIA**

**SEPTEMBER 2016**

# QUANTUM MODELING OF SOME POTENT NON-TOXIC ANTI-TUBERCULOSIS COMPOUNDS

**By**

**ERIC MUDIMA, GOWAL**

**MSC/SCI/3061/2011-2012**

**A DISSERTATION SUBMITTED TO THE SCHOOL OF POST GRADUATE STUDIES,**

**AHMADU BELLO UNIVERSITY, ZARIA**

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD**

**OF MASTER DEGREE IN PHYSICAL CHEMISTRY**

**DEPARTMENT OF CHEMISTRY,**

**FACULTY OF SCIENCE,**

**AHMADU BELLO UNIVERSITY,**

**ZARIA, NIGERIA**

**SEPTEMBER, 2016**

## DECLARATION

I declare that the work in the Dissertation entitled '**Quantum Modeling of Some Potent Non-Toxic Anti-tuberculosis Compounds'** has been carried out by me in the Department of Chemistry, under the supervision of Professor A. Uzairu and Dr. P.A.P Mamza. The information derived from the literature has been duly acknowledged in the text and a list of references provided. No part of this research was previously presented for another degree or diploma at this or any other institution.

Gowal Mudima Eric

| --- | --- | --- |
| Name of Student | Signature | Date |

## CERTIFICATION

This project Dissertation entitled **QUANTUM MODELING OF SOME POTENT NON-TOXIC ANTI-TUBERCULOSIS COMPOUNDS** by Eric Mudima GOWAL meets the regulation governing the degree of Master of Science in Physical Chemistry of the Ahmadu Bello University, and is approved for its contribution to knowledge and literary presentation.

_____ _____

PROF. A Uzairu DATE

(Chairman, Supervisory Committee)

_____ _____

Dr. P.A.P Mamza DATE

(Member, Supervisory Committee)

_____ _____

PROF. V.O AJIBOLA DATE

HEAD OF DEPARTMENT

_____ _____

PROF. K. BALA DATE

DEAN, POSTGRADUATE SCHOOL

## ACKNOWLEDGEMENTS

First and foremost, I am eternally grateful to God for making it possible for me to carry out this research work. He has been my anchor.

I also wish to express my sincere gratitude to my supervisors, Prof. A. Uzairu and Dr. P.A.P Mamza for their guide throughout the course of the work.

I thank my mother, Mrs. Victoria Gowal and my siblings for their morale and financial support, I cannot thank them enough.

I wish to also express my sincere gratitude to Mr. Arthur Ebuka David who provided some of the software I used in this study.

To all the staff and students of the Department of Chemistry, Ahmadu Bello University, Zaria, I enjoyed a good learning and working relationship with you all. God bless you.

## DEDICATION

This research work is dedicated to God Almighty, without Him I would not have been able to achieve this feat.

# TABLE OF CONTENTS

**Cover Page** ..... **i**

**Title Page** ..... **iii**

**Declaration** ..... **iv**

**Certification** ..... **v**

**Acknowledgements** ..... **vi**

**Dedication** ..... **vii**

**Table of Content** ..... **viii**

**List of Tables** ..... **xii**

**List of Figures** ..... **xiv**

**Abbreviation** ..... **xvii**

**Abstract** ..... **xviii**

**1.0 INTRODUCTION** ..... **1**

**1.1 Background to the Study** ..... **1**

**1.1.1 History of tuberculosis** ..... **2**

**1.1.2 Global epidemiology of TB** ..... **3**

**1.1.3 Pathogenesis and transmission of TB.** ..... **5**

**1.1.4 Drug-resistant tuberculosis (MDR-TB and XDRTB)** ..... **6**

1.1.5 *In Silico* Pharmacology in Drug Discovery ..... 7

**1.2 The Research Problem** ..... **9**

**1.3 Justification for the Study** ..... **9**

**1.4 The Research Questions** ..... **10**

**1.5 Aim and Objectives of the Study** ..... **10**

**1.7 The Research Hypotheses .....11**

**1.8 Significance of the Study .....12**

**1.9 Scope and Limitation of the Study .....12**

**2.0 LITERATURE REVIEW .....13**

**2.1 Theoretical Frame Work……...13**

2.1.1 Theoretical background for calculation of the quantum descriptors……14

**2.2 Treatment of TB. .....16**

**2.3 Quinolones as Anti-Mycobacterial Agents .....20**

**2.4 General Structural Features of the Quinolones .....17**

**2.5 Activity of Quinolones against *Mycobacterium Tuberculosis* .....19**

**2.6 Use of Whole Infection Animal Model Testing in Drug Discovery .....22**

**2.7 QSARs in TB Drug Design……22**

**3.0 MATERIALS AND METHODS .....29**

**3.1 Materials .....29**

**3.2 Method .....29**

**3.2.1 Data set .....29**

**3.2.2 Molecular structure generation .....31**

**3.2.3 Geometry optimization and calculation of quantum-chemical descriptors .....49**

**3.2.5 Calculation of 0D,1D, 2D and 3D descriptors .....52**

**3.2.5 Descriptor Normalization .....58**

3.2.6 Correlation Matrix……...58

**3.2.7 Variable elimination and selection.....63**

3.2.8 Development of a quantitative structure- activity relationship model……63

**3.2.8a Training and test set selection .....65**

**3.2.8b Descriptor selection and Model Building.....65**

**3.2.8.c Internal Validation .....71**

**3.2.8d External Validation.....84**

**3.2.8e Y-Randomization.....85**

**3.2.8f Applicability domain.....89**

3.2.8g Generation of the Williams plot……..91

**4.0 RESULTS.....97**

**4.1 Result of Optimization and Calculation of Quantum-chemical Descriptors. .....97**

**4.2 Result of the Calculation of 0D,1D, 2D and 3D Descriptors Using Padel. .....97**

**4.3 Result of the Variable Elimination and Selection……...98**

**4.4 Results of 8-Methylquinolone Carboxylic against *Mycobacterium tuberculosis H37Rv* (MTB). .....98**

**4.5 Results of 8 -Methylquinolone carboxylic acid against *Mycobacterium Smegmatis* ATCC 14468 MC2.....110**

**4.6 Results of a mixture of 8-Methylquinolone Carboxylic acids and Nitroquinolones against *Mycobacterium tuberculosis H37Rv* (MTB).....112**

**4.7 Result of cytotoxicity (IC50) study of 8-Methylquinolones .....134**

**5.0 DISCUSSION.....144**

**5.1 Interpretation and evaluation of model 4.2.....144**

**5.2 Interpretation and evaluation of model 4.4.....146**

**5.3 Interpretation and evaluation of model 4.7.....149**

**5.4 Interpretation and evaluation of model 4.11.....151**

**6.0 SUMMARY,CONCLUSION AND RECOMMENDATION .....153**

**6.1 Summary.....153**

**6.2 Conclusion.....154**

**6.2 Recommendation .....155**

**REFERENCES .....157**

**APPENDIX.....167**

LIST OF TABLES

**Table 2.0 Some QSAR Studies Applied to TB (2008-2013). .....24**

**Table 2.0 continued…Some QSAR Studies Applied to TB (2008-2013) .....25**

**Table 3.1 Compounds and their activities .....33**

**Table 3.2 Data matrix is divided in to a training set and an independent test set.....66**

**Table 4.1 Fitting Parameters for the three selected Models .....100**

**Table 4.2 Golbraikh and Tropsha acceptable model criteria for the three Models .....101**

**Table4.3 Comparison of Observed Activity with Predicted Activity of the three models ..102**

**Table 4.4 Observed Activity versus Predicted Activity OF Model 4.2 .....103**

**Table 4.5 Observed Activity versus Predicted Activity (Test Sets) for model 4.2 .....105**

**Table 4.6 Y-Randomization Result for model 4.2.....107**

**Table 4.7 Correlation matrix among the descriptors in models 4.2 .....109**

**Table 4.8 Fitting Parameters for models 4.4, 4.5 and 4.6.....111**

**Table 4.9 Golbraikh and Tropsha acceptable model criteria Models 4.4,4.5 and 4.6.....113**

**Table 4.10 Comparison of Observed Activity with Predicted Activity between Models 4.4, 4.5 and 4.6.....114**

**Table 4.11 Observed Activity versus Predicted Activity OF Model 4.4 .....115**

**Table 4.12 Observed Activity versus Predicted Activity for Test Sets .....117**

**Table 4.13 Y-Randomization Result for model 4.4.....119**

**Table 4.14 Correlation matrix among the descriptors.....121**

**Table 4.15 Name of the various descriptors in the three model. .....122**

**Table 4.16 Fitting Parameters For models 4.7,4.8 and 4.9 .....124**

**Table 4.17 Golbraikh and Tropsha acceptable model criteria For Models 4.7,4.8 and 4.9. .....125**

**Table 4.18 Comparison of Observed Activity with Predicted Activity between Models 4.7, 4.8 and 4.9.....126**

**Table 4.19 Observed Activity versus Predicted Activity OF Model 4.7 .....127**

**Table 4.20 Observed Activity versus Predicted Activity For Test Sets for model 4.7.....129**

**Table 4.21 Y-Randomization Result for model 4.7.....131**

**Table 4.22: Correlation matrix for descriptors in models 4.7, 4.8 and 4.9.....132**

**Table 4.23 Names of the descriptors .....133**

**Table 4.24 Fitting Parameters Of Models 4.10, 4.11 and 4.12.....135**

**Table 4.25 Golbraikh and Tropsha acceptable model criteria for models 4.7.1, 4.7.2 and 4.7.3.....136**

**Table 4.26 Comparison of Observed Activity with Predicted Activity Of Models 4.10, 4.11 and 4.12.....137**

**Table 4.27 Observed Activity versus Predicted Activity for Model 4.11.....138**

**Table 4.28 Observed Activity versus Predicted Activity For Test Sets for models 4.11 .....140**

**Table 4.29: Y-Randomization Result for model 4.11 .....142**

**Table 4.30 correlation matrix for model 4.11. .....143**

## LIST OF FIGURES

**Fig 2.1a Quinolone nuclei .....19**

Fig 2.1b Naphthyridone nucle**i.....19**

**Fig 2.2: A Quinolone Molecule .....21**

**Fig 3.1: QSAR Workflow.....30**

**Fig 3.2 Moiety of the compounds .....32**

**Fig 3.3: Structure of Compound 1 .....44**

**Fig. 3.4 :Menu bar .....45**

**Fig 3.5 Sketch Pad .....46**

**Fig. 3.6 Tool Bar .....50**

**Fig. 3.7 Calculations Dialog Box.....51**

**Fig.3.8 File browser .....53**

**Fig 3.9 Padel Descriptor Snapshot-A.....54**

**Fig 3.10 File Browser-A.....55**

**Fig 3.11 Padel Descriptor snapshot-B.....56**

**Fig 3.12 Padel Descriptor Snapshot-C.....57**

**Fig. 3.13 Normalization Dialogue box.....59**

**Fig 3.14 Excel Worksheet.....60**

**Fig. 3.15 Data Analysis Tools.....61**

**Fig 3.16 Correlation Dialogue Box.....62**

**Fig 3.17 Queries Window.....64**

**Fig. 3.18 BuildQsar interface.....68**

**Fig. 3.19 Create Dataset Dialogue Box .....69**

**Fig.3.20 BuildQsar Interface-B .....70**

**Fig 3.21 Build Genetic Search Dialogue Box.....72**

**Fig3.22 Welcome to material studio Dialogue box .....74**

**Fig: 3.23 New Project File Browser .....75**

**Fig 3.24 Material Studio Dialogue Box.....76**

**Fig 3.25 New Document Dialogue Box.....77**

**Fig. 3.26 Material Studio Study Table-A.....78**

**Fig 3.27 Material Studio Study Table-B.....79**

**Fig 3.28 Material Studio Study Table- C.....80**

**Fig 3.29 Multiple Linear Regression Dialogue Box-A.....81**

**Fig. 3.30 Multiple Linear Regression Dialogue Box-B.....82**

**Fig 3.31 Multiple Linear Regression Dialogue Box-C.....83**

**Fig 3.32 DTC MLRplusValidation Dialogue Box-A.....86**

**Fig 3.33 DTC MLRplusValidation Dialogue Box-B.....87**

**Fig. 3.34 DTC MLRplusValidation Dialogue Box-C.....88**

**Fig. 3.35 Excel Worksheet- A .....92**

**Fig 3.35 Paste special Dialogue Box .....93**

**Fig 3.36 Excel sheet-B.....94**

**Fig 3.37 Excel Worksheet-C .....96**

**Fig 4.1: Plot of observed pMIC against predicted pMIC of model 4.2.....104**

**Fig 4.2 Plot of observed pMIC against predicted pMIC of model (Test Set) model 4.2 .....106**

**Fig.4.3 The Williams Plot for Model 4.2.....108**

**Fig. 4.4 Plot of observed pMIC against predicted pMIC of model for model 4.4 .....116**

**Fig 4.5 Plot of observed pMIC against predicted pMIC (Test Sets) for model 4.4 .....118**

**Fig. 4.6 Williams plot for model 4.4 .....120**

**Fig 4.7: Plot of observed pMIC against predicted pMIC of model for model 4.7 .....128**

**Fig. 4.8: Plot of observed pMIC against predicted pMIC (Test Sets) for model 4.7 .....130**

**Fig 4.9: Plot of observed pMIC against predicted pMIC of model for model 4.11 .....139**

**Fig 4.10 Plot of observed pMIC against predicted pMIC (Test Sets) for model 4.11 .....141**

## LIST OF ABBREVIATIONS

2D QSAR 2 Dimensional Quantitative Structure-Activity Relationship

3D QSAR 3 Dimensional Quantitative Structure-Activity Relationship

ADME/Tox Absorption Distribution Metabolism Elimination Toxicity

DFT Density Functional Theory

GA Genetic Algorithm

HOMO Highest Occupied Molecular Orbital

LOO Leave-One-Out

LUMO Lowest Unoccupied Molecular Orbital

MIC Minimal Inhibitory Concentrations

$IC_{50}$ Half maximal inhibitory concentration

MLR Multiple Linear Regression

TB Tuberculosis

MTB *Mycobacterium Tuberculosis*

OECD Organization for Economic Cooperation and Development

PCA Principal Component Analysis

PSA Polar Surface Area

QSAR Quantitative Structure-Activity Relationship

pMIC Logarithm to base 10 of the minimum inhibitory concentration

pIC50 Logarithm to base 10 of half maximal inhibitory concentration

MMFF Molecular mechanics forcefield

B3LYP Becke's three-parameter hybrid functional

Cpd Compound

## ABSTRACT

A series of quinolone derivatives mainly 8-mehtylquinolones and Nitroquinolones were subjected to QSAR studies to correlate and predict their *in-vitro* biological activity- Minimal Inhibitory Concentration (MIC) - against *Mycobacterium tuberculosis* H37Rv (MTB) and *Mycobacterium Smegmatis* ATCC 14468 (MC2) as well as to predict their cytotoxicity in a mammalian vero cell reported as ($IC_{50}$). The molecular geometries of all the structures were subjected to energy minimization using molecular mechanics (MM2) and then re-optimized by the density functional theory (DFT) with Becke's three-parameter hybrid functional using LYP correlation functional and the standard Pople's 6-311G* basis set. In Addition to PADEL descriptors, the parameters from quantum-chemical calculations were used. The genetic algorithm (GA) and multiple linear regression analysis (MLRA) were used to select the descriptors and to generate the correlation models that relate the structural features to the biological activities respectively. The final models were tested internally using the Leave-One-Out cross-validation and validated against the test set. Result of the QSAR study of 8-methylquinolone Carboxylic acids against *Mycobacterium tuberculosis* H37Rv shows that the best model has $R^2 = 0.90323$; $R^2_{adjs} = 0.88123$; $F = 41.0676$; $S = 0.2301$ $Q^2_{LOO} = 0.8311$; $R^2_{Pred} = 0.78708$. Result of the QSAR study of 8-Methylquinolone carboxylic acid against *Mycobacterium Smegmatis* ATCC 14468 (MC2) shows the best model to have $R^2 = 0.9184$; $R^2_{Adjs} = 0.8990$; $F = 47.28523$; $s = 0.18902$. The model has $Q^2_{LOO} = 0.84987$. It was externally validated by testing it on the test set and performed well with $R^2_{Pred} = 0.79343$. Results of the QSAR study of combination of 8-Methylquinolone Carboxylic acids and Nitroquinolones against *Mycobacterium tuberculosis H37Rv* (MTB) shows the best model have the following statistical parameters $n = 35$; $R^2 = 0.90178$; $s = 0.22292$; $F = 42.84356$; $p < 0.0001$; SPress = 2.46008;

SDEP = 0.26512. The $Q^2_{LOO}$ = 0.82634 shows that this model is internally stable. The model was externally validated and found to have good predictive ability with $R^2_{pred}$ of 0.77462. The result of QSAR study of the *in-vivo* cytotoxicity ($IC_{50}$) study of 8-Methylquinolones carboxylic acids shows the best model to have $R^2$ = 0.97684; F =147.63356; S = 0.00716 SPress= . 0.00132 The internal validation shows the model to be internally stable with $Q^2_{LOO}$ = 0.91459. The external validation shows this model to have an excellent predictive ability with $R^2_{Pred}$ = 0.97829. The Y-randomization shows the model to have $R^2_p$ = 0.868807. Analysis of the descriptors in the final models led to the conclusion that the factors that influence the development of the activity of the active compounds analyzed against *Mycobacterium tuberculosis* H37Rv and *Mycobacterium smegmatis* ATCC 14468 (MC2) are principally steric and electronic in nature. The most important descriptor for modeling the cytotoxicity of this group of compounds was **AATS6p** Average Broto-Moreau autocorrelation - lag 6 / weighted by polarizabilities. **SpDiam_Dzv (**Spectral diameter from Barysz matrix / weighted by van der Waals volumes. This is a weighted distance matrix accounting for the presence of heteroatoms and multiple bonds)

# CHAPTER ONE

## 1.0 INTRODUCTION

### 1.1 Background to the Study

Despite the fact that tuberculosis (TB) has been recognized for thousands of years and despite the fact that the etiological agent has been identified since earliest days of medical biology, the global burden of TB continues to loom as one of the largest among infectious diseases, with an enormous toll in morbidity and mortality (Leonard and Rachael,2009).This disease which is caused by various strains of mycobacteria, usually *Mycobacterium tuberculosis* (MTB*)* isthe leading cause of death from a single infectious agent, after the human immuno deficiency virus (HIV) (WHO, 2013). Currently, one-third of the world harbors the latent form of *Mycobacterium tuberculosis*, with a lifelong risk of activation and disease development, particularly in people co-infected with HIV (Anil *et al*., 2011). The World Health Organization (2014) reported 1.5 million deaths from tuberculosis globally in 2013 andout of this number, 360,000 were HIV-positive. Ironically, available therapeutic agents are highly efficacious in TB treatment, withcure exceeding 95% during clinical trials (Frieden, 2003). The natural question is; why then has the control of TB been so problematic? The answer lies in the indolent clinical nature of the disease, whichneeds prolonged complex therapy and the unusual microbiological properties of the pathogen(Leonard and Rachael, 2009). The actual drug therapy for tuberculosis involves the administration of multiple drugs because it was clear that monotherapy led to the development of resistance (Barry and Blanchard, 2010).The current short-course TB therapy used to treat drug-susceptible MTB consists of months of treatment with four so-called first-line drugs including rifampin (RIF), isoniazid (INH),pyrazinamide (PZA) and ethambutol (EMB), followed by 4-

months treatment with RIFand INH (Sunduru *et al*.,2010). Infection by multidrug resistant tuberculosis (MTR-TB) strains requires treatment with second-line drugs such askanamycin, amikacin, capreomycin,*p*-aminosalicylate (PAS), fluoroquinolones(levofloxacin, gatifloxacin and moxifloxacin), ethionamide (ETH), and cycloserine wheretreatments often extend for up to long 2 years (Janin, 2007). This Chemotherapy is less effective, longer, expensive and more toxic than the short course therapy (Ma *et al*., 2010). Third line drugs include rifabutin, macrolides (clarithromicin), linezolid, thiacetazone, thioridazine, arginine, vitamin D are still being developed, have less or unproven efficacy and are very expensive (Lalloo and Ambaram, 2010).

### 1.1.1 History of tuberculosis

Tuberculosis (TB) is a disease of antiquity which is thought to have evolved sometime between the seventh and sixth millennia BC (Manchester, 1984). MTB may have killed more persons than any other microbial pathogen.In fact,it reached epidemic proportions in Europe and North America during the 18th and 19th centuries, earning the sobriquet, ‘‘Captain Among these Men of Death.’’. The foundation for understanding the pathogenesis of tuberculosis was laid at the beginning of the $19^{th}$ century through the work of The’ophile Laennec. This was further advanced by the demonstration of the transmissibility of Mycobacterium tuberculosis infection by Jean-Antoine Villemin in 1865 and the identification of the tubercle bacillus as the etiologic agent by Robert Koch in 1882.

Clemens von Pirquet developed the tuberculin skin test in 1907 and 3 years later used it to demonstrate latent tuberculosis infection in asymptomatic children (Daniel, 2006). The Modern regime of tuberculosis treatment and control began with the discovery and introduction of the first drugs for TB treatment some 60 years ago – streptomycin in 1944 and isoniazid in 1952. This led to optimism that the disease could be controlled if not eradicated (Dye *et al*., 1999) and

true to this optimism there was a rapid decline in incidence rate of tuberculosis in many industrialized countries. However, since the 1980's the disease has been undergoing a resurgence driven by a variety of changes in social, medical and economic factors. In conjunction with the spread of HIV infection, tuberculosis is today amongst the worldwide health threats (Hemant *et al*., 2012).

### 1.1.2 Global epidemiology of TB

The foundation of the current global tuberculosis (TB) eradication strategy began in the 1990s, when the increasing trends of TB led to the creation of directly observed treatment, short course (DOTS) strategy (Raviglione and Pio, 2002; WHO, 1991). In January 2006, the Stop TB Partnership launched its global plan to stop TB 2006–2015 to tackle these issues (Raviglione, 2007). This 6-part strategy detailed a comprehensive approach to decrease by half the 1990 death and prevalence rates of TB by 2015. The impact of current TB-control methods is measured by incidence, prevalence, and mortality indicators, which frame the ultimate aims of TB control, as determined by the Millennium Developmental Goals and Stop TB Partnership (WHO, 2014). At a global decrease rate at slightly 1% per year, Mandeep *et al*., (2010) submitted that it is highly unlikely that the goal of TB elimination by 2050 was going to be achieved. In 2010, there were an estimated 8.8 million incident cases of TB (range, 8.5 million–9.2 million) globally,equivalent to 128 cases per 100 000 population (WHO, 2011).

As in previous WHO TB reports most of the estimated number of cases in 2010 occurred in Asia (59%) and Africa (26%) The five countries with the largest number of incident cases in 2010 were India(2.0 million–2.5 million), China (0.9 million–1.2 million), South Africa (0.40 million–0.59 million), Indonesia (0.37 million–0.54 million) and Pakistan (0.33 million–0.48million). India alone accounted for an estimated one quarter (26%) of all TB cases worldwide, and China

and India combined accounted for 38%.The prevalent cases of TB in 2010 stood at 12.0 million (range,11.0 million–14.0 million) which is equivalent to 178 cases per 100 000 population. Approximately 1.4 million people (range, 1.2 million– 1.5 million) died of TB in 2010 (WHO, 2011).

In 2012, there were an estimated 8.6 million incident cases of TB (range, 8.3 million–9.0 million) globally, equivalent to 122 cases per 100 000 population whilethe estimated prevalence cases were 12 million (range, 11 million–13 million), equivalent to 169 cases per 100 000 population. There were an estimated 1.3 million TB deaths in 2012: 940 000 among HIV-negative people and 320 000 among HIV-positive people. These deaths included 410 000 women and 74 000 children(WHO, 2013)

Finally, a 2014 report on global TB control published by the WHO which describes the most recent figures available on TB epidemiology revealed that there were an estimated 9.0 million incident cases of TB (range, 8.6 million–9.4 million) globally in 2013, equivalent to 126 cases per 100 000 population. The six countries that stand out as having the largest number of incident cases in 2013 were India (2.0 million–2.3 million), China (0.9million–1.1 million), Nigeria (340 000−880 000), Pakistan (370 000−650 000), Indonesia (410 000−520 000) and South Africa ( 410 000−520 000). There were an estimated 11 million prevalent cases(range, 10 million–13 million) of TB in 2013 equivalent to 159 cases per 100 000 population. The 2013 statistics on the prevalent cases shows that the prevalence rate had fallen 41% globallysince 1990. Current forecasts suggest that the StopTB Partnership target of halving TB prevalence by 2015compared with a baseline of 1990 will not be met worldwide. There were an estimated 1.5 million TB deaths in 2013; 1.1 million among HIV-negativepeople and 360 000 among HIV-positive people these deaths included 510 000women and 80 000 children(WHO, 2014).

In conclusion, although large numbers of TB cases and deaths have been recorded in recent past, 21 years on from the 1993 World Health Organization (WHO) declaration of TB as a global public health emergency, major progress has been made. Globally, the TB mortality rate (deaths per100 000 population per year) has fallen by 45% since 1990 and TB incidence rates (new cases per 100 000 population per year) are decreasing in most parts of the world. Between 2000 and 2013, an estimated 37 million lives were saved through effective diagnosis and treatment.

**1.1.3 Pathogenesis and transmission of TB**.

Mycobacteria are non-motile, non sporulating, weakly gram positive, acid-fast bacilli that appear microscopically as straight or slightly curved rods, 1 to 4 mm in length and 0.3 to 0.6 mm wide. These bacteria also express unique mycolic acids in the cell envelope that play a critical role in the structure and function of the cell wall (Barry *et al*., 1998). The waxy cell wall is responsible for many of the unique characteristics of this genus which includes acid-fastness, extreme hydrophobicity, resistance to drying, acidity/alkalinity, and many antibiotics, as well as distinctive immunostimulatoryproperties. (Daffe and Draper, 1998). MTB is a member of the *M. tuberculosis* complex, which is defined as the etiologic agents of TB in distinct hosts, and also includes *M. bovis, M. africanum*, *M. canetti*, and *M. microti*,with *M. caprae* and *M. pinnipedii* considered variants of *M. bovis*. It was previously assumed that MTB evolved from *M. bovis* during the domestication of cattle (Stead *et al.,* 1995).

Human MTB infections usually begin by inhalation of aerosol droplets containing tubercle bacilli directly expectorated from an individual with ‘‘open’’ pulmonary disease. The infectiousdose for a person is reported to be between 1 and 200 bacilli; however, as a single aerosol droplet can contain anywhere from 1 to 400 bacilli, it is unclear what is considered a biologically relevant dose (Sakamoto, 2012).

TB infection begins when the tubercle multiply in small air sacs in the lungs, a small number enters the blood stream and spread throughout the body but the body's immune system usually keeps the bacilli under control. People who have latent TB infection, however not TB disease do not show symptoms of TB and cannot transmit it to others. However, when the immune system of an individual with latent TB cannot keep the tubercle bacilli under control and the bacilli begins to multiply and causes TB disease. This can happen immediately after TB infection or many years after infection. TB disease occurs in the lungs (Pulmonary TB) but can also occur in other parts of the body (extra pulmonary TB). Miliary TB occurs when tubercle bacilli enter the blood stream and are carried to all parts of the body, where they grow and cause disease in multiple sites (Global Tuberculosis Institute, 2015).

**1.1.4Drug-resistant tuberculosis (MDR-TB and XDRTB)**

The term multidrug-resistance tuberculosis (MDRTB) refers to simultaneous resistance to at least two or more of the five first-line anti-TB drugs (isoniazid, rifampicin, pyrazinamide, ethambutol, and streptomycin) (Ahmed *et al*.,2008).Multi-drug-resistance arises from sharing of genes between different species or genera, generally mediated by small pieces of extra-chromosomal DNA known as transposons or plasmids (Dye *et al*., 1999). Treatment for multidrug-resistant tuberculosis is long lasting, less effective, costly, and poorly tolerated.Globally, 3.5% of new and 20.5% of previously treated TB cases were estimated to have had MDR-TB in 2013. This translates into an estimated 480 000 people having developed MDR-TB in 2013 (WHO, 2014).

Extensively drug resistant tuberculosis (XDR-TB) by definition is resistance to at least isoniazid and rifampicin in addition to any quinolone and at least one injectable second-line agent (capreomycin, amikacin, and kanamycin) (Ahmed *et al*., 2008). On average, an estimated 9.0%

of patients with MDRTB had extensively drug resistant TB (XDR-TB) in 2013 (WHO, 2014). The principles of treatment for MDR-TB and XDR-TB are the same; the main difference is that XDR-TB is associated with a much higher mortality rate than MDR-TB, because of reduced number of effective treatment options. Hence there is an urgent need for novel drugs that are active against *M. tuberculosis* in order to shorten the duration of tuberculosis therapy.(Ahmed *et al*., 2008)

### 1.1.5*In Silico* Pharmacology in Drug Discovery

The term ‘in silico’ is a modern word usually used to mean experimentation performed by computer and is related to the more commonly known biological terms *in-vivo* and*in-vitro*. *In silico* pharmacology (also known as computational therapeutics, computational pharmacology) is a rapidlygrowing area.More specifically, it defines the use of this informationin the creation of computational models or simulations that can be used to make predictions, suggest hypotheses, and ultimately provide discoveries or advances in medicineand therapeutics. A lot of information needs to be processed during drug discovery, thus there is a need to be able to discover shortcuts or the rules that will point as quickly as possible to the targets and molecules that are likely to proceed to the clinic then onto the market. Computational or *in silico* methods help us to make decisions and simulate virtually every facet of drug discovery and development (Swaan and Ekins, 2005). In the hit identification phase, drug discovery teams are provided with many novel chemicals in other to find potential lead molecules that possess the desired drug properties. (Sofie, 2009). This challenging tasked makes the benefits of *in silico* method immediatelyobvious. Quantitative Structure Activity Relationships (QSAR), the *in silico* method used in this study, is able to elucidate activities in a fast manner thereby shortening the cycle time needed to identify compounds that possess the desired drug like properties.

QSARs are mathematical models that attempt to relate the structure-derived features of a compound to its biological or physicochemical activity (Yew *et al*., 2011). It is based on the assumption that the structure of a molecule (for example, its geometric, steric, and electronic properties) must contain features responsible for its physical, chemical, and biological properties and on the ability to capture these features into one or more numerical descriptors QSAR models can be used to predict the biological activity (or property, reactivity, etc.) of a new designed or untested chemical from the molecular structure of similar compounds whose activities (properties, reactivities, etc.) have already been assessed (Todeschini and Consonni, 2000). In other to make an *in silico* prediction of an untested molecule, an atomistic model is built using available experimental data. These models must be built using some representation of a molecule's physical properties and its chemical characteristics, which is then related to their biological activity. *In silico* models can in principle be built using 1D representations of molecules including descriptors such as molecular weight or atom counts, 2D descriptors, which incorporate information about a molecule's topological structure, such as predicted lipophilicity, connectivity indices or fingerprints, or 3D descriptors such as pharmacophoric points, 3D surfaces or grids. The relationship between structure and response are then fitted (or trained) using the available dataset which can be achieved using relatively simple methods such as "read-across" or multiple linear regression to more complex non-linear machine learning methods such as non-linear Support Vector Machines (SVM) or Artificial Neural Networks (NN).

## 1.2The Research Problem

The increasing problem of multi-drug resistant strains is the major challenge for the investigation and designs of novel drug candidates which are not only active against stable drug resistant MTB

but also shorten the length of therapy.However, the daunting cost of drug development both in terms of money and time has stalled TB drug development. QSAR-the method employed in this study- is a very veritable tool in cutting down both the cost of drug production and the length of time involved since it can predictif a  molecule can be a good drug candidate or not prior to its synthesis

### 1.3 Justification for the Study

A more recent and alarming global public health threat has emerged with the identification of extensively drug-resistanttuberculosis (XDR-TB) – a highly pathogenic strain of TB developing from MDR-TB (Raviglione and Smith, 2007). This has raised the fear that a relapse into a situation similar to the pre-antibiotic TB era is likely (Anil*et al*., 2011). To achieve global control of this epidemic, there is an urgent need to develop novel, affordable anti-tubercular drugs with high potency(Kumar*et al*., 2011) and less toxic effect which can among other things

- Reduce the time of treatment.
- Provide treatment options for all forms of TB.
- Simplify treatment by reducing the daily pill burden.
- Reducing the dosage frequency, example once in a week regiment.
- Is co- administered with HIV medications.

The process of drug development is a herculean task. This is not only because of the length of time it takes which is an average of 12years,but also because of high cost, somewhere around USD868million per marketed drug (Dimasi *et al.*, 2003). The high cost and lengthy time for drug development is due to the high risk of drug failure. A study of cases of drug attrition during development in year 2000 revealed that 10% of drug attrition was a result of poor

pharmacokinetic and bioavailability while in the chemical stage,another 30% was caused by toxicity or clinical safety issues (Kola and Landis,2004). This underscores the need to predict these failures prior to the clinical stage, so as to reduce drug development cost. In other to develop a novel anti tuberculosis drug which will embody one or more of the characteristics mentioned above, and at the same time reduce the cost and length of time of development, there is a need to adopt a more rational approach to drug development. As has been discussed earlier, the Quantitative-Structure Activity Relationship (QSAR) is the method of choice here.

## 1.4 The Research Questions

The questions this study tries to answer include

i) What are the structural features(descriptors) of the class of quinolone compounds used in this study thatcorrelates with the biological activities of interest - potency and cytotoxicity?

ii.) Which of these structural features correlates best with the biological activities?

iii.) How can these descriptors which correlate highly with the biological activitiesbe used to develop mathematical models for predicting these biological activities for untested compounds?

## 1.5 Aim and Objectives of the Study

This study is aimed at using QSAR methodology to develop mathematical models which could be used to predict the anti-tuberculosis potency and toxicity of molecules within the family of quinolones specifically, 8-methylquinolones and Nitroquinolones, with the hope of discovering highly potent and non-toxic anti-tuberculosis drug candidates. This shall be achieved through the following objectives:

I) To collect from literature,quinolone compounds with experimentally measured cytotoxicity($IC_{50}$)and anti-tubercular as activity minimum inhibitory concentration (MIC).

II) To optimize the structures of the compounds at the DFT level of theory.

III) To calculate the respective molecular descriptors of the compounds using Spartan14 and Padel software, then find out which of these calculated molecular descriptors correlatesbest with the biological activities of interest.

IV) To established a mathematical relationship in form of an equation, between the molecular descriptors and the biological activities of interest.

V) To use the equations in predicting the biological activities of interest for compounds that has not yet been tested.

**1.6The Research Hypotheses**

The study was guided by the following hypotheses

a) The null hypothesis ($H_0$) which states that the anti-tuberculosis potency (MIC) and the cytotoxicity ($IC_{50}$) of the quinolone compounds in this study are determined by factors other than their molecular descriptors.

b) The alternate hypothesis ($H_1$) which states that the anti-tuberculosis potency (MIC) and cytotoxicity ($IC_{50}$) of the quinolone compounds in this study are a function of their molecular descriptors.

**1.7 Significance of the Study**

The study was aimed at establishing the relationship between the anti-tuberculosis potency and cytotoxicity of compounds within the 8-methylquinolones and Nitroquinolones congeneric series

and their molecular structures. This relationship can be exploited for predicting the suitability of molecules within this class of compounds to serve anti-tuberculosis drugs, prior to the synthesis and screening of these molecules; this will reduce drug development cost and failure during development. These prediction techniques could be used as a replacement for animal testing.

### 1.8 Scope and Limitation of the Study

This study focused on developing valid models capable of predicting the in-vitro antimycobacterial activity (MIC) of some quinolones against *Mycobacterium tuberculosis H37Rv* (MTB) and *smegmatis* ATCC 14468 (MC2) as well as their cytotoxicity. These models could subsequently be used to design new potential anti-tuberculosis drug candidates with higherpotency and less toxicity.However, molecular modelers typically analyze data generated byother (experimental) researchers. Consequently, when itcomes to the experimental data quality they are always atthe mercy of the data providers, the availability of these quality experimental data was a major challenge during this study.

## CHAPTER TWO

## 2.0 LITERATURE REVIEW

### 2.1 Theoretical Frame Work

After nearly five decades ‘in the making’, QSAR modeling has established itself as one of the major computational molecular modeling methodologies. As any mature research discipline, QSAR modeling can be characterized by a collection of well-defined protocols and procedures that enable the expert application of the method for exploring and exploiting ever growing collections of biologically active chemical compounds (Alexander, 2000).The QSAR method involves recognizing that a molecule is really a three dimensional distribution of properties (Winkler,2002).The development of any QSAR model involves

1. Collection of data for property of interest while taking into consideration the quality of the data
2. Converting molecular structures into mathematical descriptors that encapsulates the key properties of the molecules relevant to the activity or property being modeled (Winkler, 2002). Molecular descriptors are results of a logical and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into useful numbers or the result of some standardized experiment (Todeschini andConsonni, 2000). Some of the earliest descriptors used in QSAR include molecular weight or molecular volume and log of the octanol-water partition coefficient (Log P). Log P was found to be important in many QSAR studies. Another important class of molecular descriptors is the topological descriptors. This involves treating molecules as topological objects where atoms become vertices and bonds the edges. It is possible to characterize molecular graphs using a number of indices. The most well-known of this are the Randic indices and Kier and Hall electropological indices (Winkler, 2002).

The topological descriptors encode 2D representation of molecules. Several other molecular descriptors are calculated from 3D properties of molecules, the coMFA (comparative field

analysis surrounds 3D structures of molecules by an array of grid points. QSAR models derived from 3D molecular descriptors are more easily interpreted than are those from 2D descriptors (Winkler, 2002). They are able to account for the effect of conformation and isomerism on activity.

1. Splitting of the data set into training and test sets.
2. Selecting the best descriptors from a larger set of accessible relevant descriptors.
3. Developing the model using relevant statistical methods such as multi linear regression analysis (MLR) etc.
4. Validating the model to determine how predictive it is, and how well it can generalize to new molecules not in the data set used to develop the model.

### 2.1.1 Theoretical background for calculation of the quantum descriptors

For a molecule of N electrons, the electronic chemical potential $\mu$ is defined as (Todeschini and Consonni, 2000)

$$\mu = \left(\frac{\partial E}{\partial N}\right)_{V(\vec{r})} \dots\dots(1)$$

Iczkowski and Margrave (1961) identified chemical potential (μ) as the negative of the electronegativity ($\chi$) and defined as

$$\chi = -\mu = \left(\frac{\partial E}{\partial N}\right)_{V(\vec{r})} \dots\dots.(2)$$

The quantitative definition of hardness ($\eta$) of an N-electron system with total energy E and external potential v(r) using DFT can be expressed (Parr and Pearson, 1983) as:

$$\eta = \frac{1}{2}\left(\frac{\partial^2 E}{\partial N^2}\right)_{V(\vec{r})} = \frac{1}{2}\left(\frac{\partial \mu}{\partial N}\right)_{V(\vec{r})} = \int h(r)\, dr = \frac{1}{2.S} \dots\dots\dots(3)$$

where $\eta$ is the electronic chemical potential, h(r) is called local hardness (or hardness density),and S is the total softness(defined by equation 9). The global electrophilcity index ($\omega$) which has been proposed as a measure of the electrophilic power of the molecule and called electrophilicity index (Parr, *et al*., 1999; Chattaraj and Roy, 2007) is expressed in terms of chemical potential and hardness

$$\omega = \frac{\mu^2}{2\eta} \ldots\ldots\ldots\ldots (4)$$

According to the finite difference approach, global hardness and chemical potential can be approximated (Todeschini and Consonni, 2000) as:

$$\text{Chemical potential} \quad \mu = -\left(\frac{IP+EA}{2}\right) \ldots\ldots\ldots (5)$$

$$\text{Global hardness} \qquad \eta = \frac{IP-EA}{2} \ldots\ldots\ldots\ldots (6)$$

where IP and EA, are the first vertical ionization potential and electron affinity of the chemical system, respectively. Further approximation using Koopmans' theorem (Koopmans, 1933), the above parameters can be expressed by taking IP and EA as negative of the HOMO and LUMO energies, respectively.

$$\mu = \frac{E_{LUMO} + E_{HOMO}}{2} \ldots\ldots\ldots\ldots (7)$$

$$\eta = \frac{E_{LUMO} - E_{HOMO}}{2} \ldots\ldots\ldots\ldots (8)$$

Where $E_{LUMO}$ is the energy of the lowest unoccupied molecular orbital and $E_{HOMO}$ is the energy of the highest occupied molecular orbital.The total softness S is defined as

$$S = \left(\frac{\partial N}{\partial \mu}\right)_{V(\vec{r})} = \int S(r)\, dr = \frac{1}{2\eta} \ldots\ldots\ldots\ldots (9)$$

**2.2 Treatment of TB.**

Effective chemotherapy for tuberculosis started in the 1940s with the discovery and use of streptomycin (STR) and para-aminosalicylic acid (PAS)(Iseman 2002, Murray 2004, Thoren and Hinshaw 1952). However, streptomycin was effective in the short term but eventually so many patients developed STR-resistant TB andhearing loss (Marshall, 1948).Experimental models showed that PAS was bacteriostatic against MTB (including STR-resistant strains) and was able to prevent the development of STR resistance (Youmans *et al.,* 1947;Graessle and Pietrowski, 1949). Another advantage of PAS was that it could be administered to pulmonary TB patients as monotherapy. However the development of resistanceand poor tolerance of the drug mainly due to gastrointestinal side effects and occasionally hepatitis (Dooneief *et al*., 1950) was a major drawback.

The 1950s saw the discovery and use of isonizaid (INH), although there was a very significant tolerance of this drug in patients, some experienced rash or hepatitis. Several attempts were made to optimize treatment combinations of INH with STR and PAS. INH treatment led to rapid improvement over the first month of therapy, but recrudescence of disease was common due toacquired resistance (Bowen and Collins 1952;Joiner *et al*., 1954). It was shown that combination of INH with STR was superior to the combination of INH with PAS as measured by radiographic, microbiologic, and clinical investigations. In addition, such studies showed that STR was more effective than PAS in preventing the emergence of INH resistance (Marshall,1955). The triple drug combination of INH, STR, and PAS was found to be better than therapy with INH and PAS combined (Capon, 1954). Other drugs discovered during the 1950s include cycloserine (Hudgins *et al.,* 1955), ethionamide and the closely related prothionamide (Hutton *et al.*,1960), viomycin (Bartz *et al*., 1951), kanamycin (Patnode and Hudgins, 1958), and pyrazinamide (PZA) (Kaida and Sugiyama., 1959). At the time of their discovery, these drugs

were thought to be inferior to INH, PAS, and STR and were used only in patients with disease refractory to standard therapy (Somner and Brace, 1962).

TB drugs introduced during the 1960s include thiacetazone (Bienenstock and Shaldon, 1963), capreomycin (Cuthbert and Bruce 1964), and clofazimine Ethambutol (EMB) supplanted PAS in thestandard drug regimen since this drug was better tolerated than PAS and alsoallowed the treatment regimen to be shortened to 18 months(Doster*et al*., 1973; Lees *et al*., 1970).Rifampicin (RIF), was one of the last drugs to be introduced into clinicalpractice. It revolutionized TB therapy (Fisher, 1971).Watershed clinical trials during the 1970s inEast Africa and Hong Kong revealed that introduction of RIF to the standard INH/EMB/STR or INH/STR drug plan allowed the time of treatment to be reducedfrom 18 to 9 months without increasing the relapse rate (Mitchison 2005 andControlled clinical trial of four short-course (6-month), 1974). At the end of the 1970s and in the 1980s clinical trials were conducted to investigate the use of PZA in differentcombinations and treatment time with STR, INH, RIF, EMB, and thiacetazone.PZA was responsible for decreasing TB chemotherapy to 6 months(Long-term follow-up of a clinical trial, 1986). The culmination of these studies was the introduction of modern "short-course" chemotherapy for drug susceptible TB where PZA forms part of the drug regimenfor the first 2 months (the "intensive phase") in combination with INH, EMB, andRIF, but then is not included in a subsequent continuation phase of 4 months ofadditional treatment with INH and RIF (sometimes in combination with EMB) toobtain a cure rate greater than 95% (Mitchison., 2005). Contemporary treatment is exceedinglylengthy and takes 6-9 months (Blumberg*et al*., 2003). Currently, TB chemotherapy is made up of a cocktail of first-line drugs, isoniazid (INH), Rifampicin (RIF), pyrazinamide (PZA) and ethambutol (EMB), which are given for six months (Blumberg *et al*., 2006). If this treatment fails as a result of bacterial drug resistance or

intolerance to one or more drugs, second-line drugs such as *para*-aminosalicilate (PAS), kanamycin, fluoroquinolones, capreomycin, ethionamide and cycloserine are used. These are generally less effective or more toxic with serious side effects.

**2.3 Quinolones as Anti-Mycobacterial Agents**

Quinolones and their derivatives occur in excellent anti-bacterial agents (Chu and Fernandes 1991; Crumplin and Smith, 1975) they are active against aerobic gram-negative microorganisms but less active against gram-positive microorganisms (Boteva and Krasnykh, 2009; Chu and Fernandes, 1991). They are extremely useful for the treatment of a variety of infectious diseases.The first quinolone, nalidixic acid, (1-ethyl-7-methyl-1,4-dihydro-4-oxo-1,8-naphthyridine-3-carboxilic acid , was isolated as a by-product of the synthesis of chloroquine by George Lesher and his coworkers in 1962. It was found to be effective in the treatment of treatment of urinary tract infections (UTIs)(Lescher*et al*., 1962; Norris and Mandell, 1988).

**2.4General structural features of the quinolones**

Quinolone agents basically possess a bicyclic aromatic core; this can contain a carbon at the 8-position, yielding a true quinolone (Fig.2.1a), or nitrogen which provides a ring system technically termed a naphthyridone (Fig 2.1b).It is common practice, however, for both quinolone and naphthyridone structures to be encompassed in the class descriptor "quinolone antibacterial agents." The pyridone ring on the right-hand side is an essential necessity for antibacterial activity (Gootz andBrighty, 1998).

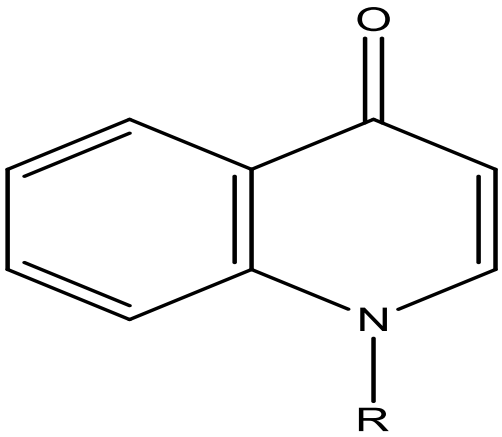



**Fig2.1aQuinolone nuclei**

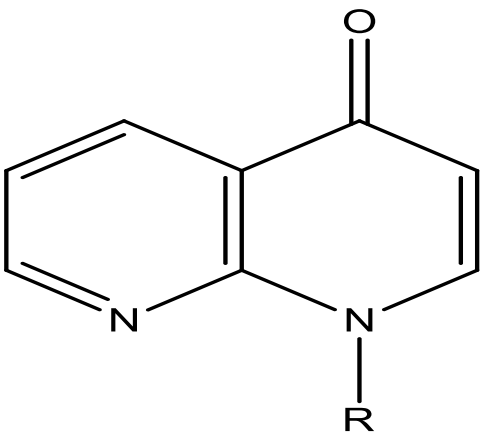



**Fig 2.1b Naphthyridone nuclei**

According to structure-activity relationship (SAR) studies, the carboxylic acid at the 3-position, the ketone at C-4 (see Fig. 2.2) and the R1-substituted nitrogen at the 1-position are all essential. Substitution at C-2 is generally deleterious, although exceptions exist in which the C-2 substituent forms a ring with R1. In the left-hand ring, the fluorine at C-6 is found in essentially all the modern agents; it is because of this substituent that the class is often referred to as the fluoroquinolones. A cyclic diamine R2 is most often present, attached through one of its nitrogens to the 7-position.Some variation is permitted at C-5 and the 8-position. Although the older quinolone agents were generally un-substituted at C-5 (R3 = hydrogen), several more recent compounds bear small substituents such as amino or methyl at this site. At the 8-position, when X = carbon, a number of small substituents such as fluorine, chlorine, and methoxy have been found to provide improved potency(Gootz and Brighty, 1998).

Among the second-generation quinolones,ciproflaxen and oxoflaxin have received examination for utility against *M.tuberculosis*( Gootz and Brighty, 1998), but newer agents like Sparfloxacin, BAY y 3118, moxifloxacin, sitafloxacin, and gatifloxacin, have been reported to have amplified *in vitro* activity against this pathogen (Yew *et al*., 1994; Sirgel.*et al* 1999; Saito *et al.,* 1994; Gillespie and Billington, 1999; Gillespie,1999).

**2.5Activity of quinolones against mycobacterium tuberculosis**

Quinolones act by blocking bacterial DNA synthesis through inhibition of bacterial topoisomerase-II (DNA gyrase) and topoisomerase IV. Inhibition of DNA gyrase prevents the relaxation of positively super coiled DNA that is required for normal transcription and replication. Inhibition of topoisomerase-IV interferes with separation of replicated chromosomal DNA into the respective daughter cells during cell division (Alangaden and Lerner, 1997; Flamm *et al.*, 1995)

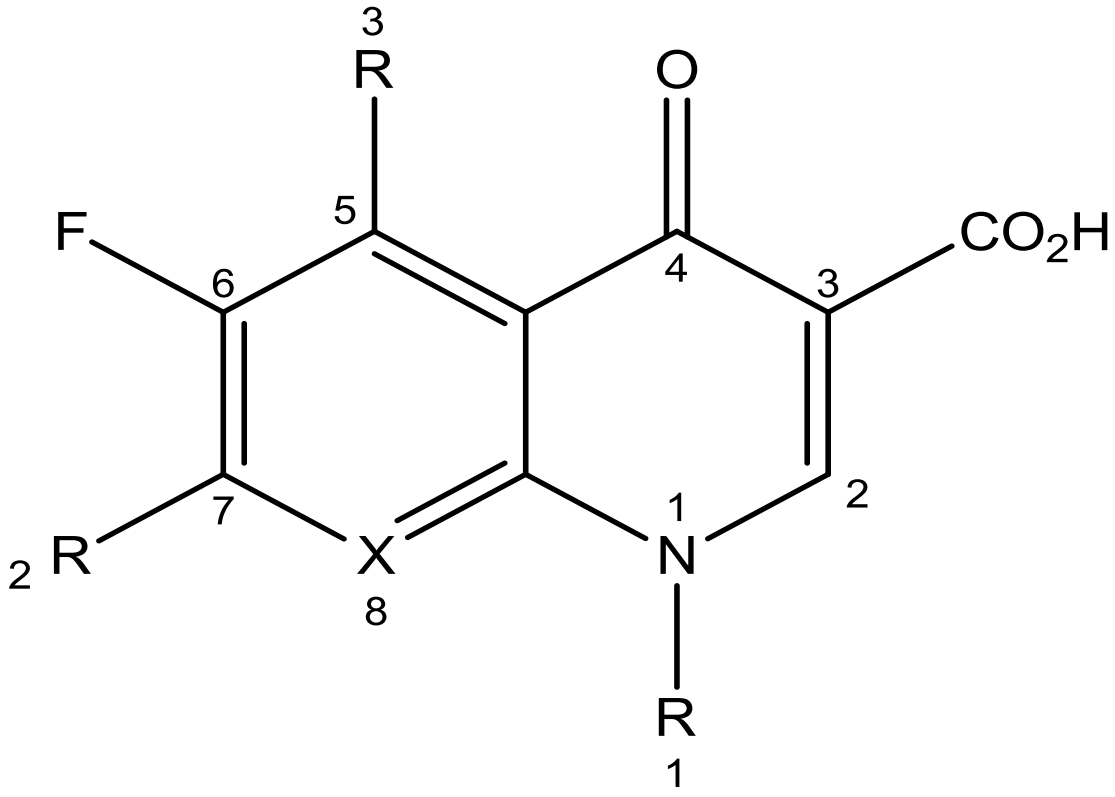



**Fig2.2**: **A Quinolone Molecule.**

## 2.6Use of Whole Infection Animal Model Testing in Drug Discovery

The use of specific chemicals to treat specific diseases and disorders dates to 1910 when PaulEhrlich and Sahachiro Hata discovered that salvarsan, also known as arsphenamine andcompound 606, killed the microorganism that caused syphilis. Drug development involves three stages Stage I: Hit and lead compound development phase(Identification of lead compound amongst the million compounds). Stage II: Preclinical studies (*in vitro* and *in vivo* experiments). Stage III: Clinical studies (experiments in humans)2. The preclinical stages studies for drug screening involves the use of animals, however it is often very time consuming and expensive and most times leads to the suffering of those organisms. Animal right activists across the world have intensified their campaign against the use of animals for this purpose. This has compelled researchers not only to find ways for reducing the time involved in drug screening procedure but also decrease the number of animals used.

## 2.7QSARs in TB Drug Design

Different studies using *in silico* method to help in designing new active anti-tubercular have been carried out extensively, (see Table 2.0).However in most cases these studies do not give much attentionto external validation and where it is done some researchers do not observe the most recently proposed stricter statistical criteria recommended for to assess the true prediction ability of the established models.A study was carried by Ponnurengam *et al*., (2007) on chalcones, chalcone-like compounds, flavones and flavanones to understand the relationship between anti-tuberculosis activity and structural features. The study employed the Genetic function approximation (GFA) method to identify the descriptors that would lead to good regression equation. They identified the best molecular descriptors to be Jurs descriptors (Jurs charged partial surface area), hydrogen bond donor, principal moment of inertia, molecular energy, dipole

magnetic, molecular area, absorption, distribution, metabolism andexcretion (ADME) properties and Chi indices (Kier and Hall chi connectivity indices). Theyconcluded that the compounds reported were more lipophilic in nature and hence, as expected exhibit good activity since *M. tuberculosis* has a high concentration of lipid layer. This is however in contrast to the finding of Filomina *et al*., (2010) in their study titled " QSAR based design of New Anti-tubercular compounds: improved Isonaizid against Multidrug-resistant TB" where in they reported that lipophilicity as measured by CLogP of target Isoniazid derivative against MTB is not relevant in explaining the anti-tubercular activity of these compounds. On the other hand, steric factors and geometrical features such as some inter-atomic distances were found to be very significant in modeling of the anti-tubercular behavior of this group of compounds. In a similar study Foroumadi*et al*., (2007) carried out a QSAR study of some novel 2-(nitroaryl)-5-(nitrobenzylsulfinyl and sulfonyl)-1,3,4-thiadiazole derivatives and the optimum 3-D geometry of the molecules were obtained using AM1 semi-empirical method.The final QSAR model showed that the anti-tubercular activity for the compounds used in the study are affected by:

(i) Mean atomic polarizability (Mp) which is a constitutional descriptor (ii) Maximum Positive Charge. They concluded that electronic distribution is among the most important factors in determining the activity of this group of compounds.Another study by Madhusudana *et al.,* (2014) led to the conclusion that activity of 1,3,5-triazine-bis-azomethines as anti-tubercular agents can be successfully modeled with molecular descriptors (molecular weight, wiener index, Andrews affinity and polar surface area).

**Table 2.0Some QSAR Studies Applied to TB (2008-2013).**

| Compounds | Target | Dependant variable (Y) | Descriptors Pool | Original Matrix Data | Training set Compounds | Descriptors Used | Computational method | Reported Statistics | | | ref |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Internal Validation | Test Set Compounds | External Validation | |
| Nitrofuranylamide | mtb | pMIC | Gobal Molecular and 3D Physicochemical descriptors | 110 | 95 | - | CoMFA/PLS and CoMSIA/PLS | $Q^2$ =0.585 SE = 0.625 | 15 | $R^2$= 0.769 SE = 0.524 | Henver *et al.*, 2008 |
| Cumarin-4-acetic acid benzylidene hydrazide | Mtb H37Rv strain | 'Activity' | - | 25 | 19 | Steric and electrostatic field | CoMFA/PLS | $r^2$ = 0.983, $Q^2$ = 0.636 | 6 | $R^2$=0.388, SE =0.052 | Manvar *et al*.,2008 |
| N-Penyl-6-methyl-2-oxo-4-phenyl-1,2,3,4-tetrahydrohydropyramidine-5-carboxomides | Mtb H37Rv strain | "Log activity" | 3D-cubic Lattice. grid spacing, probe atom, energy cutoff | 30 | 23 | 6 | CoMFA/PLS Regression | $r^2$ = 0.98 $q^2$ = 0.636 SE=0.08 | 7 | $R^2$=0.41 | Virsodia *et al*.,2008 |
| | | | 5 Physicochemical properties(steric, electrostatic,hydrophobic and hydrogen bond donor and acceptor | | | | CoMSIA/PLS Regression | $r^2$ = 0.95 $q^2$ = 0.58 SE=0.15 | 7 | $R^2$=0.32 | |
| Hydrazides | mtb | Log(1/MIC) | Physicochemical, steric, geometrical and electronic descriptors | 173 | 30 | 2 | MLR | $R^2$ = 0.758,$Q^2$=0.759 SE=0.320 | 29 | $R^2$=0.774, $Q^2$=0.746 SD=0.350 | Ventura *et al*.,2008 |
| Isoniazid derivatives | | | | | 22 | 3 | | $R^2$ = 0.830,$Q^2$=0.900SE =0.335 | 11 | $R^2$=0.896 $Q^2$=0.942 SD=0.263 | |
| Quinoxaline-2-carboxylate-1,4 –di N-oxide derivatives | Mtb H37Rv strain | Log(MIC) | Theoretical Structural descriptors | 43 | 38 | 5 | Multivariable Linear regression | $R^2$ = 0.946,$Q^2$=0.928s= 0.229 | 5 | - | Vicente *et al*.,2009 |
| 5-aryl-2-thio-1,3,4-oxadiazoles | Mtb H37Rv strain | % inhibition | Steric and electrostatic grid energy | 41 | 33 | 4 | KNN molecular field analysis | kNN =5 Df=26 $q^2$= 0.502 | 8 | $R^2$=0.2898 | Jhakk*et al*.,2009 |

**Table 2.0 Continued**

| compounds | Target | Dependant variable (Y) | Descriptors Pool | Original Matrix Data | Training set Compounds | Descriptors Used | Computational method | Reported Statistics | | | ref |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Internal Validation | Test Set Compounds | External Validation | |
| c-(3-aminomethyl-cyclohexyl)-methylamine derivatives | mtb H37Rv Strain | pMIC | - | 44 | 38 | - | PLS | $r^2$=0.9225; $Q^2$=0.9508; SE=0.1645 | 6 | $R^2$=0.9585; $Q^2$=0.7495; RMSE=0.28 | Kumar *et al.*,2013 |
| 3-phenyl-2H-1,3-benzoxazine-2,4-(3H)-diones | mtb H37Rv Strain | pMIC | 55 theoretical molecular descriptors and simulated NMR chemical shift | 55 | 29 | 55 | ANN/bank propagation | Without parameters | 13 | R=0.992; Ccc=0.986 | Nemecek *et al.*,2013 |
| N-phenyl-1,3-(4-fluorophenyl)-4-substituted pyrazoles | mtb H37Rv Strain | $pIC_{50}$ | 3D-steric,electrostatic,hydrogen bonding hydrophobic fields | 51 | 33 | 2 | CoMFA | $q^2$=0.64; SE=0.1 | 18 | $r^2$=0.6 | Khunt *et al.*,2012 |
| | | | | | 33 | 5 | CoMSIA | $q^2$=0.63; SE=0.15 | 18 | $r^2$=0.6 | |
| 3-ethyoxy-4-hydraoxybenzylidine/4nitrobezylidine hydrazides | mtb | pMIC | TSAR descriptors | 20 | 20 | 1 | MLR | $r^2$=0.812;$Q^2$=0.571; SE=0.11 | - | - | Kumar *et al.*,2011 |
| | | | | | 20 | 2 | | $r^2$=0.889 $Q^2$=0.611; SE=0.089 | - | - | |
| Oxozaline benzyl esters derivatyives | mtb | pGAST | Several 2D-Descriptors | 33 | 23 | 4 | Multiple | $r^2$= 0.8845; $Q^2$=0.8371; | 10 | $R^2$=0.759 | Khade *et al.*,2011 |

Molecular weight is an important parameter that signifies the size of the molecule. Wienerindex is atopological index of a molecule, defined as the sum of the numbers of edges in the shortest paths in a chemical graph between all pairs of non-hydrogen atoms in a moleculerelated to molecular branching. Andrews's affinity defines the functional group contributions to drug-receptor interactions.

The polar surface area (PSA) is defined as the surface sum over all polar atoms, (usually oxygen and nitrogen), including also attached hydrogens. However optimization of the compounds by only molecular mechanics (MM2) was not enough to give an accurate result.In separate 2D QSAR studies of some novel quinazolinone derivatives asAnti-tubercular agents, Rajasekaran*et al.* (2011) developed 34 computational models base on the activities of 19 compounds. The energy minimization of the molecules was done using MM2 force field followed by semi empirical AMI,9 of the models were found to be to be extremely significant. The model with descriptors involving diameter, ovality, partition coefficient and radius was found to be extremely significant. Also the model with ovality, partition coefficient and radius was also found to be extremely significant implying the importance of these parameters for the design of new pharmacophores containing quinazolinone moiety for anti-tubercular activity.

Minovski*et al.*(2011) investigated the activity 6-fluoroquinolones where they employed the MLR method d for the construction of a robust model for prediction of the inhibitory activity (*p*MIC) against *M. tuberculosis*. After optimizing the geometry of the compounds using the *ab initio*method, the best descriptors which were used for the model development were selected by Heuristic algorithm, which is a suitable option for selection of most important molecular descriptors. The selected most frequent theoretical molecular descriptors in that study described some of the crucial inter-molecular interactions between the 6-fluoroquinolones and the GyrA subunit and DNA: π-π

stacking interaction of the main 6-fluoroquinolone scaffold and the planar aromatic systems of the bacterial DNA, as well as possible electrostatic interactions with the amino acid residues of the GyrA subunit active site.

In the Multiple Linear Regression and Principal Component Analysis-Based study of the Anti-Tuberculosis Activity of some 2-aryl-1,3,4-Thiadiazole Derivatives, Hemmateenejad *et al.*,(2006), did a QSAR on the anti-tuberculosis activity of some newly synthesized 2-aryl-1,3,4-Thiadiazole derivatives. Different sets of molecular descriptors were calculated to predict the anti-tuberculosis activity of a set of 27 thiadiazole derivatives using the multiple linear regression (MLR), Free-Wilson analysis (FWA) and principal component analysis (PCA) methods. Fusing the pool of all calculated descriptors afour-parametric equation including, Ms as constitutional descriptor, sum of surface area of polar atoms (PSA) as electronic parameter, number of sulfurs (nRSR), a functional group parameter, and a valence connectivity topological index (X2Av), was obtained. Further attempts were made to classify the molecules to the active and inactive compounds by application of the principal component analysis (PCA) on the descriptors data matrix. In this study, the PCA was run two times. First, all of the calculated descriptors were subjected to the PCA and then PCA was run on the descriptorsused in the MLR equation.For the entire set of descriptors, the first three PCs explain 78.3% of variances in the descriptors data matrices, and for the subset of 4 descriptors selected by MLR, the 97% of variances were explained by the first three PCs. The resulting equations describe the significant roles of the hydrophobic and electronic interaction on the anti-tubercular activity. Also, they revealed the effect of the molecular topology and transient interactions on the anti-tuberculosis activity. In addition, it was reported that to obtain compounds with higher activity, molecules with lower number of nitrogen atoms and sulfuxide group should be synthesized. In the PCA, different patterns of distributions were obtained by plotting the first three

scores of the descriptors data matrices against each other. Better classification patterns were obtained when PCA was performed on the subset of descriptors relative to all of the calculated descriptors.

Ravichandran *et al.* (2014)undertook a QSAR study, eight 7-chloroquinolines was used to develop the model using step-wise multiple linear regression technique. The developed QSAR model was validated andevaluated for statistical significance. Physiochemical descriptors serve as the first line guidance for the design of potent and novel anti-tubercular agents. The key conclusion of this study is that the number of hydrogen bond acceptors significantly influences the anti-tubercular activity of novel 7- chloroquinoline derivatives.

# CHAPTER THREE

## 3.0 Materials and Methods

### 3.1 Materials

All calculations in this study were performed using an Intel Celeron core Hp Laptop Computer with 4 GB of memory and 350 GB of scratch disk space**.** The following softwares were used;

(i) SPARTAN "14 v 1.1.0 software

(ii) PADEL software package

(iii)BuildQSAR software

(iv)Material studio software

(v) NanoBRIDGES softwares (vWSP program andthe NormalizeData1.0).

### 3.2 Method

QSAR modeling can be characterized by a collection of well-defined protocols and procedures that enable the expert application of the method (Alexander, 2000). The summary of these protocols is given in the work flow in Fig 3.1

#### 3.2.1 Data set

The biological assay data used in this study were obtained from classical *in vitro* tests for inhibitory activity reported as MIC against *M. tuberculosis* and *M.smgmatis* as well as cytotoxicity activity in a mammalian vero cell reported as $IC_{50}$ .The dataset was taken from the

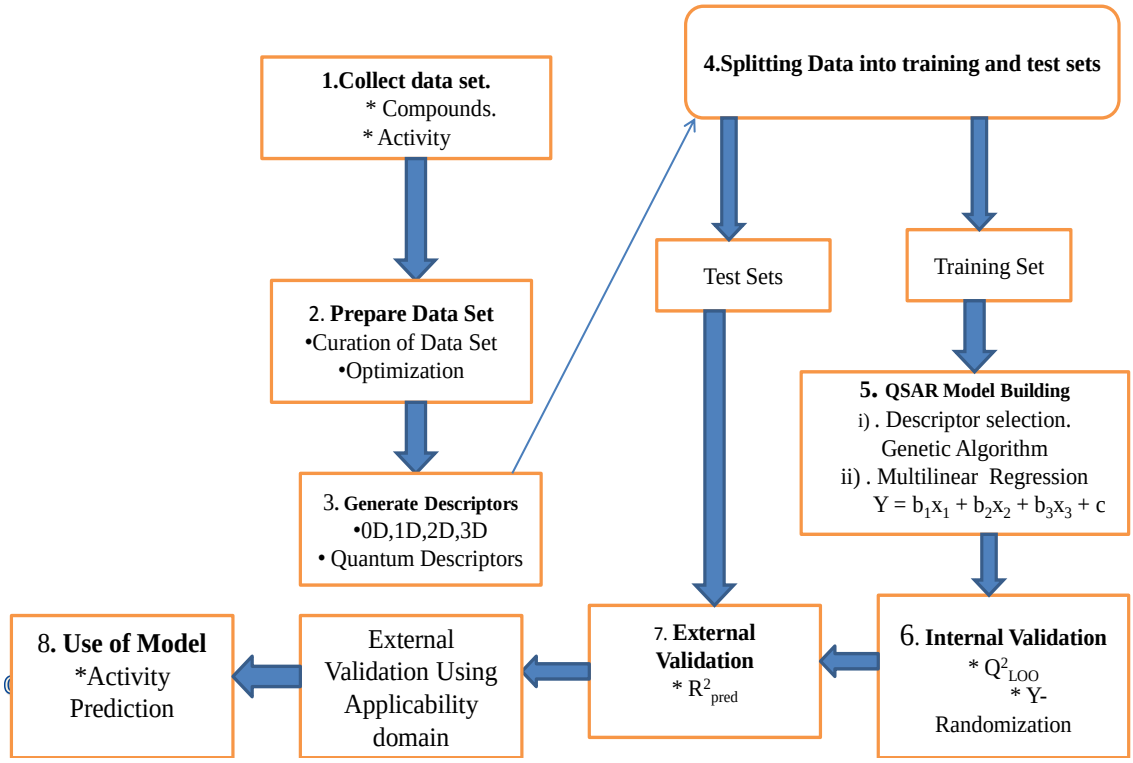
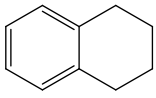



**Fig3.1:QSAR Workflow**

works of Senthilkumar *et al*. (2009) and Sriram *et al*. (2007). It consists of fifty ninequinolones(see Fig. 3.2 for the moiety of the compounds and Table 3.1 for detailed structures of the compounds respectively).The inhibitory data (MIC in µM) and cytotoxicity activity ($IC_{50}$ in µM) were converted to negative logarithmic dose in mole pMIC and $pIC_{50}$ respectively because a QSAR is a linear free energy relationship, and from the Van't Hoff isotherm, free energy change during a process is proportional to the logarithm of the rate or equilibrium constant of the process ($\Delta G$ = 2.303 RTlog K) (Gupta *et al.*, 2007).

### 3.2.2 Molecular structure generation

The first step in performing any QSAR studies is to specify the structure of the molecules**.** The correct 3D geometry of the compounds is very important to find a possible relationship between molecular descriptors of the compounds and their biological activity (Minovski*et al*., 2011) as well as for subsequent calculation of 3D molecular descriptors. SPARTAN "14 v 1.1.0 software packages (Warren and Sean, 2013), contain tools to sketch organic molecules in 2D and then automatically convert them into realistic 3D structures The procedure for building compound 1 in Fig. 3.3 is as follows: The **SPARTAN14** software package was launched by *clicking* the **Spartan 14** icon on the desktop of the Laptop Computer. From the **Menu bar**(Fig 3.4) the **File** menu and**New Sketch**  were selected respectively to bring up the **sketch pad**(Fig.3.5). From the sketch pad, a benzene ring was selected by *clicking* on ( ), the screen was then *double-clicked* to place the benzene ring. Next, cyclohexane (  ) was selected from the sketch pad and the 5-6 bond (see numbering in Fig 3.3) was *double clicked*. These operations resulted in the structure .

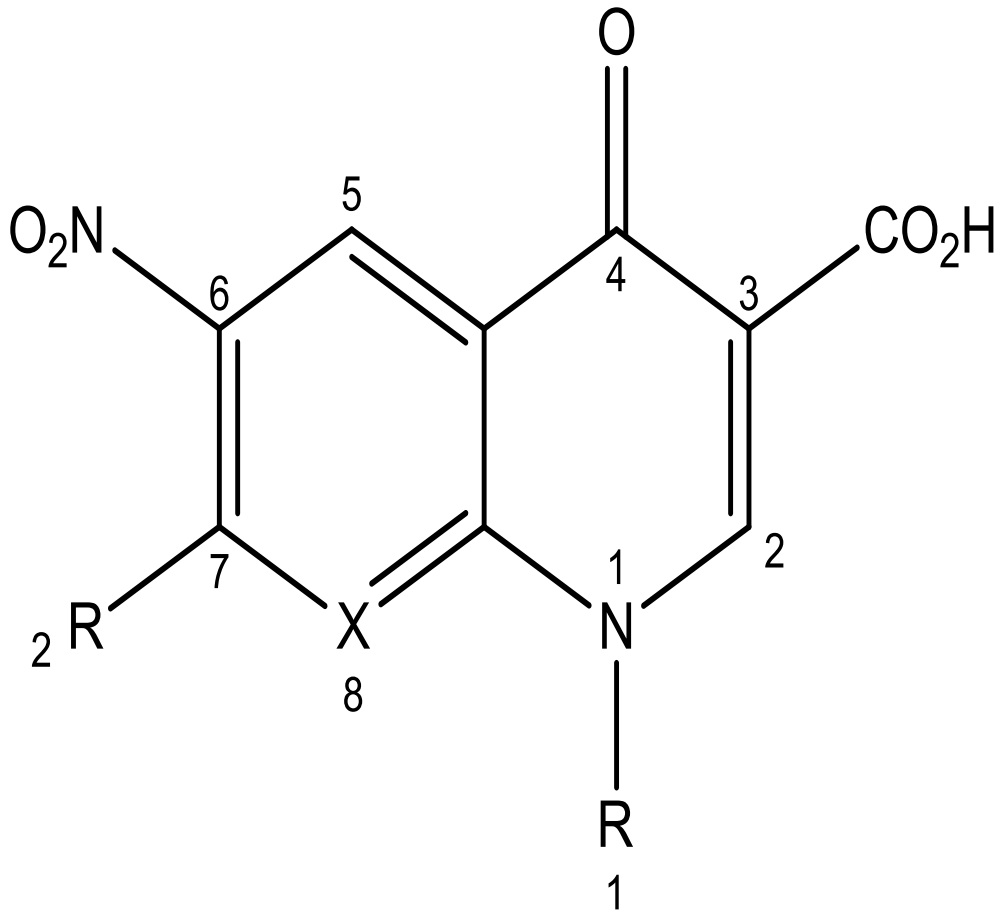
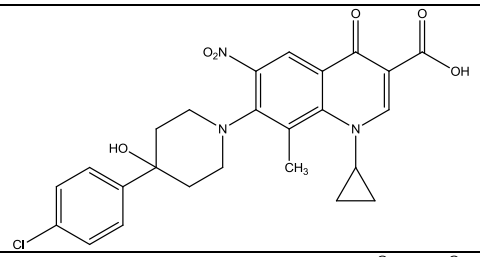
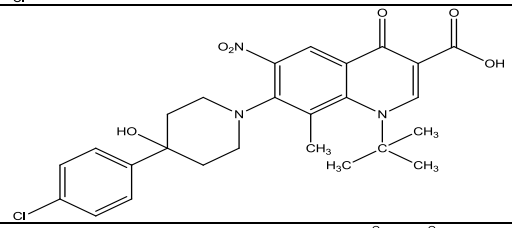
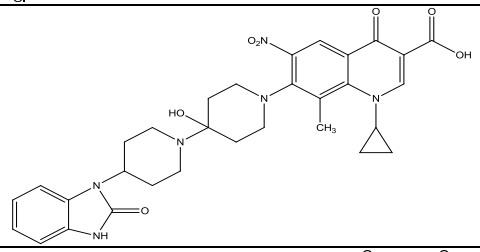
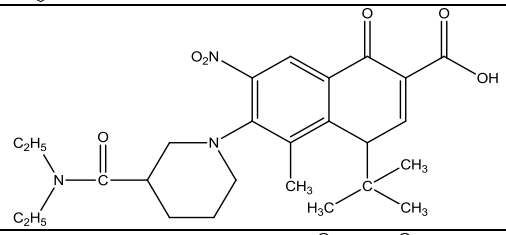
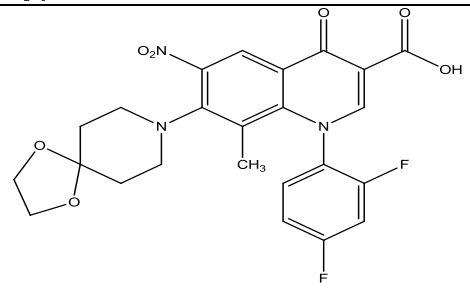
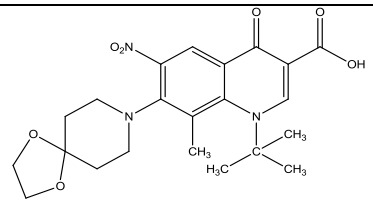
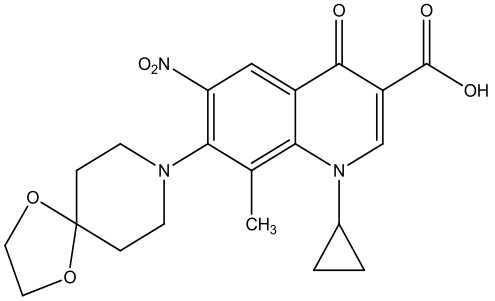
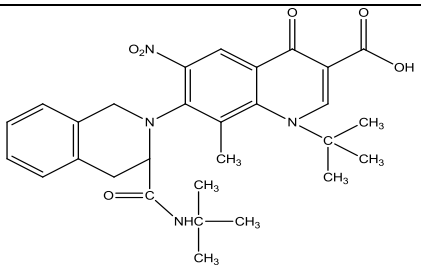
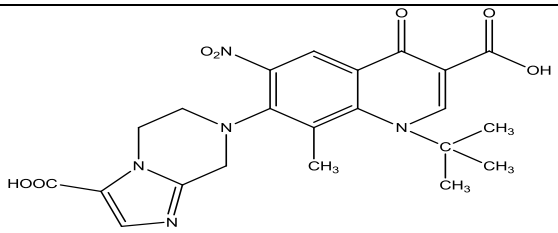
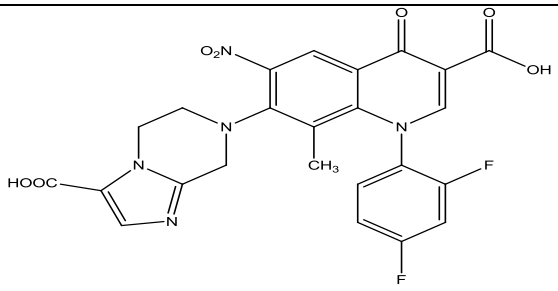
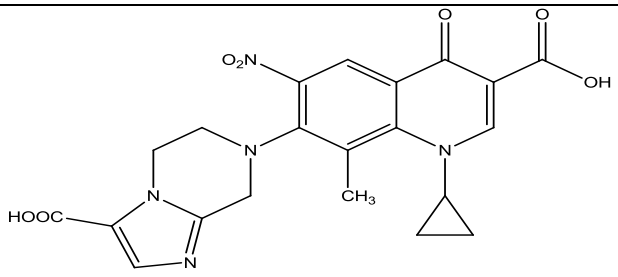
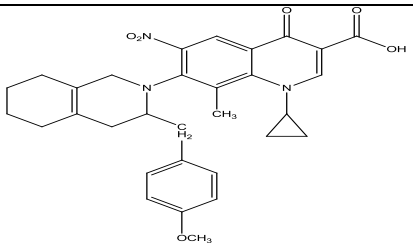
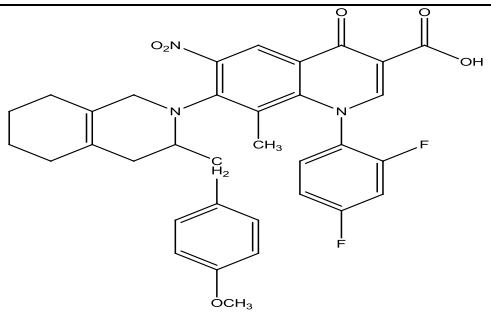
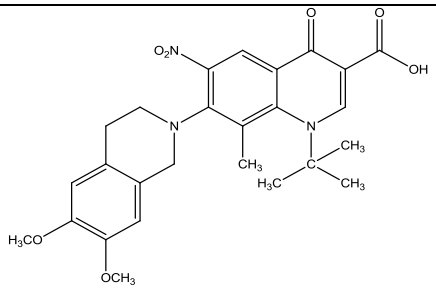
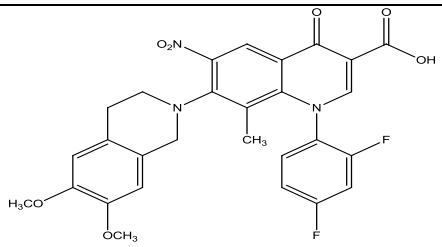
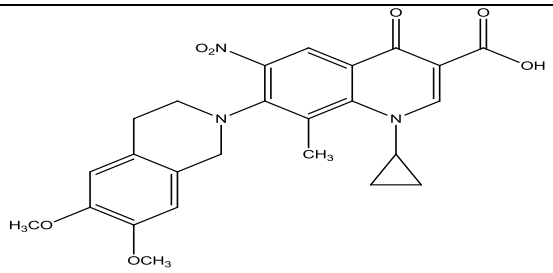
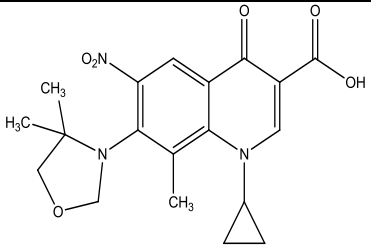
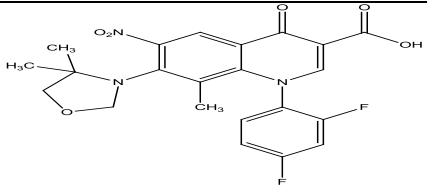


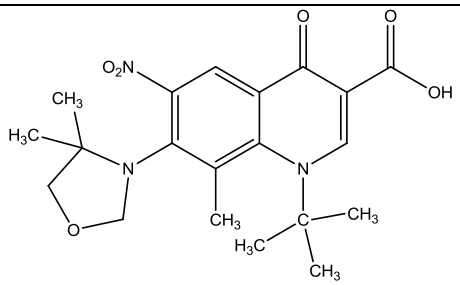
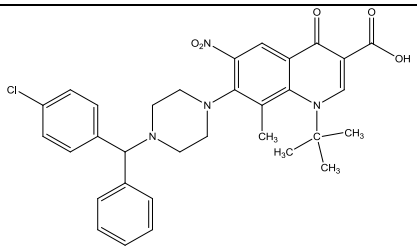
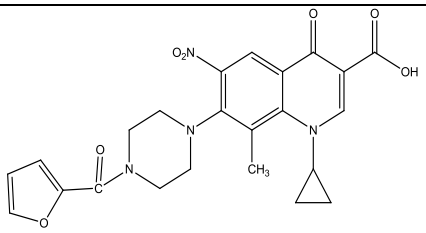
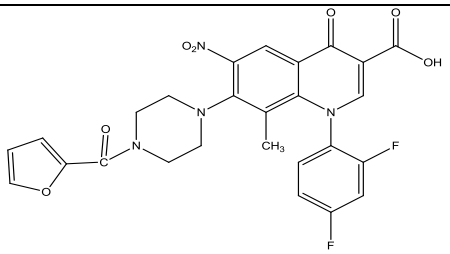
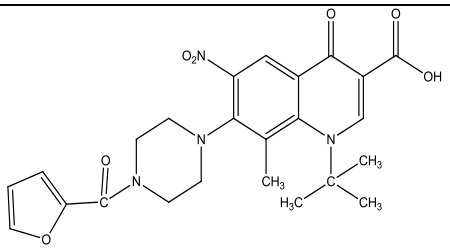
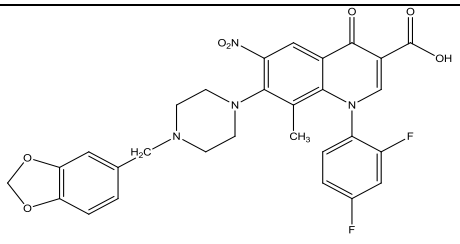

**Fig 3.2:Moiety of the compounds**

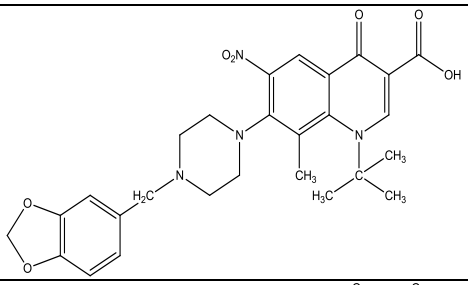
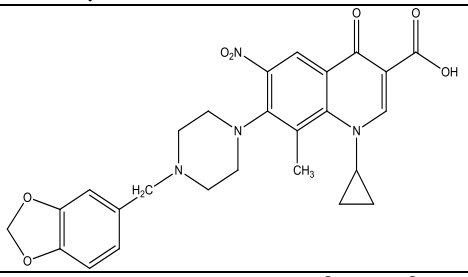
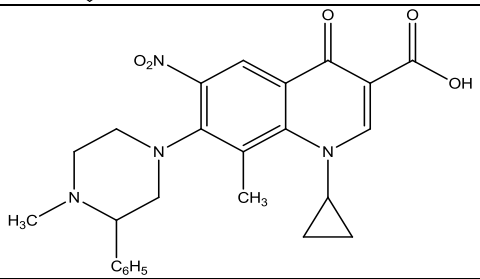
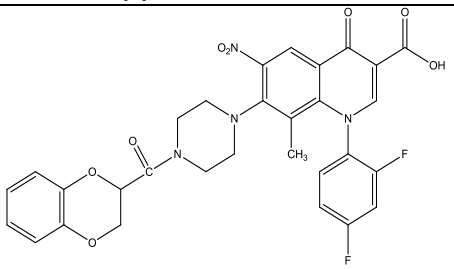
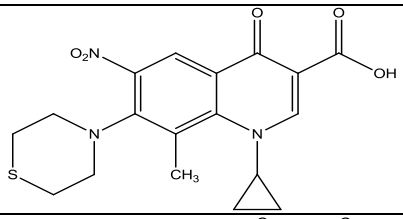
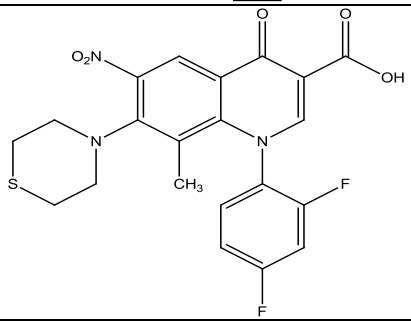
**Table 3.1 The Quinolone Compounds and their activities**

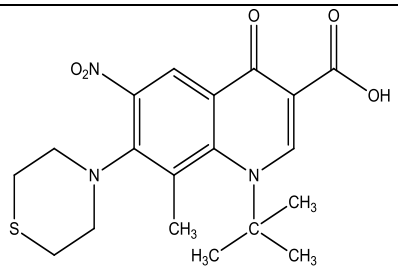
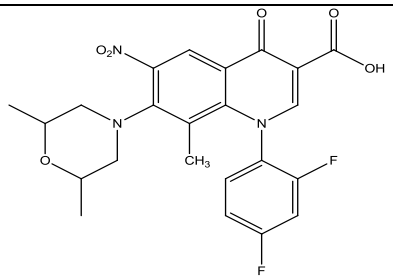
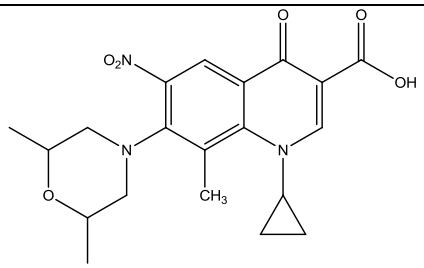
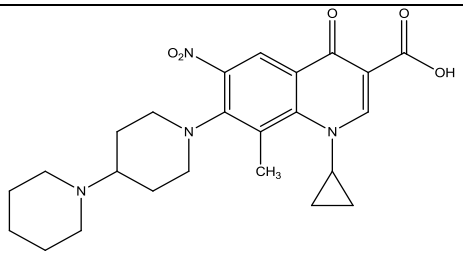
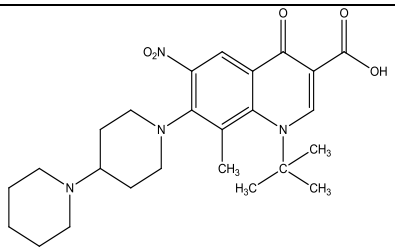
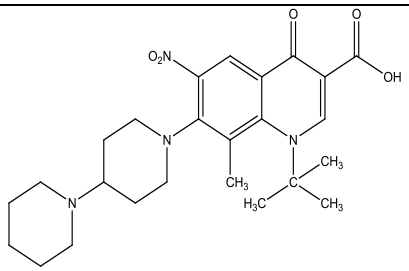
| ID | MOLECULE | pMIC[a] | pMIC[b] | $pIC_{50}$ |
|---|---|---|---|---|
| 1 |  | 4.90 | 5.20 | |
| 2 |  | 4.36 | 4.61 | |
| 3 |  | 4.64 | 5.54 | 3.94 |
| 4 |  | 4.29 | 4.59 | |
| 5 |  | 4.60 | | 3.90 |
| 6 |  | 3.95 | 4.85 | |

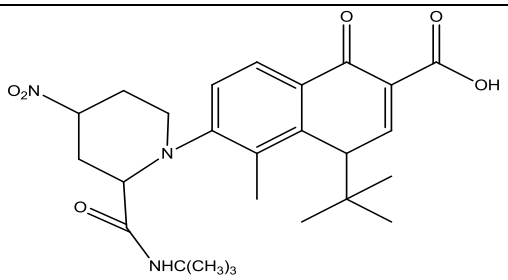
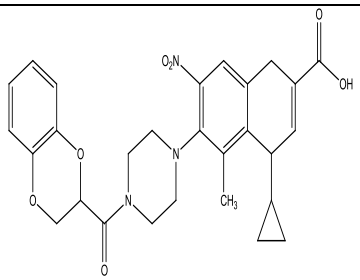
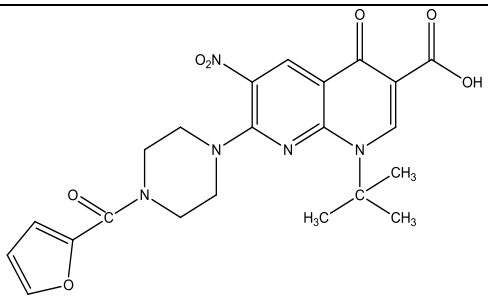
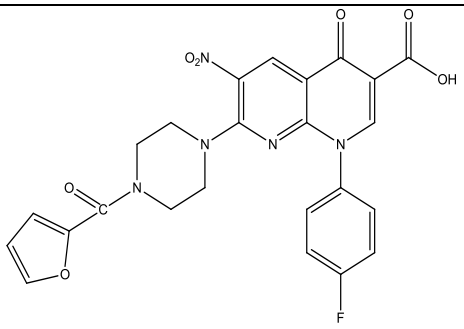
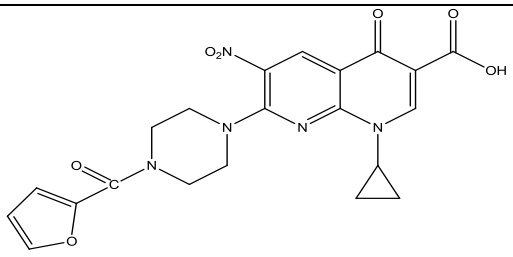
| ID | MOLECULE | pMIC[a] | pMIC[b] | $pIC_{50}$ |
|---|---|---|---|---|
| 7 |  | 5.44 | 5.74 | 3.84 |
| 8 |  | 5.82 | 5.82 | |
| 9 |  | 5.22 | 4.57 | 3.92 |
| 10 |  | 3.97 | 5.22 | |
| 11 |  | 4.32 | 6.38 | |
| 12 |  | 5.76 | 5.51 | 3.86 |

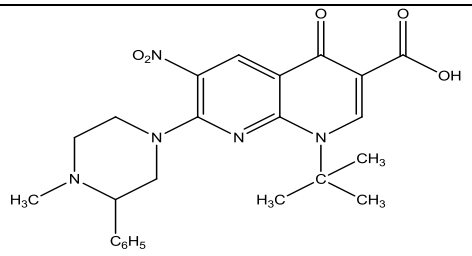
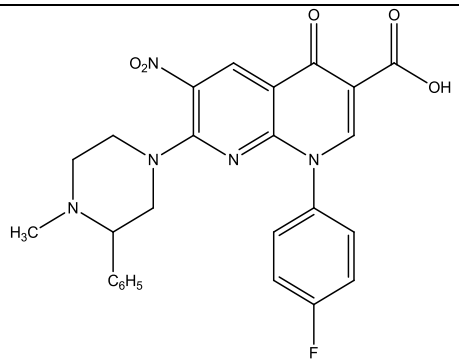
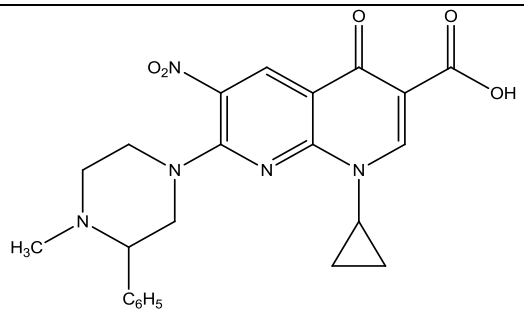
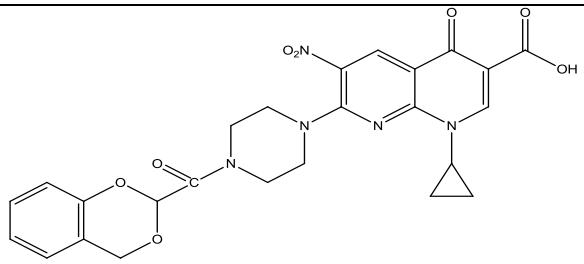
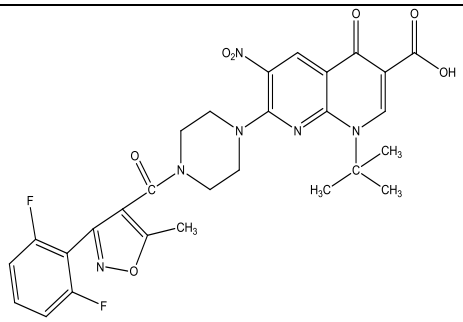
| ID | MOLECULE | pMIC[a] | pMIC[b] | pIC50 |
|---|---|---|---|---|
| 13 |  | 5.23 | 4.70 | 3.91 |
| 14 |  | 4.39 | 4.60 | |
| 15 |  | 4.30 | 4.95 | |
| 16 |  | 4.34 | 6.41 | |
| 17 |  | 6.09 | 5.70 | 3.89 |
| 18 |  | 5.40 | 5.17 | 3.79 |

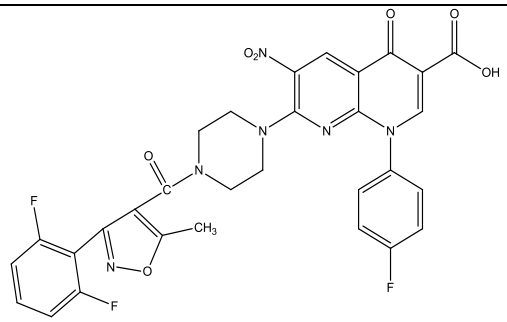
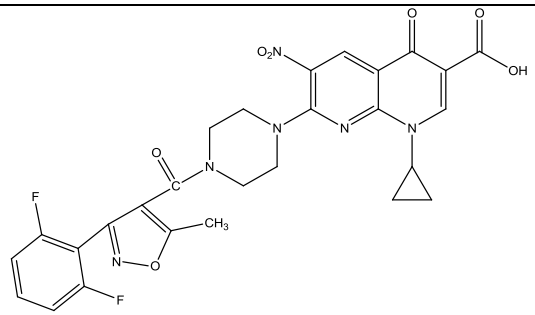
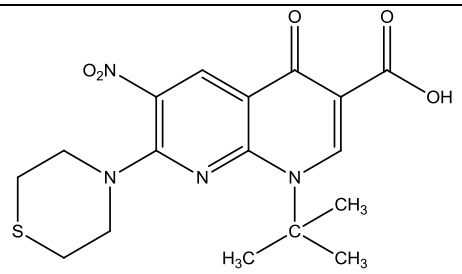
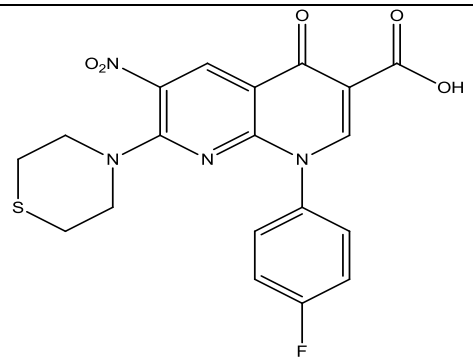
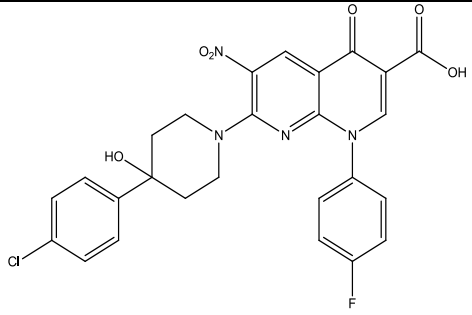
| ID | molecule | pMIC[a] | pMIC[b] | $pIC_{50}$ |
|---|---|---|---|---|
| 19 |  | 4.57 | 4.81 | |
| 20 |  | 4.21 | 6.17 | |
| 21 |  | 6.08 | 6.07 | 3.96 |
| 22 |  | 5.78 | 4.39 | 3.87 |
| 23 |  | 4.63 | 4.59 | |
| 24 |  | 4.66 | 4.67 | |

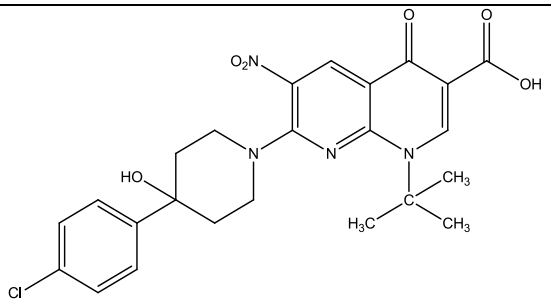
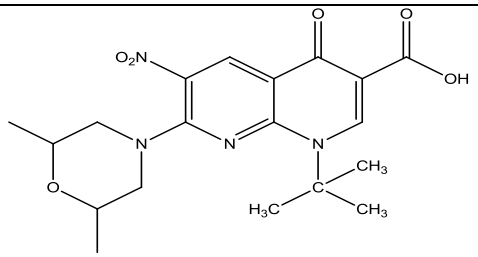
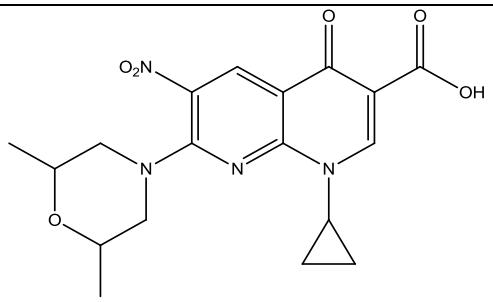
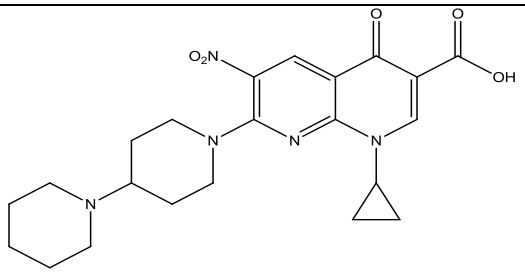
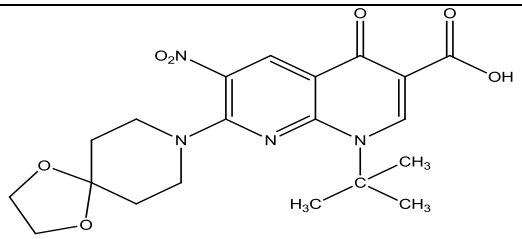
| ID | Molecule | pMIC[a] | pMIC[b] | $pIC_{50}$ |
|---|---|---|---|---|
| 25 |  | 4.36 | 4.62 | |
| 26 |  | 4.32 | 6.11 | |
| 27 |  | 5.51 | 6.08 | 3.91 |
| 28 |  | 5.47 | 5.53 | 3.87 |
| 29 |  | | 5.70 | |
| 30 |  | 5.40 | 4.87 | 3.79 |

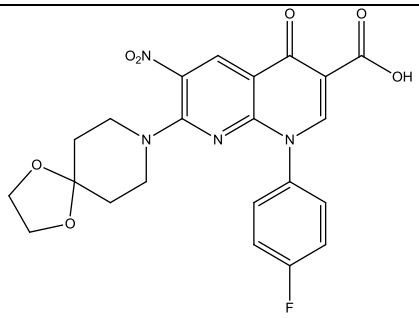
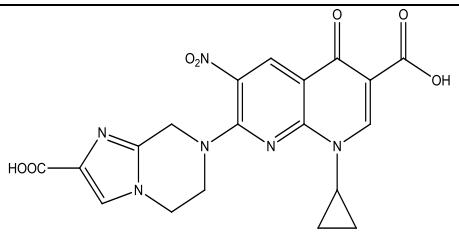
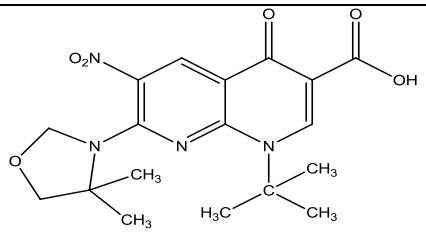
| ID | Molecules | pMIC[a] | pMIC[b] | $pIC_{50}$ |
|---|---|---|---|---|
| 31 |  | 4.27 | 4.51 | |
| 32 |  | 4.21 | 5.18 | |
| 33 |  | 4.58 | 6.01 | |
| 34 |  | 5.71 | 5.16 | 3.81 |
| 35 |  | 5.16 | 4.58 | |
| 36 |  | 3.97 | 5.53 | |

| ID | molecules | pMIC[a] | pMIC[b] | $pIC_{50}$ |
|---|---|---|---|---|
| 37 |  | 4.33 | 4.07 | |
| 38 |  | 4.93 | 5.53 | 3.93 |
| 39 |  | 5.48 | | |
| 40 |  | 5.51 | | |
| 42 |  | 5.16 | | |

| ID | Molecule | pMIC[a] | pMIC[b] | $pIC_{50}$ |
|---|---|---|---|---|
| 43 |  | 6.08 | | |
| 44 |  | 6.11 | | |
| 45 |  | 5.16 | | |
| 45 |  | 4.92 | | |
| 46 |  | 5.58 | | |

| ID | Molecule | pMIC[a] | pMIC[b] | $pIC_{50}$ |
|---|---|---|---|---|
| 47 |  | 5.01 | | |
| 48 |  | 4.67 | | |
| 49 |  | 6.00 | | |
| 50 |  | 6.36 | | |
| 51 |  | 5.54 | | |

| ID | Molecule | pMIC[a] | pMIC[b] | $pIC_{50}$ |
|---|---|---|---|---|
| 52 |  | 5.51 | | |
| 53 |  | 5.11 | | |
| 54 |  | 5.40 | | |
| 55 |  | 5.15 | | |
| 56 |  | 6.05 | | |

| ID | Molecule | pMIC[a] | pMIC[b] | $pIC_{50}$ |
|---|---|---|---|---|
| 57 |  | 5.18 | | |
| 58 |  | 5.45 | | |
| 59 |  | 7.00 | | |

* pMIC[a] is the negative logarithm to base ten of *in vitro* tests for inhibitory activity reported as MIC against *M. tuberculosis*

pMIC[b]is the negative logarithm to base ten of *in vitro* tests for inhibitory activity reported as MIC against*M.smegmatis*

$pIC_{50}$is the negative logarithm to base ten of cytotoxicity activity in a mammalian vero cell reported as $IC_{50}$

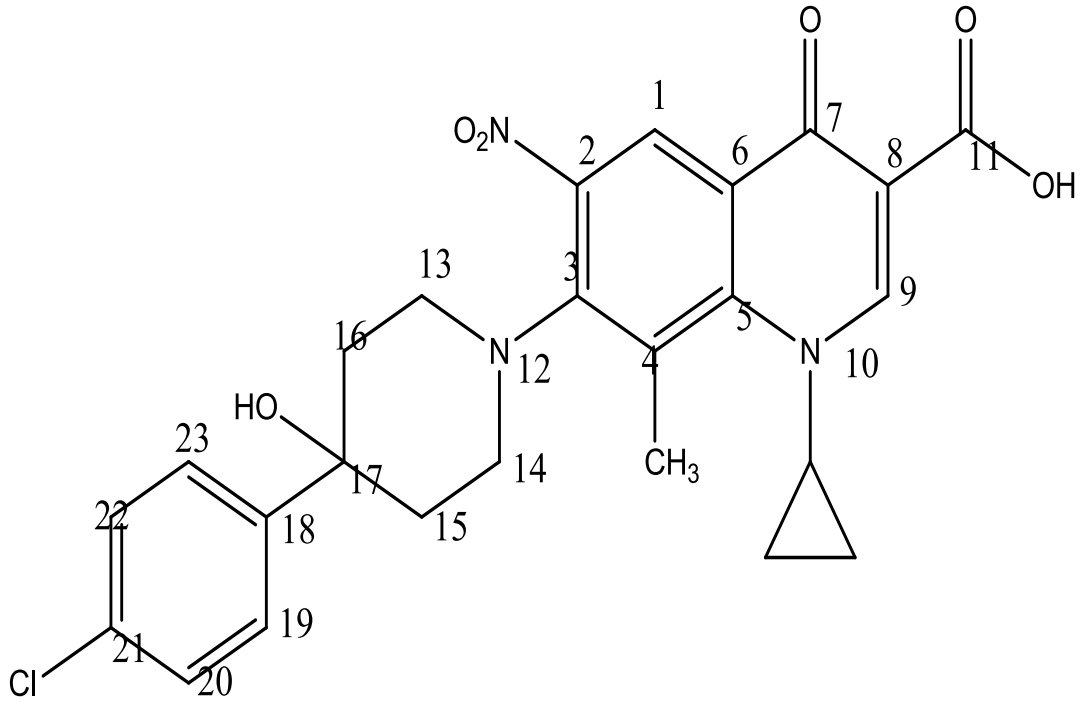



**Fig 3.3: Structure of Compound 1**

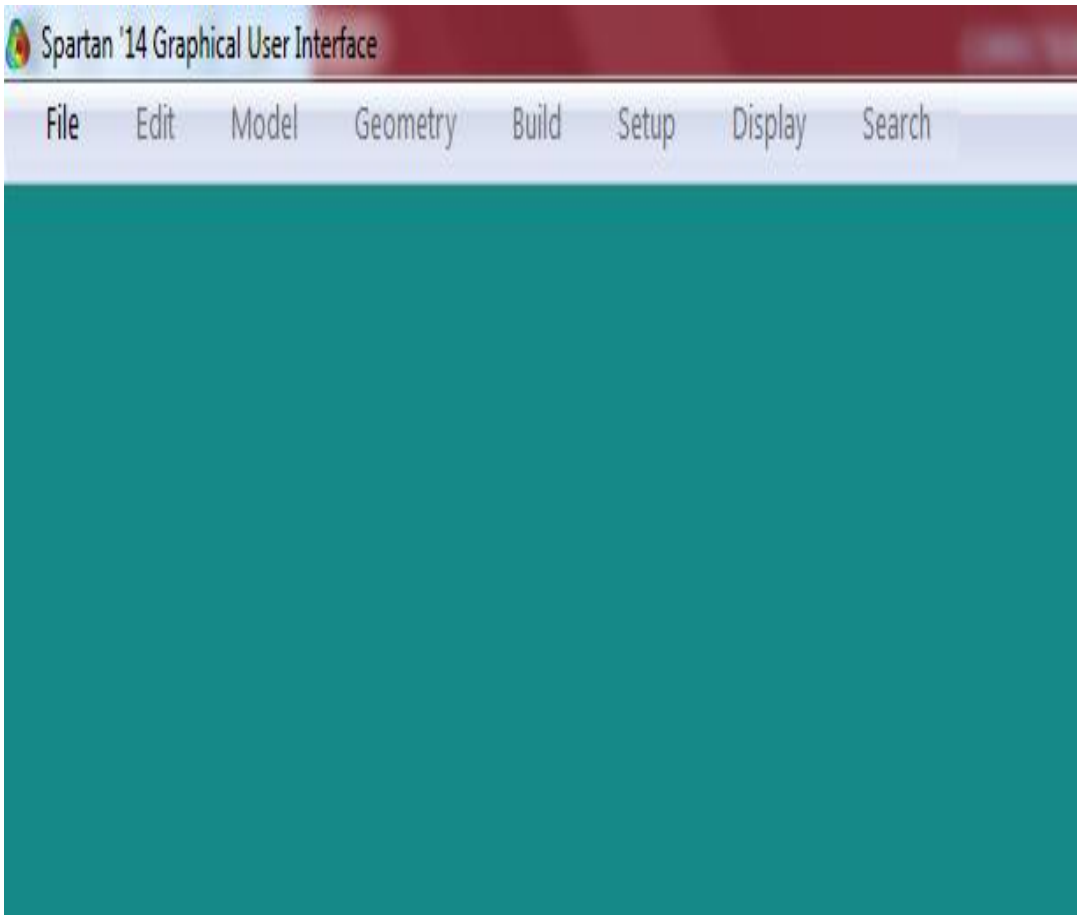



**Fig.3.4: Graphical User Interface showing the Menu bar**

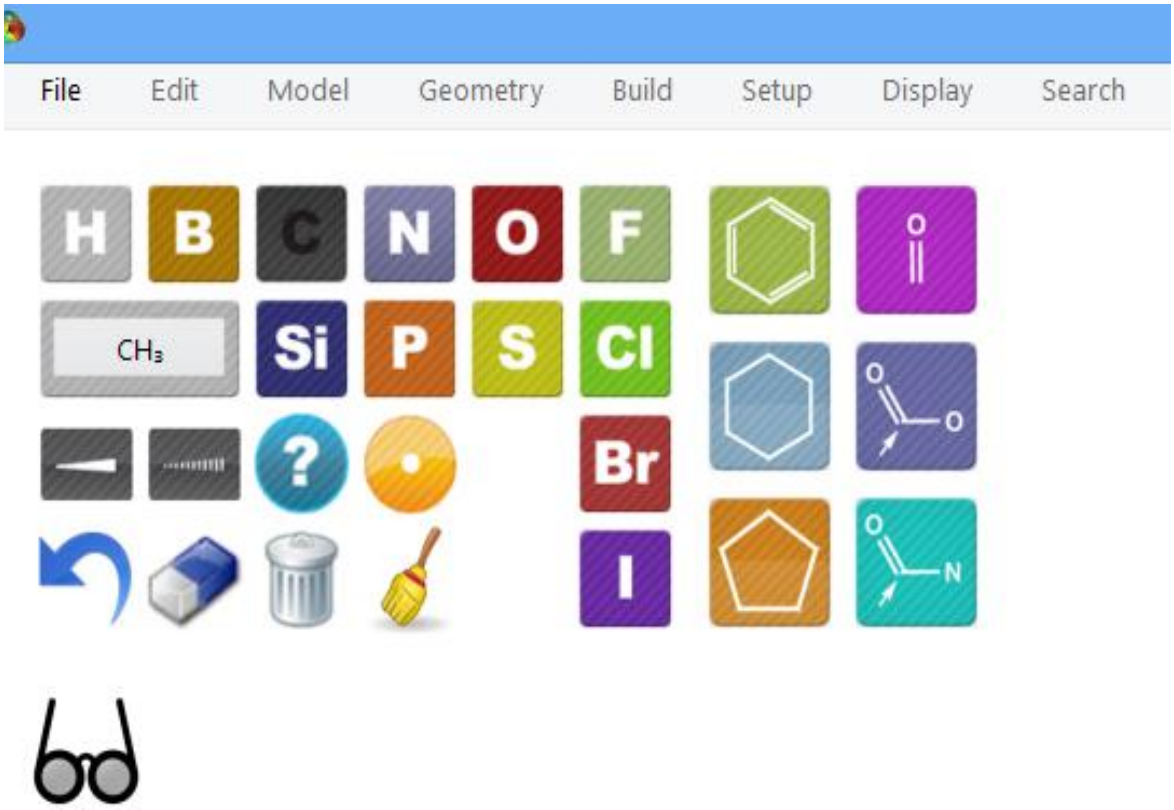
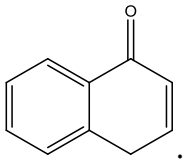
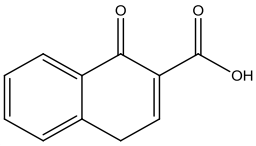



**Fig3.5: Sketch Pad**

With carbon ( C ) selected on the sketch pad, the cursor was positioned over $C_9$, *dragged* away from the ring and released to form a double bond between $C_9$ and $C_{10}$ .Oxygen atom was selectedby *clicking* O on the sketch pad, the cursor was placed above $C_7$, *dragged* away from the ring and released forming a CO single bond between $C_7$ and the oxygen atom, again the cursor was positioned over $C_7$, *dragged* along the CO bond to the oxygen atom and released, converting the CO single bond to a CO double bond to give the structure

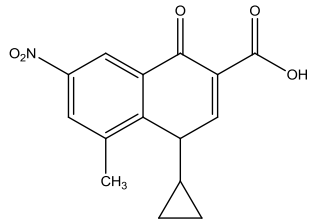


Carbon C was selected on the sketch pad and the cursor was positioned over $C_{10}$ at approximately $45^o$ to it, *dragged* towards the right and released, O was selected and the cursor positioned above $C_{11}$, *dragged*away from it in an upward direction and released forming a CO single bond between $C_{11}$ and the oxygen atom, the CO single bond was converted to a CO double bond as described earlier. With O still selected on the sketch pad, the cursor was positioned over $C_{11}$, dragged down to the right and released. The following structure was produced

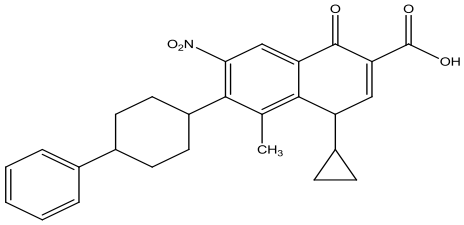



With C selected on the sketch pad, the cursor was positioned over $C_4$, *dragged* down and released to give a methyl substituent on $C_4$, still with C selected, the cursor was positioned over $C_8$*dragged* down, then dragged again to form a triangle which now yields a cyclopropyl substituent on $C_8$.The wildcard icon Ph located immediately below H and B in the palette was *Double-clicked,* $NO_2$ was typed in the box in the dialog that appeared, OK was *clicked*; $NO_2$appeared in

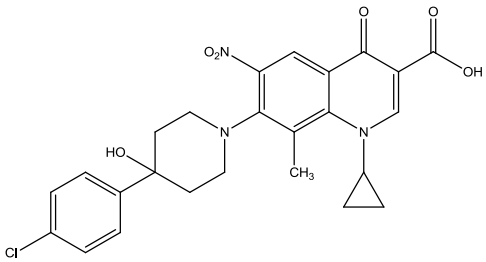
place of the wildcard icon. The cursor was positioned over $C_4$, *dragged* away from the ring and released to give a $NO_2$ substituent on $C_2$




Cyclohexane ring was selected on the sketch pad, the cursor was positioned over $C_3$ and *dragged* away from the ring and released, a benzene ring was also selected, and the cursor was positioned over $C_{17}$, *dragged* and released.



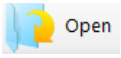

On the sketch pad, O was selected,the cursor was placed over $C_{17}$ and dragged away to the left, Cl was selected the cursor was placed over C21 dragged and released, finally (N) was selected the cursor was placed one after another C8 and C12 were double-clicked.




The brush icon on the sketch pad was clicked to clean up the drawing and also on the sketch pad was clicked to convert the structure to 3D which was checked by visual inspection in

order to ensure that the 3D geometry was correct, and saved for geometry optimization. This procedure was applied in building the 3D structures of all the compounds in this study, the structures were saved as Spartan 14 input file format (.spartan).

### 3.2.3 Geometry optimization and calculation of quantum-chemical descriptors

The molecular geometry of all the structures were optimizedusing the density functional theory (DFT) with Becke's three-parameter hybrid functional (Becke, 1988) using LYP correlation functional (Lee *et al*., 1988). The standard Pople's 6-311G* basis set was used. This was done by using **SPARTAN14** software package. The compound to be optimized was retrieved from the folder containing the 3D structures that were built earlier. This was done by clicking on 

on the **tool bar** (see Fig.3.6). Once the 3D structure of the compound was displayed on the screen the **Build** menu on the menu bar was selected (see Fig 3.6)then 

was *clicked* to minimize the energy of the molecule. Next, from the **Menu Bar** (see Fig 3.4),the **Setup** menu was *clicked*,followed by selecting**calculation** which resulted in the display ofthe **calculations Dialog**inFig.3.7;this contains a number of pull-down menus, buttons, check boxes and a space to enter additional information ("options").**Equilibrium Geometry** from the top menu to the right of **Calculate** was selected; this specifies optimization of equilibrium geometry. By default **Ground** appeared in the menu to the left of **State** this was left as it was. **Density Functional**, **BLY3, 6-311G***and **Vacuum** from the four bottom menus to the right of **Calculate** were selected.

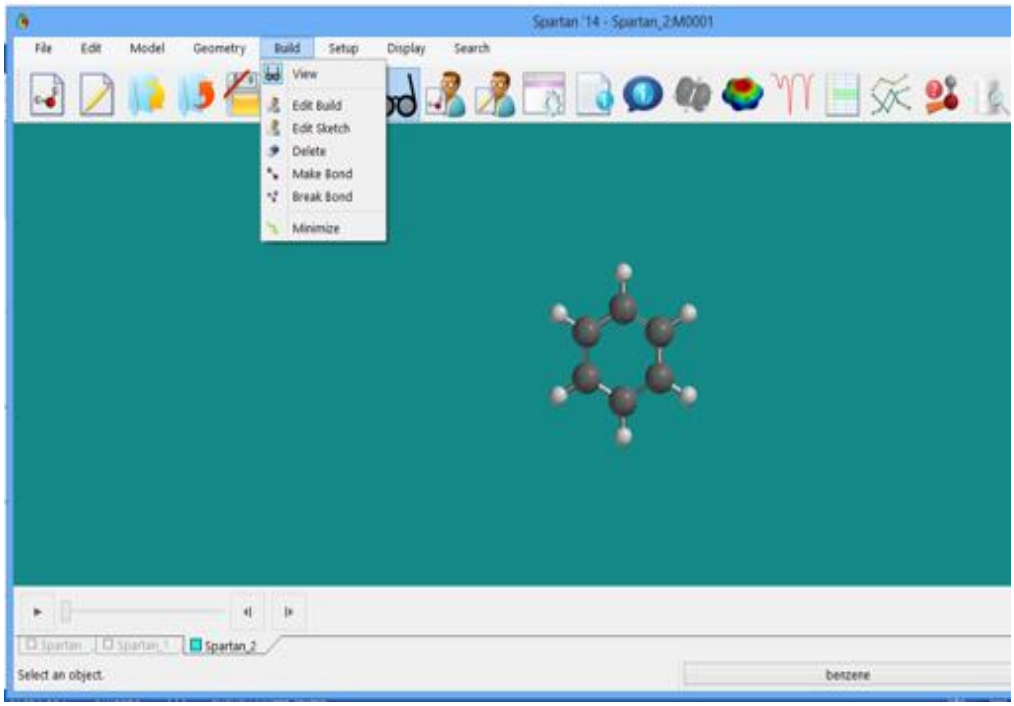



**Fig. 3.6 Tool Bar**

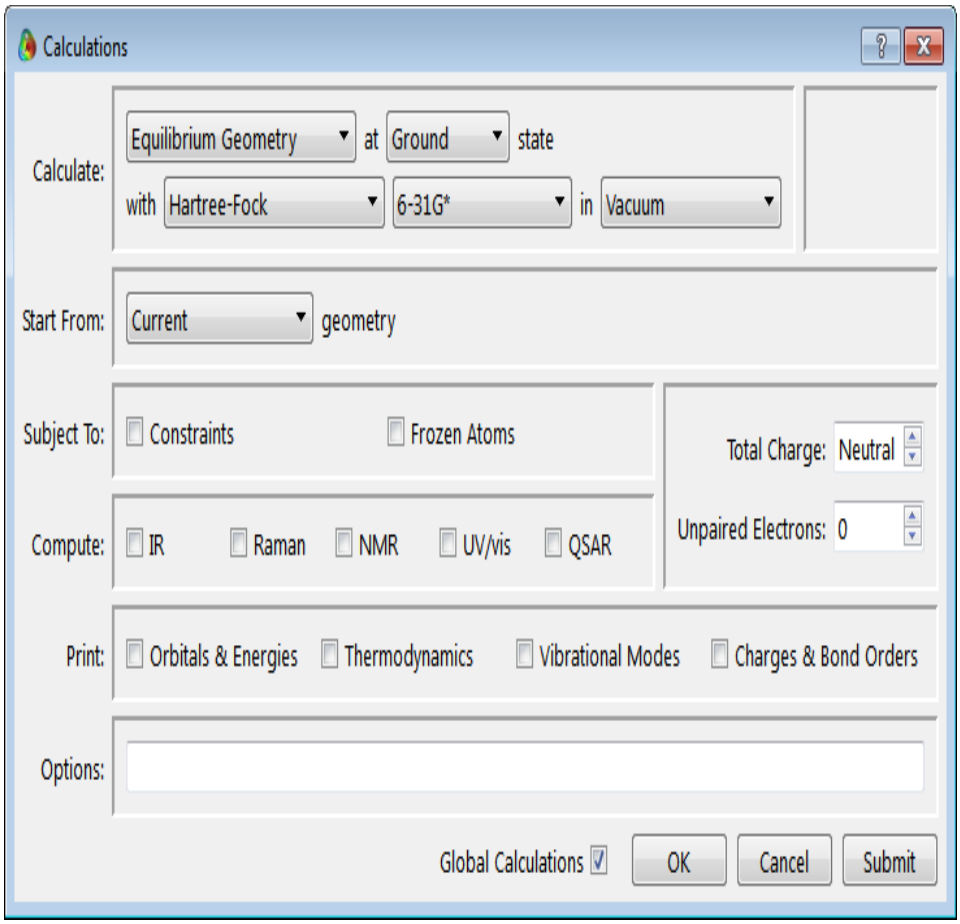



**Fig. 3.7**: **Calculations Dialog Box**

This specifies that the BLY3/6-311G* density functional model was to be used for the calculation. The **QSAR BOX** was*checked* to enable the calculation of a number of QSAR descriptors such as polarizability and logP etc. The **Thermodynamics box** was*checked* **(**thisrequires that vibrational frequencies **IR** under **Compute** have been checked)to calculate some thermodynamic properties such as enthalpy, entropy, heat capacity and Gibbs energy which also served as molecular descriptors in this study. The **Total Charge** is Neutral for all calculations in this study.**Submit** at the bottom of the dialog was *clicked*; a **file browser**Fig 3.8 appeared. In the section **File name** the name of the compound was typed and**Save** was *clicked*. A notification appeared that the calculation has been submitted. This procedure was applied to optimize all other compounds in this study.

### 3.2.4 Calculation of 0D, 1D,2D and 3D descriptors

For calculation of the 0D, 1D, 2D, 3D molecular descriptors for each compound of the training/validation sets, PADEL (Yap, 2011) software package was used.The jar file in the Padel folder was *double clicked* to launchthe software and access the interface Fig 3.9.The folder containing the earlier optimized molecules was selected by *clicking* the **molecules directory/file**(Fig 3.9). A folder to which the calculated descriptors would be saved was then selected by *clicking* on the **Descriptor output file**; this brings up a **file browser** Fig.3.10. In the **File name** section on the file browser the document was named accordingly. On the **Files of types** section,comma separated value (CSV) file format was selected to indicate that the calculated descriptors would be saved in this file format.The options "**1D &2D**" was*elected* to calculate 1D and 2D descriptors as shown in Fig.3.11. The **3D** option was also selected to calculate the 3D descriptors as shown in Fig 3.12. The options "**Remove salt**", **"Detect aromaticity**" and "**Standardize tautomers**", were all *selected* in other to remove salts like NaCl, remove existing aromaticity

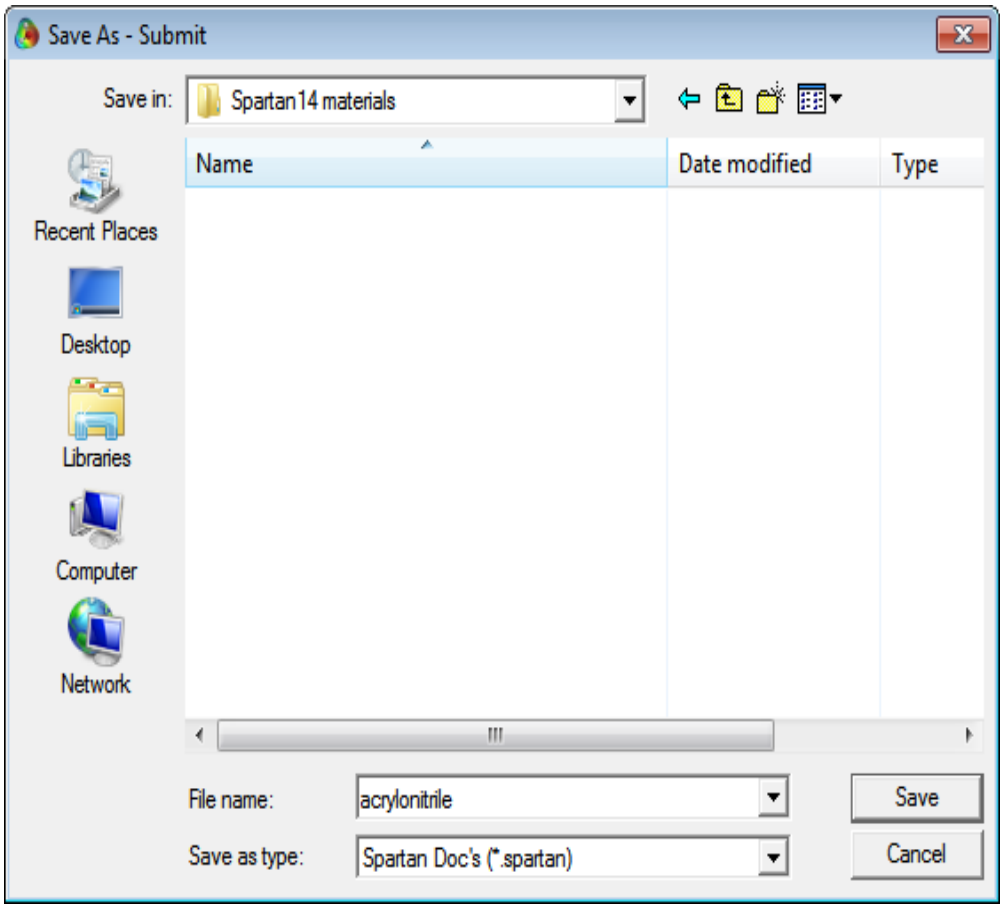



**Fig.3.8: File browser**

.

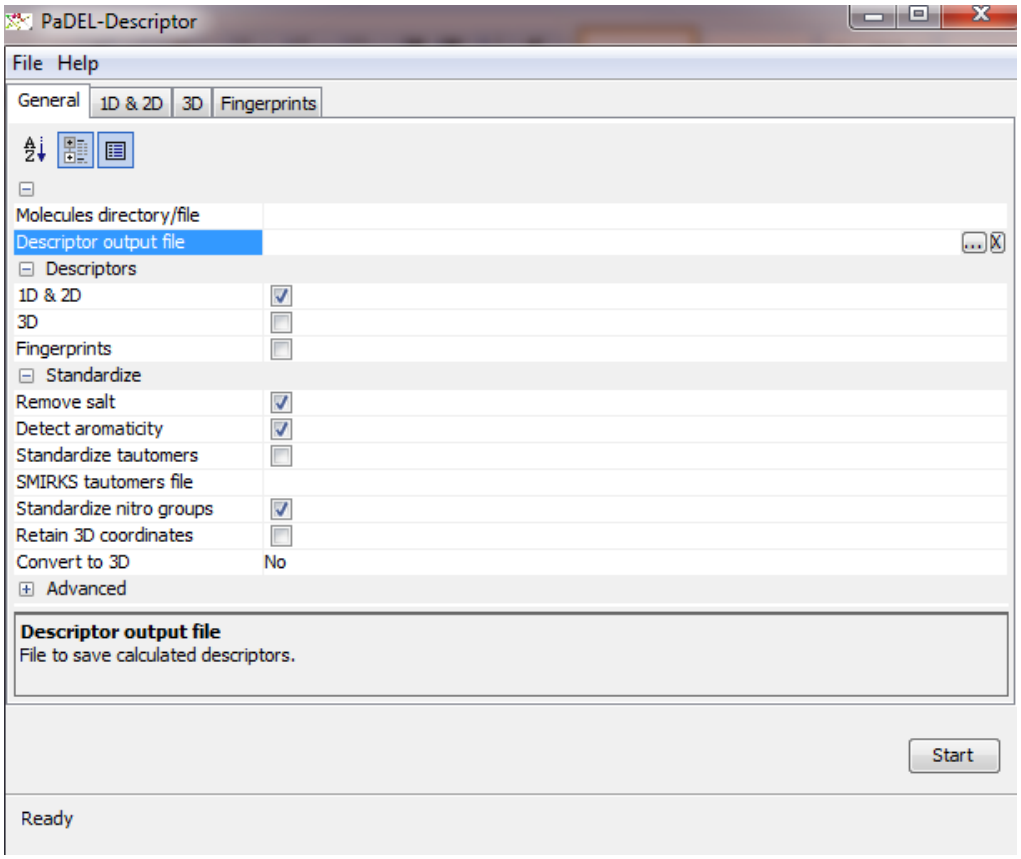



**Fig 3.9:Padel Descriptor Snapshot-A**

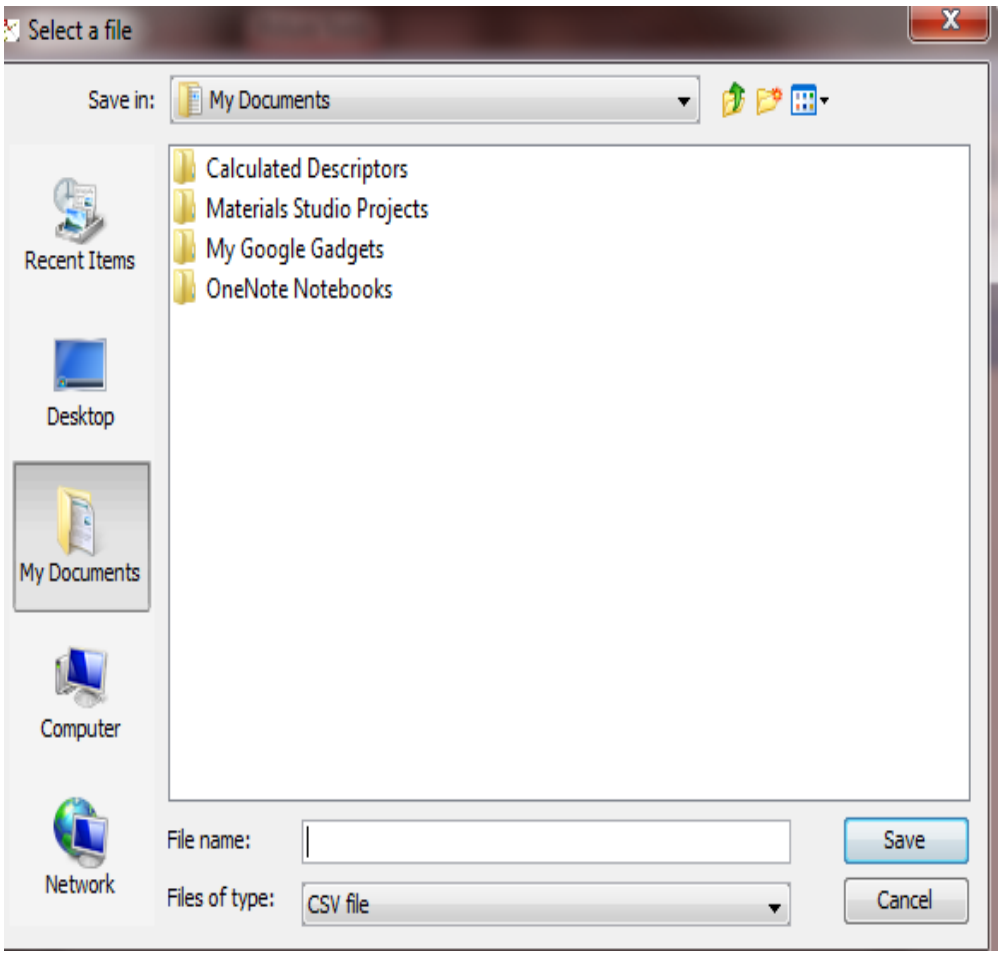



**Fig 3.10**: **File Browser-A**

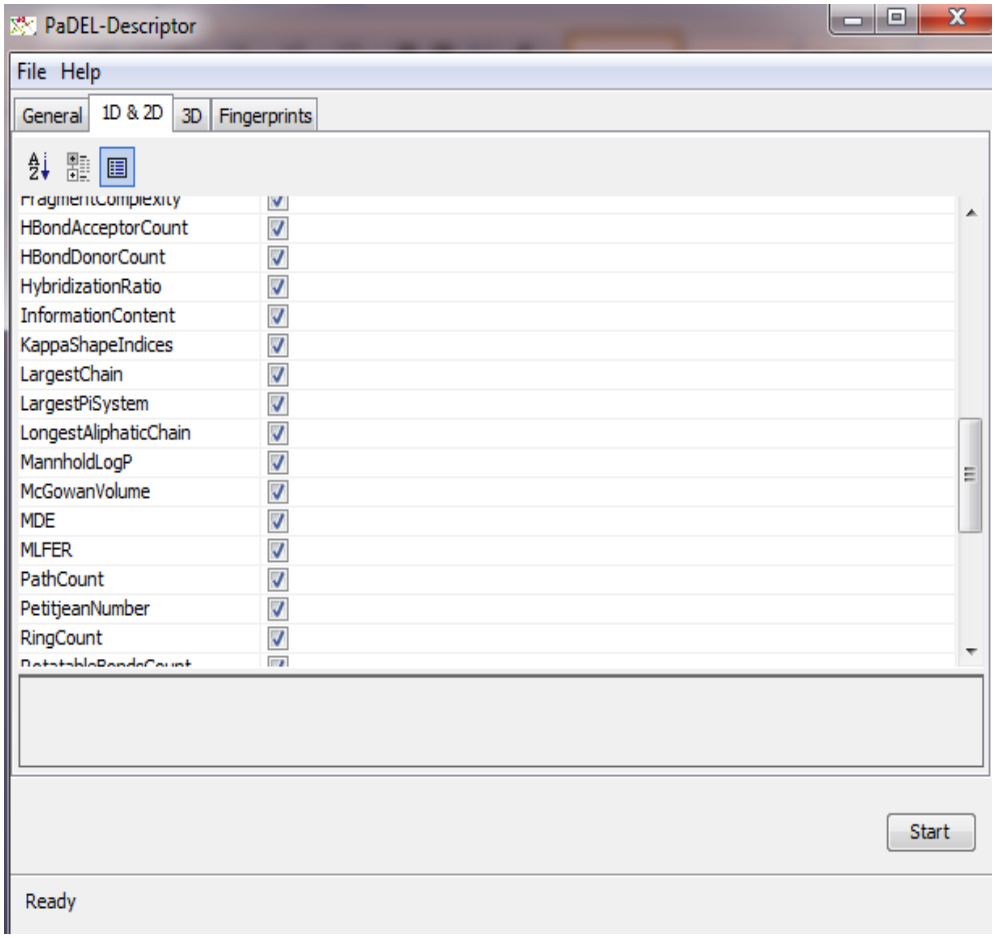



**Fig 3.11: Padel Descriptor snapshot-B**

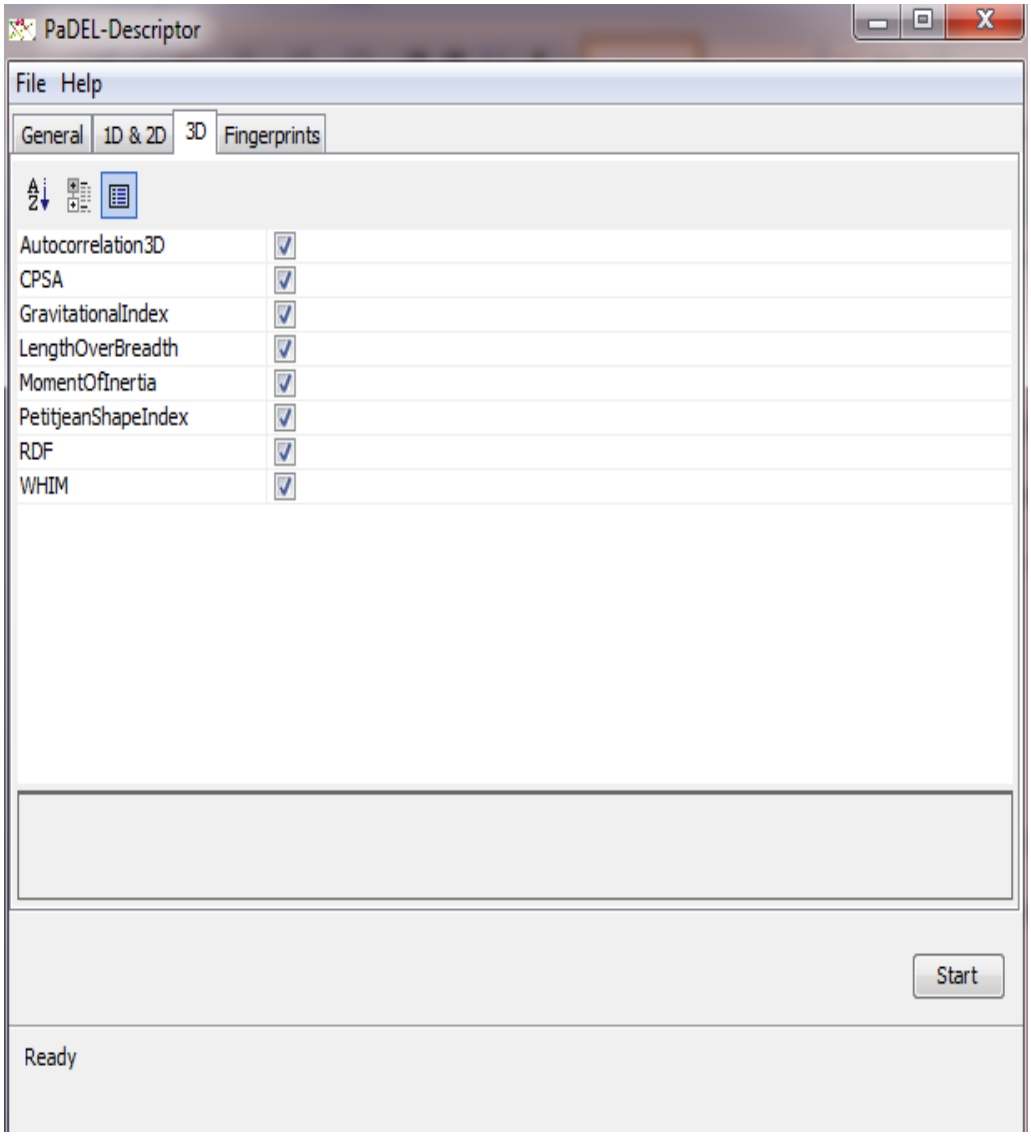



**Fig 3.12:Padel Descriptor Snapshot-C**

from the molecule and standardize tautomers respectively before calculation of descriptors, finally the job was then submitted for calculation by *clicking* on **start.** This was repeated for all the molecules.

### 3.2.5 Descriptor Normalization

The calculated descriptors were normalized using NormalizeData 1.0 software (Ambure*et al.*,2015). The excel file containing the calculated descriptors was exported to the folder containing the NormalizeData 1.0 software. The executable jar file contained in the folder was launched by *double clicking* on it to give the dialog box in Fig. 3.13. The name of the excel file containing the descriptors was *typed* and **ok** was *clicked* to give the scaled descriptors in a separate excel worksheet. This worksheet was named and saved accordingly.

### 3.2.6Correlation Matrix

The file containing the calculated descriptors of the molecules earlier saved in csv file format was opened in Excel worksheet. The corresponding biological activities (pMIC) of the molecules were pasted in the first column. From **the Menu Bar** Fig 3.14, **Data analysis**which is found at the extreme right of the **Menu Bar** was *selected*, this brought up **Data tools Dialogue box**(Fig.3.15) containing a number of **Analysis Tools**. From the **Analysis tools**, **Correlation** was *selected*. **Ok** was *clicked* which resulted in the **dialogue box** in Fig.3.16. On the **input section** the input range was specified and '**Labels in first row'** option *checked*. On the **Output section'New Worksheet Ply'** was selected. **Ok** was *clicked* to generate the correlation matrix

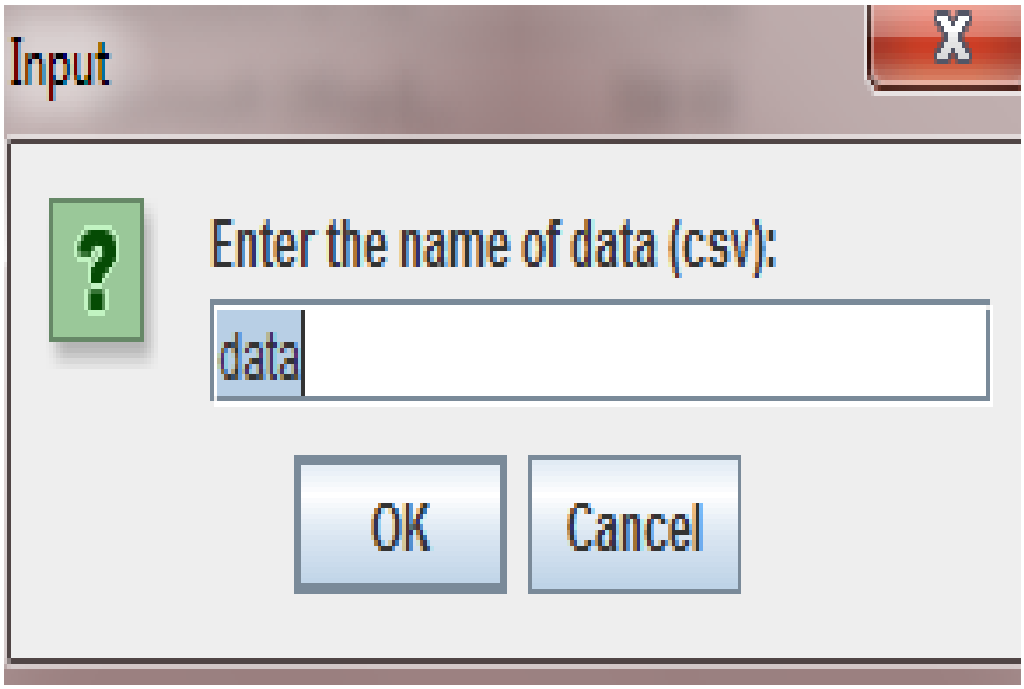
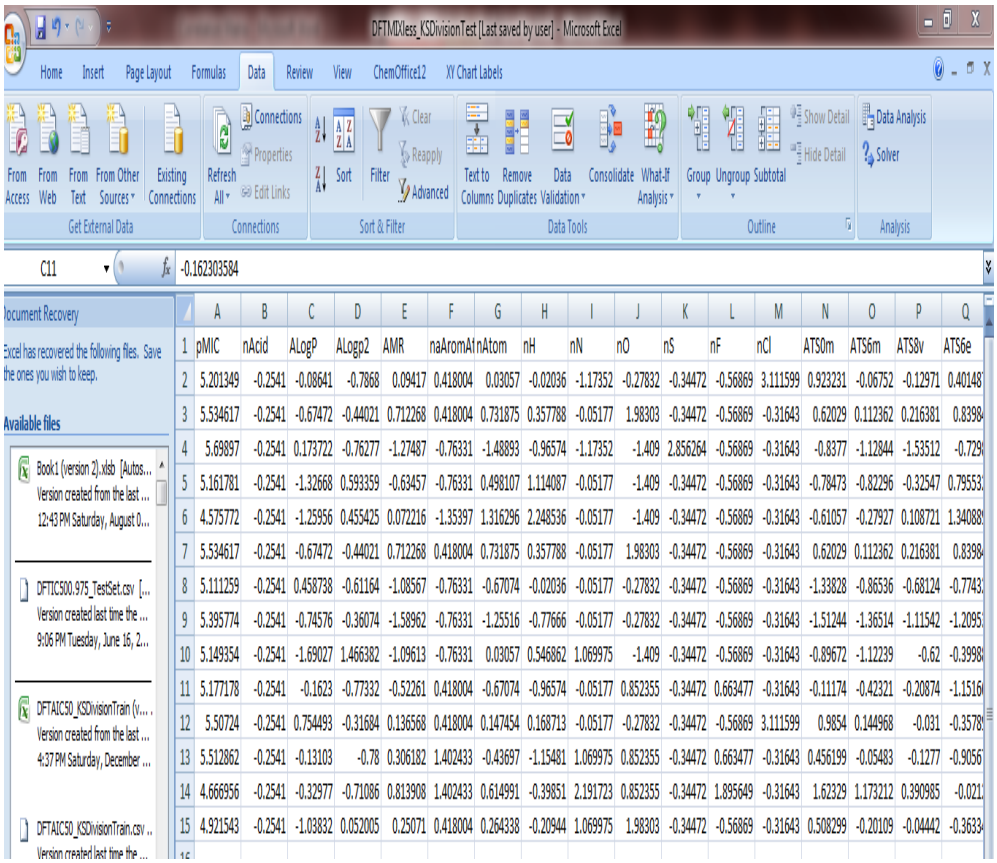



**Fig. 3.13**: **Normalization Dialogue box**




| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | pMIC | nAcid | ALogP | ALogp2 | AMR | naAromAt | nAtom | nH | nN | nO | nS | nF | nCl | ATS0m | ATS6m | ATS8v | ATS6e |
| 2 | 5.201349 | -0.2541 | -0.08641 | -0.7868 | 0.09417 | 0.418004 | 0.03057 | -0.02036 | -1.17352 | -0.27832 | -0.34472 | -0.56869 | 3.111599 | 0.923231 | -0.06752 | -0.12971 | 0.40148 |
| 3 | 5.534617 | -0.2541 | -0.67472 | -0.44021 | 0.712268 | 0.418004 | 0.731875 | 0.357788 | -0.05177 | 1.98303 | -0.34472 | -0.56869 | -0.31643 | 0.62029 | 0.112362 | 0.216381 | 0.8398 |
| 4 | 5.69897 | -0.2541 | 0.173722 | -0.76277 | -1.27487 | -0.76331 | -1.48893 | -0.96574 | -1.17352 | -1.409 | 2.856264 | -0.56869 | -0.31643 | -0.8377 | -1.12844 | -1.53512 | -0.729 |
| 5 | 5.161781 | -0.2541 | -1.32668 | 0.593359 | -0.63457 | -0.76331 | 0.498107 | 1.114087 | -0.05177 | -1.409 | -0.34472 | -0.56869 | -0.31643 | -0.78473 | -0.82296 | -0.32547 | 0.79553 |
| 6 | 4.575772 | -0.2541 | -1.25956 | 0.455425 | 0.072216 | -1.35397 | 1.316296 | 2.248536 | -0.05177 | -1.409 | -0.34472 | -0.56869 | -0.31643 | -0.61057 | -0.27927 | 0.108721 | 1.34088 |
| 7 | 5.534617 | -0.2541 | -0.67472 | -0.44021 | 0.712268 | 0.418004 | 0.731875 | 0.357788 | -0.05177 | 1.98303 | -0.34472 | -0.56869 | -0.31643 | 0.62029 | 0.112362 | 0.216381 | 0.8398 |
| 8 | 5.111259 | -0.2541 | 0.458738 | -0.61164 | -1.08567 | -0.76331 | -0.67074 | -0.02036 | -0.05177 | -0.27832 | -0.34472 | -0.56869 | -0.31643 | -1.33828 | -0.86536 | -0.68124 | -0.7743 |
| 9 | 5.395774 | -0.2541 | -0.74576 | -0.36074 | -1.58962 | -0.76331 | -1.25516 | -0.77666 | -0.05177 | -0.27832 | -0.34472 | -0.56869 | -0.31643 | -1.51244 | -1.36514 | -1.11542 | -1.2095 |
| 10 | 5.149354 | -0.2541 | -1.69027 | 1.466382 | -1.09613 | -0.76331 | 0.03057 | 0.546862 | 1.069975 | -1.409 | -0.34472 | -0.56869 | -0.31643 | -0.89672 | -1.12239 | -0.62 | -0.3998 |
| 11 | 5.177178 | -0.2541 | -0.1623 | -0.77332 | -0.52261 | 0.418004 | -0.67074 | -0.96574 | -0.05177 | 0.852355 | -0.34472 | 0.663477 | -0.31643 | -0.11174 | -0.42321 | -0.20874 | -1.1516 |
| 12 | 5.50724 | -0.2541 | 0.754493 | -0.31684 | 0.136568 | 0.418004 | 0.147454 | 0.168713 | -0.05177 | -0.27832 | -0.34472 | -0.56869 | 3.111599 | 0.9854 | 0.144968 | -0.031 | -0.3578 |
| 13 | 5.512862 | -0.2541 | -0.13103 | -0.78 | 0.306182 | 1.402433 | -0.43697 | -1.15481 | 1.069975 | 0.852355 | -0.34472 | 0.663477 | -0.31643 | 0.456199 | -0.05483 | -0.1277 | -0.9056 |
| 14 | 4.666956 | -0.2541 | -0.32977 | -0.71086 | 0.813908 | 1.402433 | 0.614991 | -0.39851 | 2.191723 | 0.852355 | -0.34472 | 1.895649 | -0.31643 | 1.62329 | 1.173212 | 0.390985 | -0.021 |
| 15 | 4.921543 | -0.2541 | -1.03832 | 0.052005 | 0.25071 | 0.418004 | 0.264338 | -0.20944 | 1.069975 | 1.98303 | -0.34472 | -0.56869 | -0.31643 | 0.508299 | -0.20109 | -0.04442 | -0.3633 |

**Fig 3.14**: **Excel Worksheet**

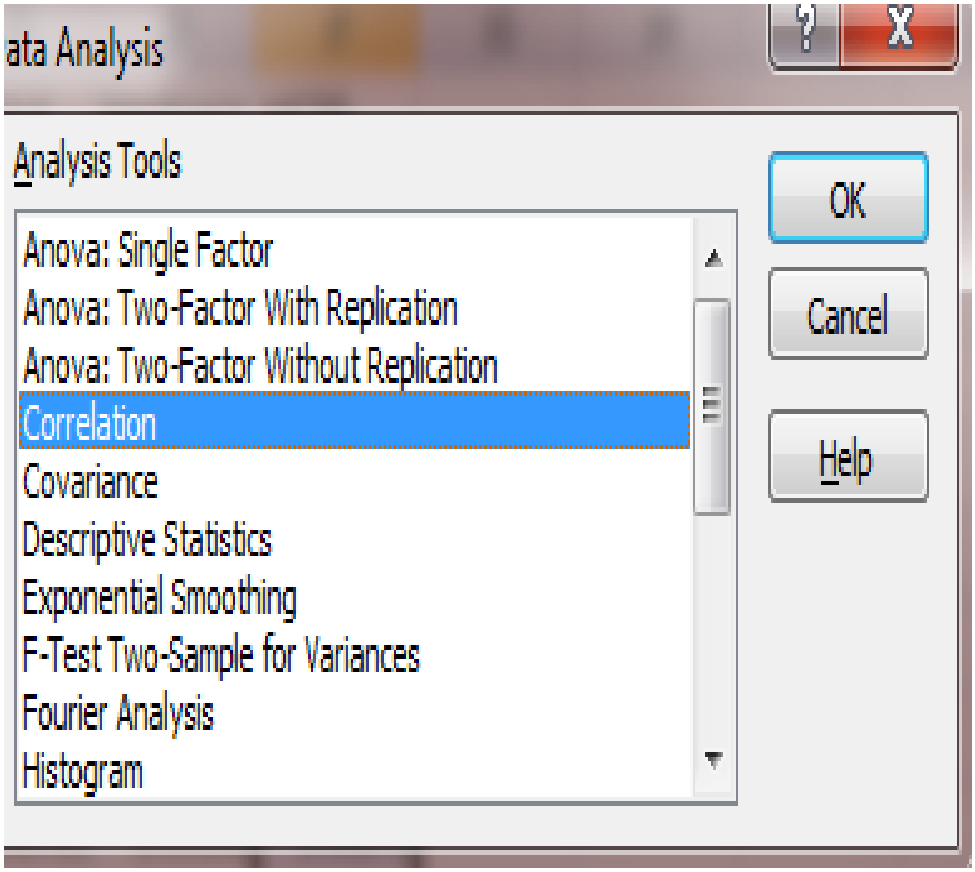



**Fig. 3.15**: **Data Analysis Tools**

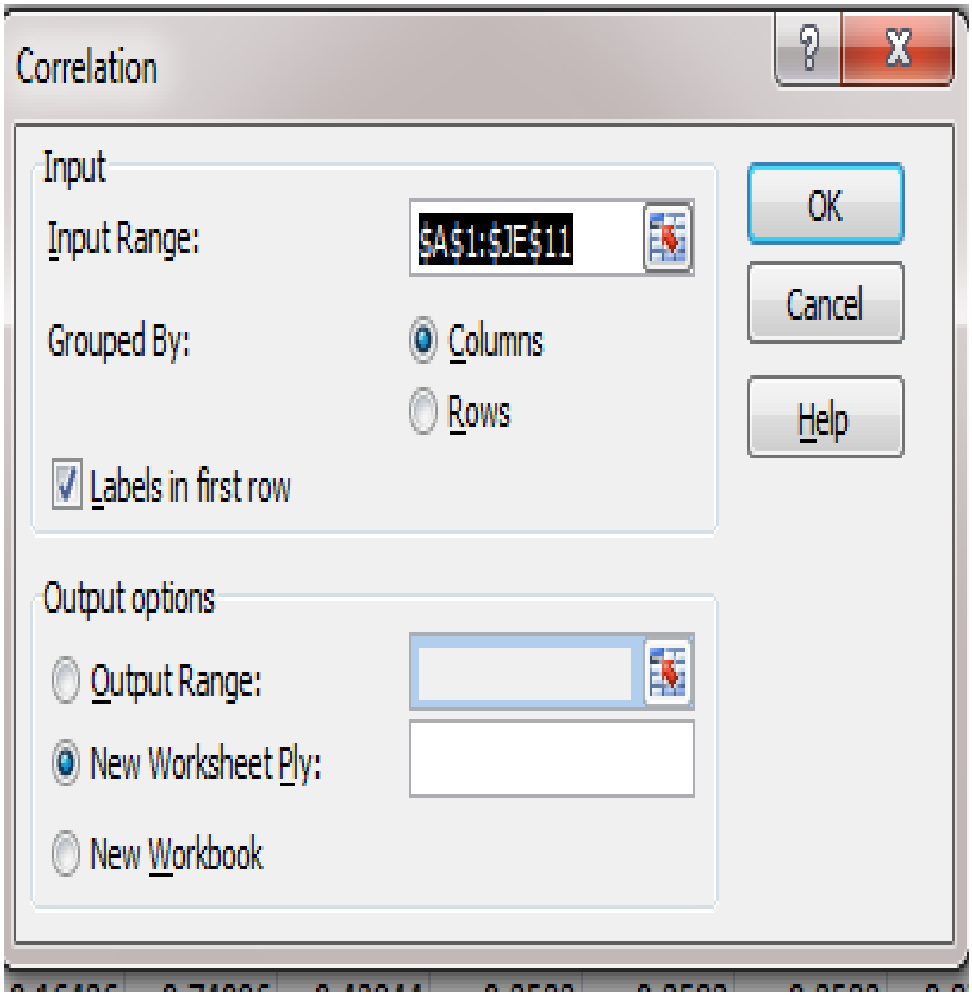



**Fig 3.16**: **Correlation Dialogue Box**

### 3.2.7 Variable elimination and selection

Variable elimination is the process where variables are physically removed from the dataset. Variables are eliminated for two reasons. First, they are eliminated if they have a small variance, below some threshold value. The second reason for variable removal is the existence of redundancy (exact linear dependencies between subsets of the variables) and multicollinearity (high multiple correlations between subsets of the variables) in QSAR data sets. The elimination of the independent variables (i.e., descriptors) was done by removing invariable (constant column) and any descriptors that are cross-correlated by R = 0.6 or more .The jar file (vWSP.*jar*) present in the vWSP program (Ambure *et al.,* 2015) folder was *double clicked* opening a window (Fig 3.17) with few queries. On the "***Select Query Input File***" the 'browse' button was *clicked* to select the input file. On the "***Select Output Directory***" the 'browse' button was *clicked* to select the destination/output file directory and define output file name. On the "**Enter Variance cut-off"** section,the variance cut-off value which determines the constant variables to be removedwas set to *0.0001*. On the "**Enter correlation coefficient cut-off"** section, the inter-correlation coefficient cut-off value which determined the inter-correlated variables that were removed was set to *0.60*. The start button was clicked to complete the task.

### 3.2.8 Development of a quantitative structure- activity relationship model

This study was done in four categories;

i) QSAR study of *in vitro* antimycobacterial activity (MIC) 8-methylquinolone carboxylic *against Mycobacterium tuberculosis H37Rv* (MTB).

ii) QSAR study of *in vitro* antimycobacterial activity (MIC) of some 8 -Methylqinolone carboxylic acid against *Mycobacterium Smegmatis* ATCC 14468

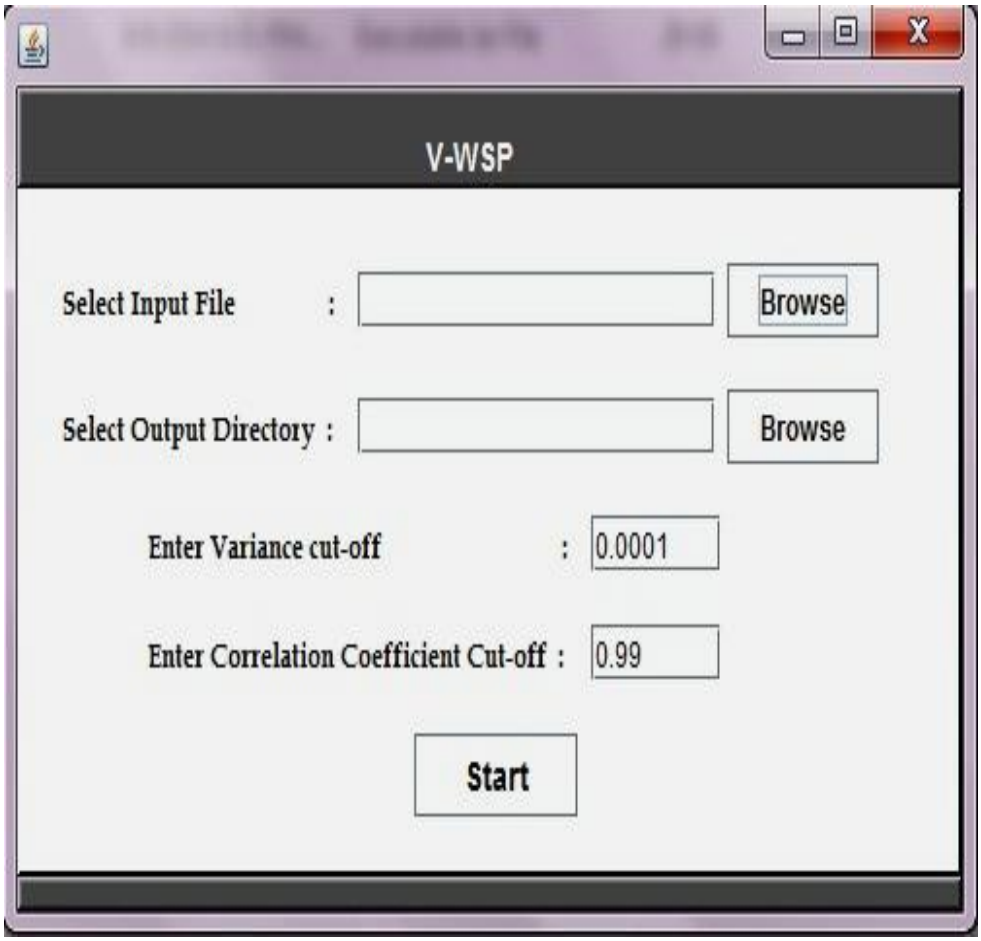



**Fig 3.17 Queries Window**

*iii)* QSAR study of the *in vitro* antimycobacterial activity (MIC ) of a mixture of 8-Methylquinolone and Nitroquinolones carboxylic acids against *Mycobacterium tuberculosis* H37Rv(MTB)

iv) QSAR study of the *in-vivo* cytotoxicity (IC50) study of some 8-Methylquinolones.

**3.2.8a Training and test set selection**

Prior to using a quantitative structure activity relationship (QSAR) model for external

Predictions, its predictive power should be established and validated. In the absence of a trueexternal data set, the best way to validate the predictive ability of a model is to perform its statistical external validation. In statistical external validation, the overall data set is divided into training and test sets. Commonly, this splitting is performed using random division (Todd *et al.*, 2012).In this study each data set was randomly divided into trainingand external evaluation set that included 80% and 20% of theentire data set, respectively, seeTable 3.2

**3.2.8bDescriptor selection and Model Building**

The selection of significant descriptors, which constructs a relationship between the biological activity of the data and its molecular structures, is an important step in QSAR modeling.For this purpose, variable selection Genetic Algorithm (GA) and Multiple Linear Regression Analysis (MLRA) methods were used.

The GA has been applied widely in many areas, and specifically, in recent studies as a powerful tool to address many problems in cheminformatics.GA is an optimization algorithm based on the evolutionary mechanisms as proposed by Darwin, it uses random mutation, crossover and selection procedures to produce better models or solutions from an originally random starting population or sample. (Davis, 1991; Devillers, 1996).

**Table 3.2Data matrix is divided in to a training set and an independent test set**

| Compounds | Target | Dependent variable (Y) | Descriptors Pool | Original Matrix Data | Training set Compounds | Test Set Compounds |
|---|---|---|---|---|---|---|
| 8-methylquinolone carboxylic acid analogues | *Mycobacterium tuberculosis* H37Rv (MTB) | pMIC | 1598 | 37 | 28 | 9 |
| 8-methylquinolone carboxylic acid | *Mycobacterium smegmatis* ATCC 14468 ($MC^2$) | pMIC | 1598 | 34 | 27 | 7 |
| 8-methylquinolone carboxylic acid/ nitroquinolones | *Mycobacterium tuberculosis* H37Rv (MTB | pMIC | 1598 | 45 | 34 | 11 |
| 8-methylquinolone carboxylic acid analogues | *Invivo* cytotoxicity | $pIC_{50}$ | 1598 | 15 | 10 | 5 |

At the end the MLRA is used to correlate the descriptors selected by the GA with the biological activity of interest. The MLR, also known as linear free-energy relationship method, is an extension of simple linear regression analysis to more than one dimension (Berk, 2003).

MLR generates QSAR equations by performing standard multivariable regression calculations to identify the dependence of a drug property on any or all of the descriptors under investigation (Thangapandian *et al*., 2012). The combination of the GA-MLRA technique was employed for preliminary model selection, as implemented in the BuildQSAR software(De Oliveira and Gaudio, 2001). The selected equations were analyzed using the following statistical data: correlation coefficient ($R$), squared correlation coefficient ($R^2$), standard deviation ($s$), Fischer's test ($F$), level of confidence ($p$), squared cross– correlation coefficient ($Q^2$) and standard deviation of sum of square of difference between predicted and observed values (Spress).

The BuildQsar programe was launched by double clicking on the icon  present on the desktop of a laptop. From the **Menu Bar** Fig.3.18**File menu** was *selected* followed by selection of the **New**from the drop-down menu, this resulted in the **dialog box** in Fig.3.19.On the **Dataset size** section, the number of activities,descriptors and compounds were *specified*accordingly. The **Ok** was *clicked* which brought up an excel-like worksheet with a number of columns and rows equal to the number of descriptors and compounds respectively. The file containing the calculated descriptors of the molecules of the training set with their corresponding activity was opened; the values of the activity were copied and pasted in the first column(yellow) of the excel-like sheet in Fig. 3.20. The descriptor values were copied and pasted in the subsequent columns. On the **Menu Bar** the **QSAR Menu** was selected, followed by **Variable Selection** then **Genetic Algorithm** Fig. 3.20.

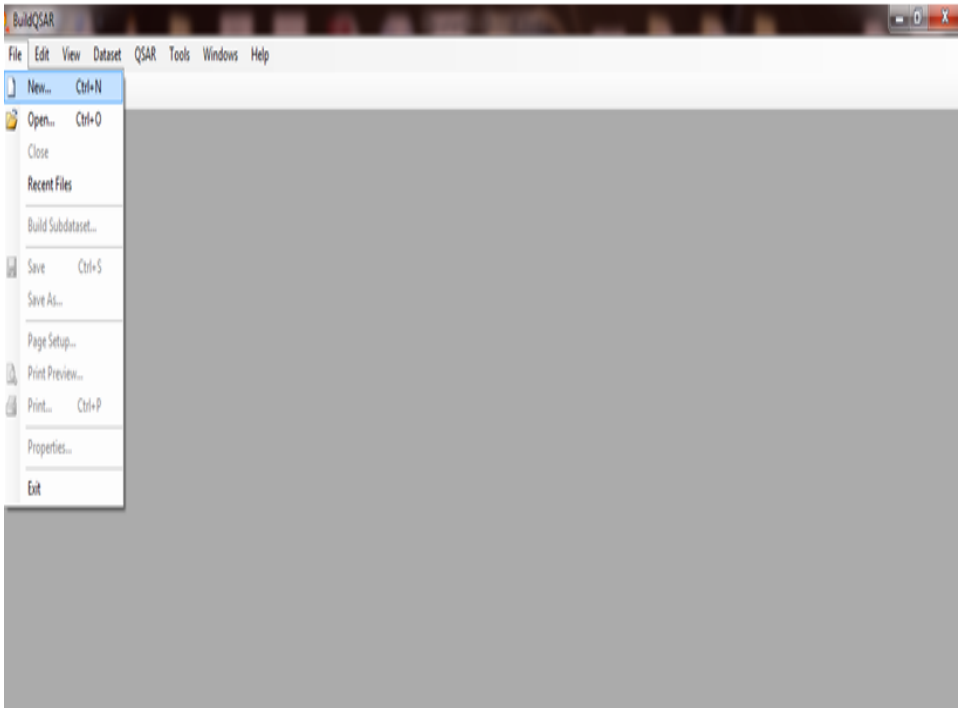



**Fig. 3.18BuildQsar interface**

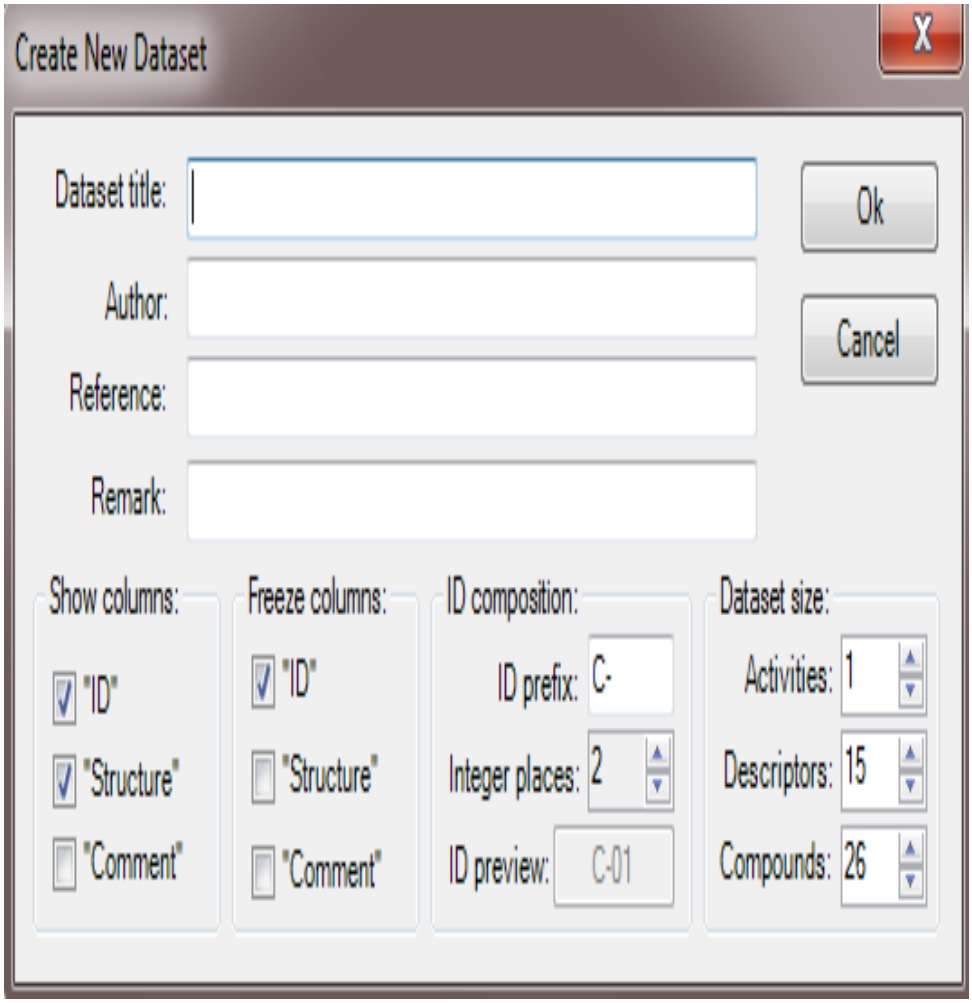



**Fig. 3.19:Create Dataset Dialogue Box**

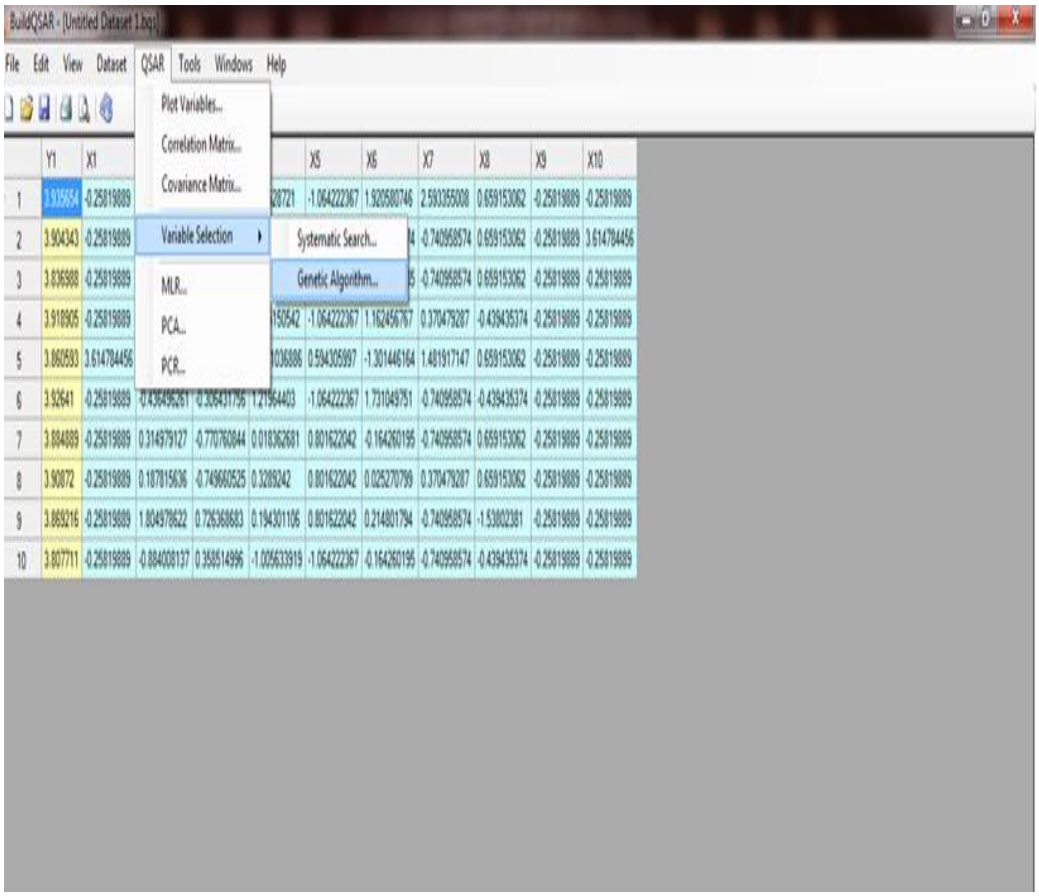



**Fig.3.20**:**BuildQsar Interface-B**

This brought up the **Dialog box** in Fig.3.21. On the **Build genetic search****Dialog Box** (Fig 3.21) the **number of descriptors per model, number of generation** and **models per generation** were all *specified*. For the QSAR study of 8-Methylquinolone Carboxylic against *Mycobacterium tuberculosis H37Rv* (MTB) the number of descriptors per model was set at five, four and three respectively. For the QSAR study of 8 -Methylquinolone carboxylic acid against *Mycobacterium Smegmatis ATCC 14468 (MC$^2$)* the number of descriptors per model was first set at five, four and three respectively . For the QSAR study of a mixture of 8-Methylquinolone Carboxylic acids and Nitroquinolones against *Mycobacterium tuberculosis H37Rv* (MTB) the number of descriptors per model was set at six, five, four, three and two respectively. For the QSAR study of the of cytotoxicity ($IC_{50}$) study of 8-Methylquinolones, the number of descriptor per model was set at two and one respectively.In all these, the **number of generation** was set at 1000, **Models per generation** was set at 3, and correlations higher than 0.6 were forbidden. The **Runoption** in Fig.3.21 was *clicked* to generate the models in each case. The best three models were selected in each case for further studies.

### 3.2.8c Internal Validation

The selected models were internally validated using the cross-validation leave-one-out procedure. In the leave-one-out (LOO) method of cross validation, the process of removing a molecule, and creating and validating the model against the individual molecules is performed for the entire training set. Once completed, the mean of all the $Q^2$ values is taken. (Ravichandran*et al*., 2011). The $Q^2_{LOO}$ was calculated using the equation (10) Material Studio software 7.0

$$Q_{tr}^2 = 1 - \frac{\sum_{i=1}^{N_{tr}} (y_{i,exp} - y_{i,pred})^2}{\sum_{i=1}^{N_{tr}} (y_{i,exp} - \bar{y}_{i,exp})^2} \quad \ldots\ldots\ldots\ldots (10)$$

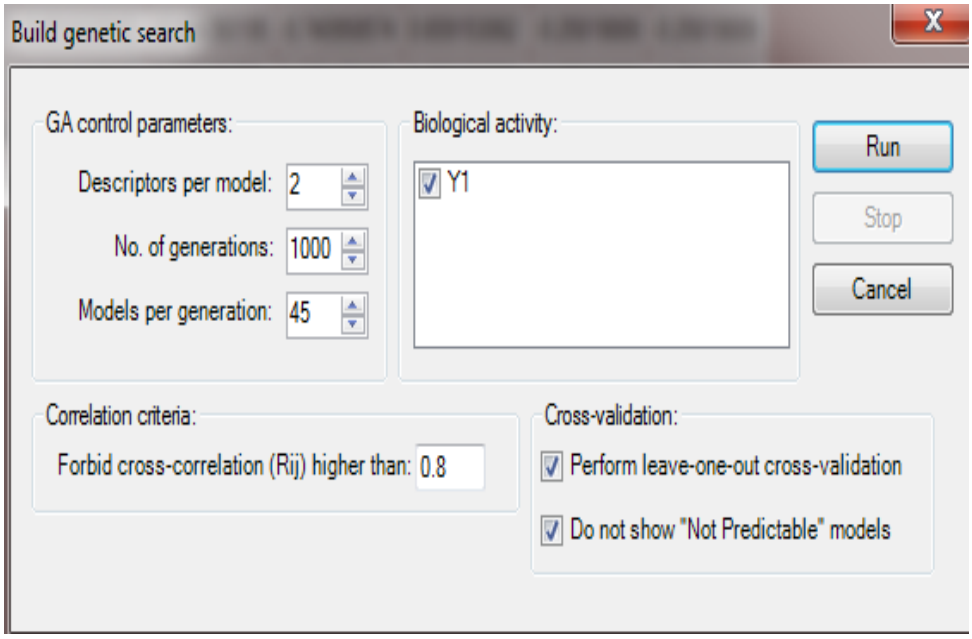



**Fig 3.21**:**Build Genetic Search Dialogue Box**

.Where $N_{tr}$is the totalnumber of training set objects; $y_{i,exp}$ and$y_{i,pred}$ are the experimental and predicted values, respectively; $\bar{y}_{i,exp}$ is the average response value of the training set. The Material studio software was launched from the the desktop of a laptop.The **'welcome to material studio'** dialog box shown in Fig.3.22 appeared, from which the **create a new project** option was *selected* and **ok** was *clicked.* This action brought up the **File browser** in Fig.3.23, on the **File name** section of this**File browser**, the project was named accordingly and *stp was chosen in the section **Save as type**. On the **Menu bar** (Fig.3.24) the **File menu** was selected to bring a drop-down menu from which **New**was *selected*. This operation produced the dialog box in Fig.3.25, and on this Dialogbox,**Study Table** was *chosen,***Ok** was *clicked* and the worksheet in Fig.3.26appeared. The values of the descriptors were copied and pasted in subsequent columns as shown in Fig.3.27. The Column containing the values of the activities was highlighted. **Model Building** was *selected,* this resulted in another drop-down menu from which **Multiple Linear Regression** was *selected* see Fig.3.28. This action produced the dialog shown in Fig. 3.29. On the **Menu bar**, the**Statistics Menu** was *selected* resulting in a drop-down Menuin Fig.3.29. From **the dialog box**(Fig.3.29)**Variable Selection** was chosen and the boxes corresponding to each descriptor were *checked*. Next the **stepping** option on the dialog box was *selected* and the dialog now appeared as is shown in Fig.3.30. On the **Type of Stepping** section **2-way** was *selected*, both **F to enter** and **F to leave** were set at **4.** On the **Exclude if Correlation greater than** section, 0.6 was *imputed*. The Maximum number of steps was set at **10**. Next the **Validation** option on the dialog box was *selected* and the dialog now appeared as shown in Fig.3.31.

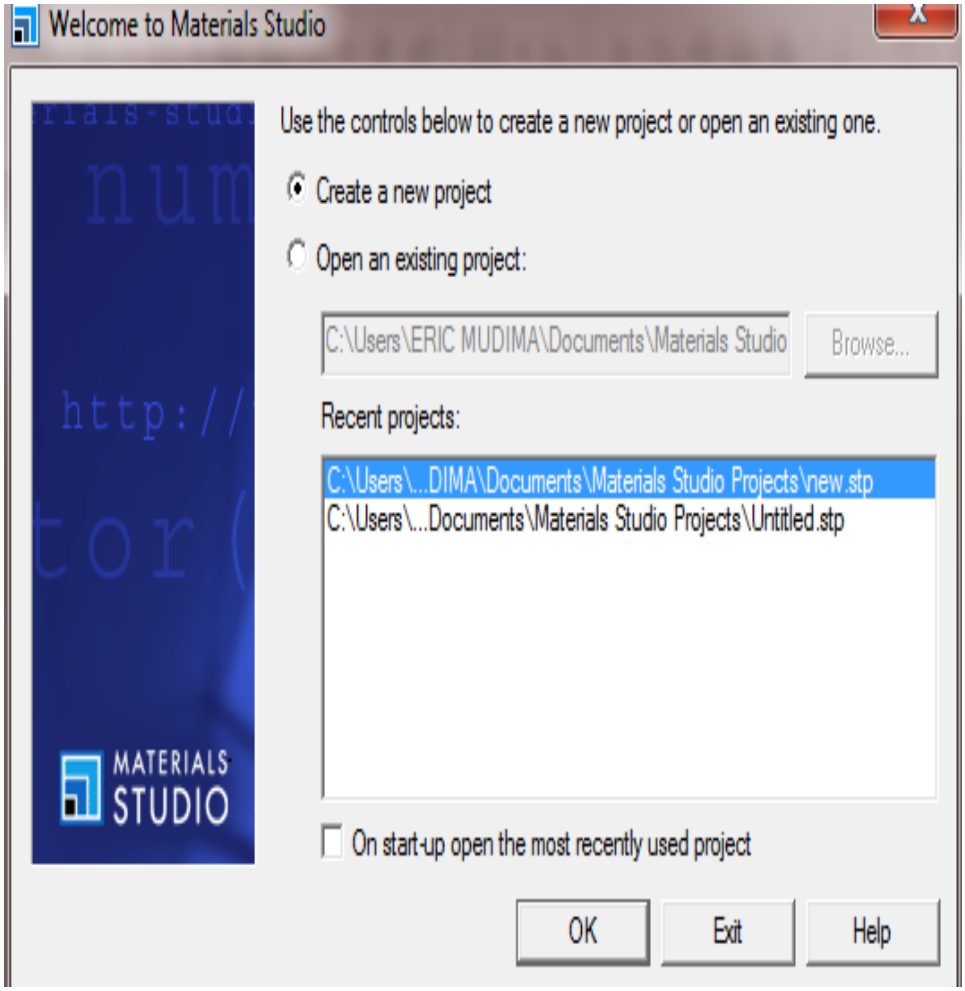



**Fig. 3.22**: **Welcome to material studio Dialogue box**

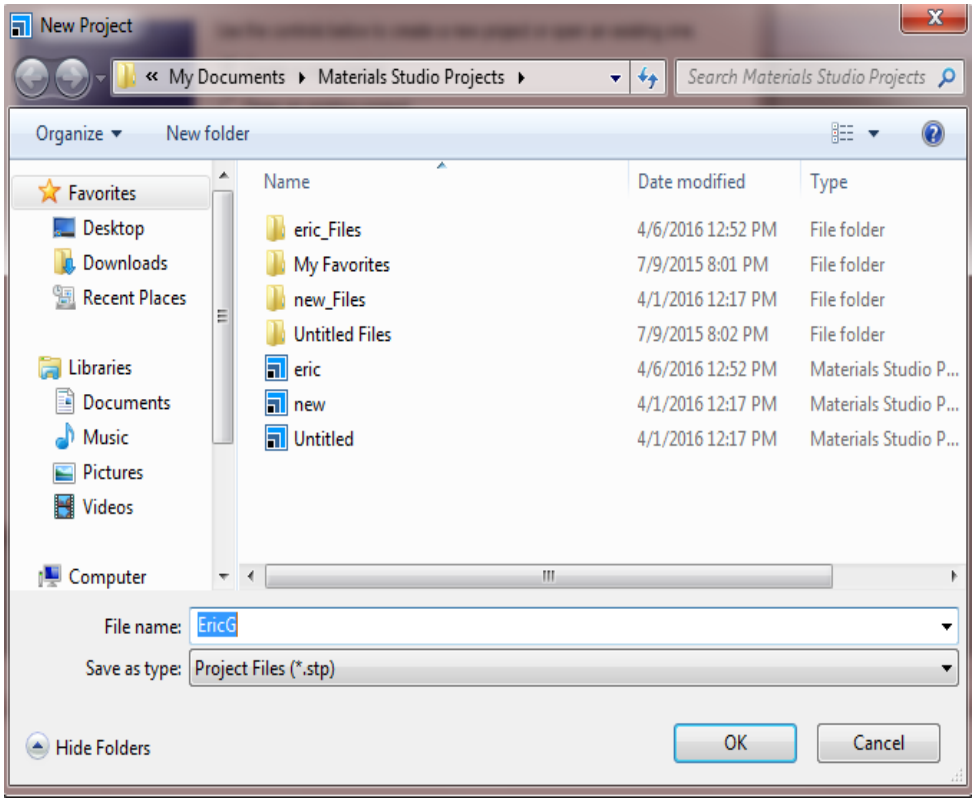



**Fig. 3.23**:**New Project File Browser**

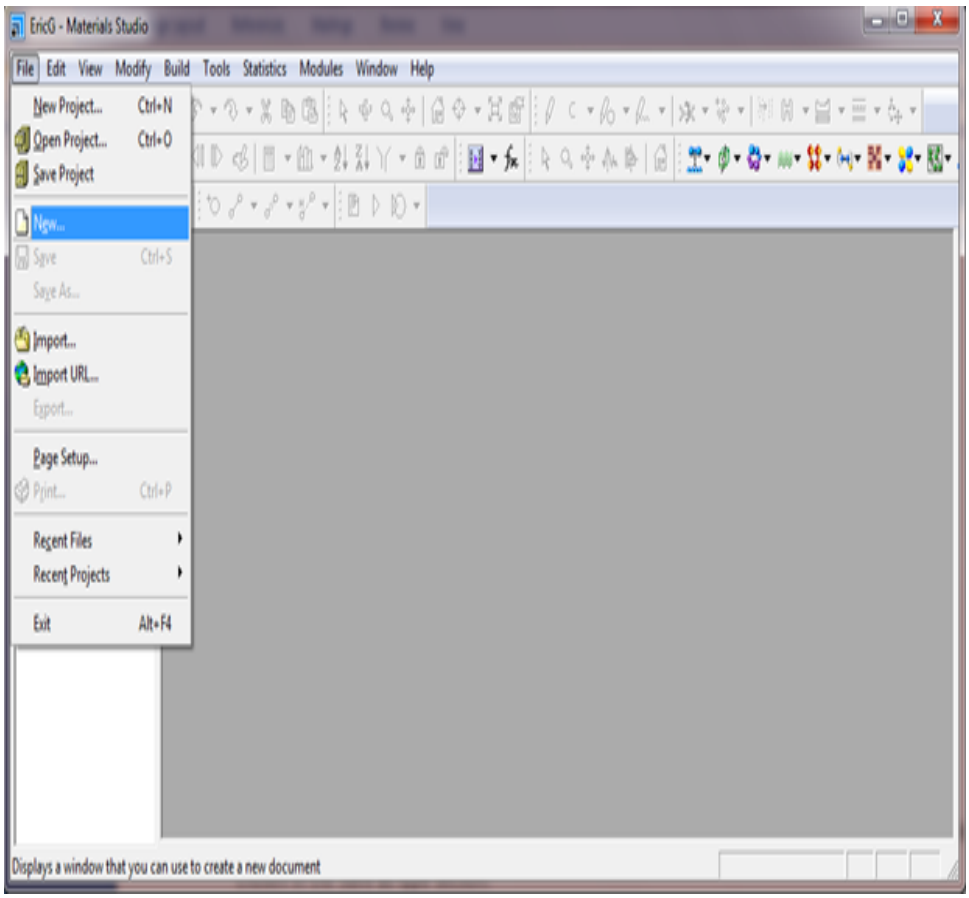



**Fig. 3.24**:**Material Studio Dialogue Box**

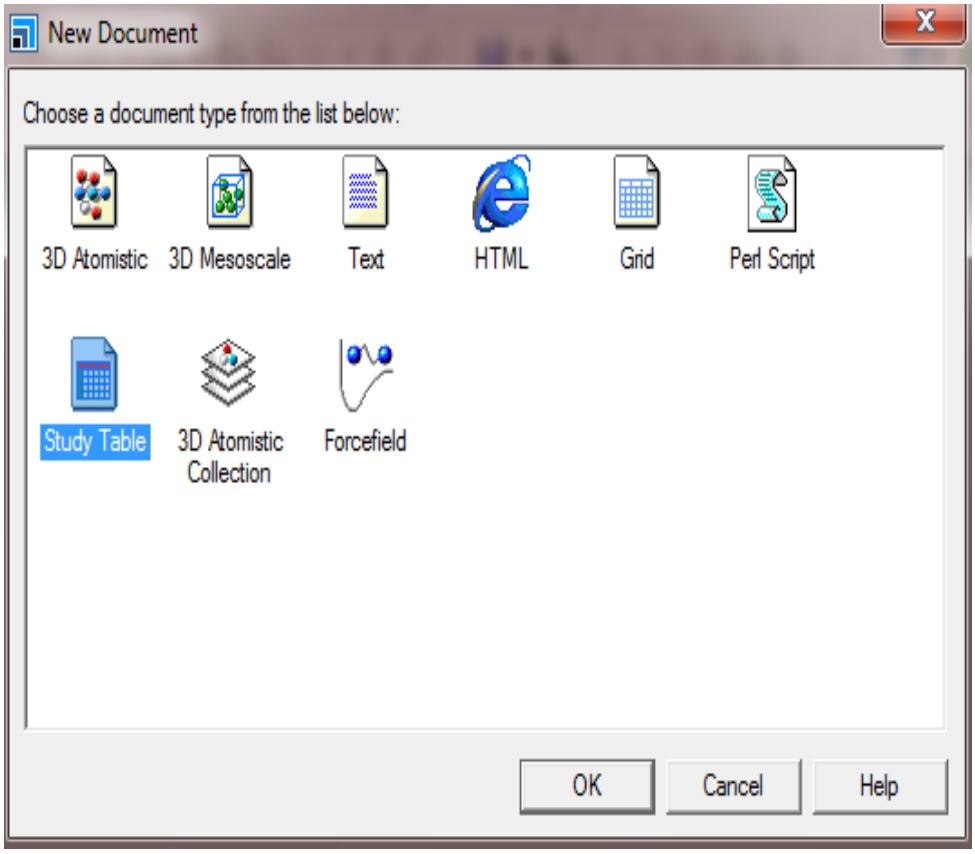



**Fig. 3.25:New Document Dialogue Box**

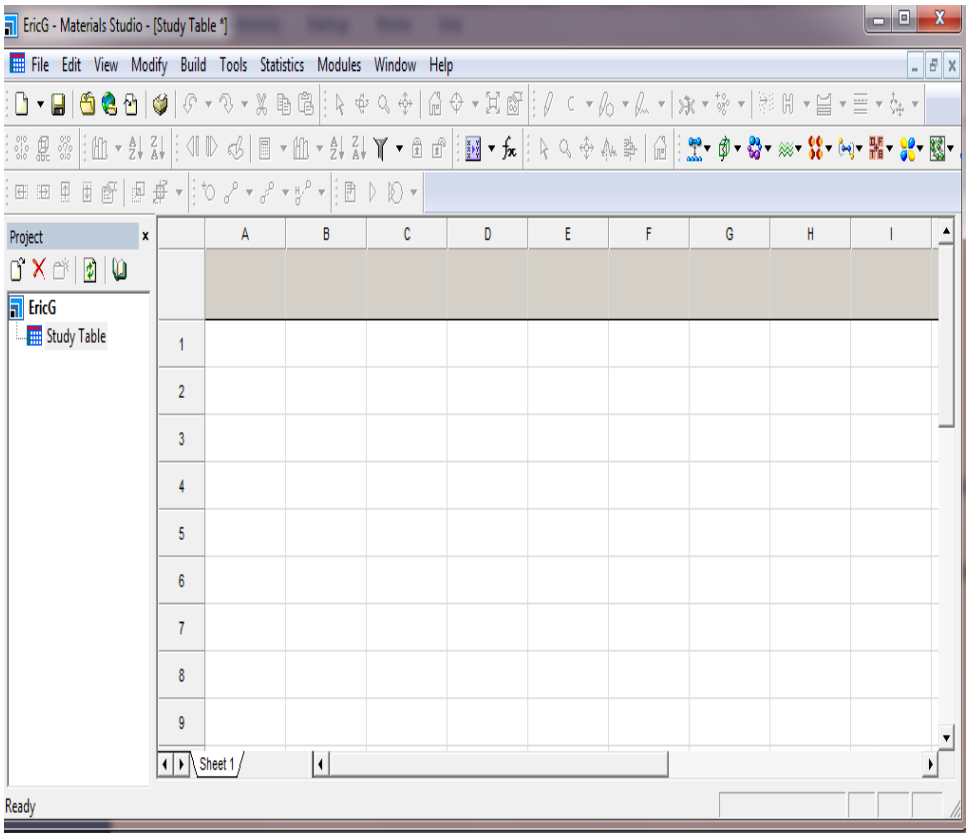
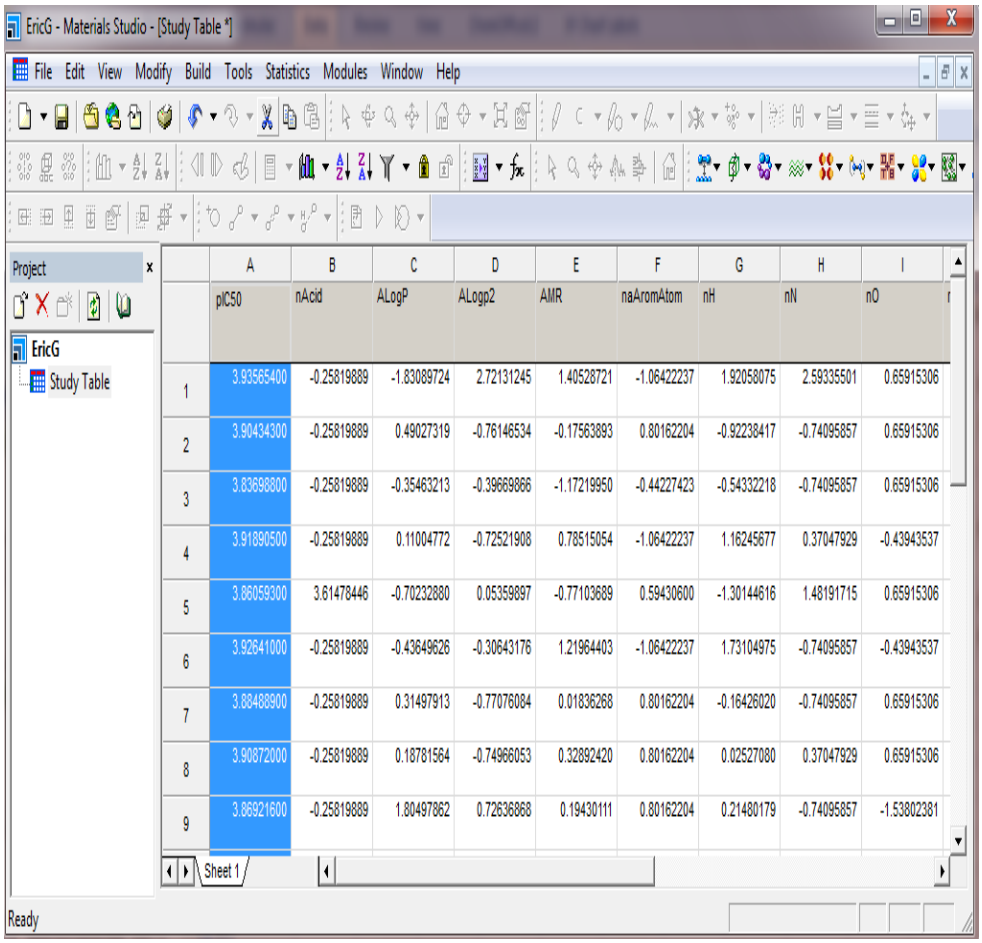



**Fig. 3.26:Material Studio Study Table-A**




| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| | pIC50 | nAcid | ALogP | ALogp2 | AMR | naAromAtom | nH | nN | nO |
| 1 | 3.93565400 | -0.25819889 | -1.83089724 | 2.72131245 | 1.40528721 | -1.06422237 | 1.92058075 | 2.59335501 | 0.65915306 |
| 2 | 3.90434300 | -0.25819889 | 0.49027319 | -0.76146534 | -0.17563893 | 0.80162204 | -0.92238417 | -0.74095857 | 0.65915306 |
| 3 | 3.83698800 | -0.25819889 | -0.35463213 | -0.39669866 | -1.17219950 | -0.44227423 | -0.54332218 | -0.74095857 | 0.65915306 |
| 4 | 3.91890500 | -0.25819889 | 0.11004772 | -0.72521908 | 0.78515054 | -1.06422237 | 1.16245677 | 0.37047929 | -0.43943537 |
| 5 | 3.86059300 | 3.61478446 | -0.70232880 | 0.05359897 | -0.77103689 | 0.59430600 | -1.30144616 | 1.48191715 | 0.65915306 |
| 6 | 3.92641000 | -0.25819889 | -0.43649626 | -0.30643176 | 1.21964403 | -1.06422237 | 1.73104975 | -0.74095857 | -0.43943537 |
| 7 | 3.88488900 | -0.25819889 | 0.31497913 | -0.77076084 | 0.01836268 | 0.80162204 | -0.16426020 | -0.74095857 | 0.65915306 |
| 8 | 3.90872000 | -0.25819889 | 0.18781564 | -0.74966053 | 0.32892420 | 0.80162204 | 0.02527080 | 0.37047929 | 0.65915306 |
| 9 | 3.86921600 | -0.25819889 | 1.80497862 | 0.72636868 | 0.19430111 | 0.80162204 | 0.21480179 | -0.74095857 | -1.53802381 |

**Fig. 3.27:Material Studio Study Table-B**

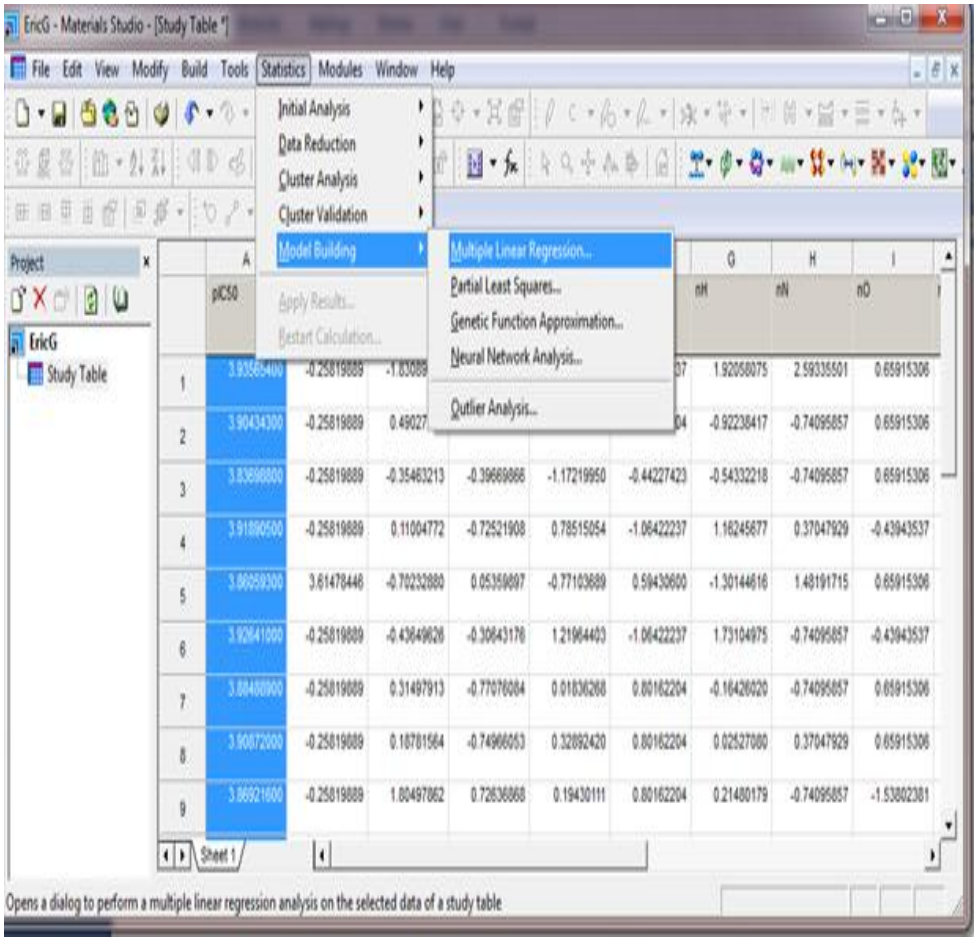



**Fig. 3.28:Material Studio Study Table- C**

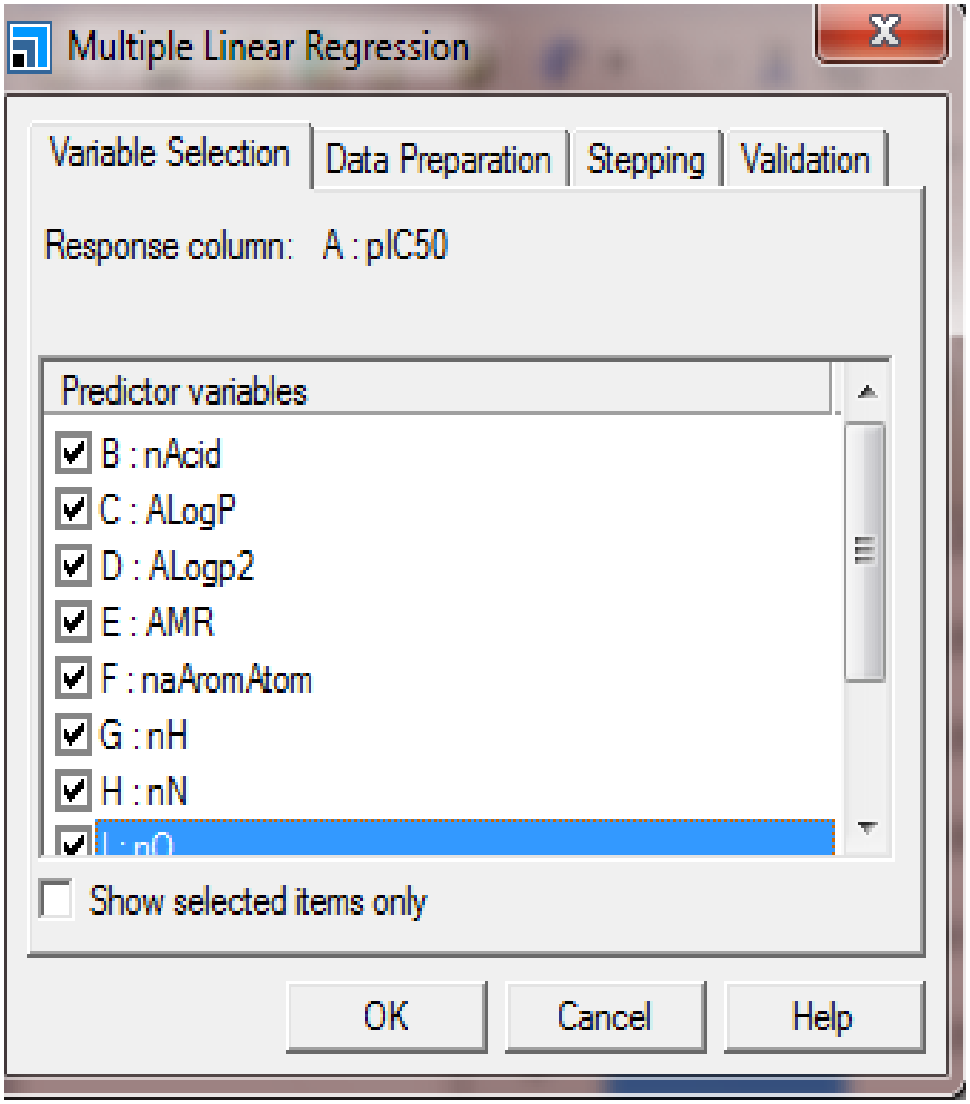



**Fig. 3.29**:**Multiple Linear Regression Dialogue Box-A**

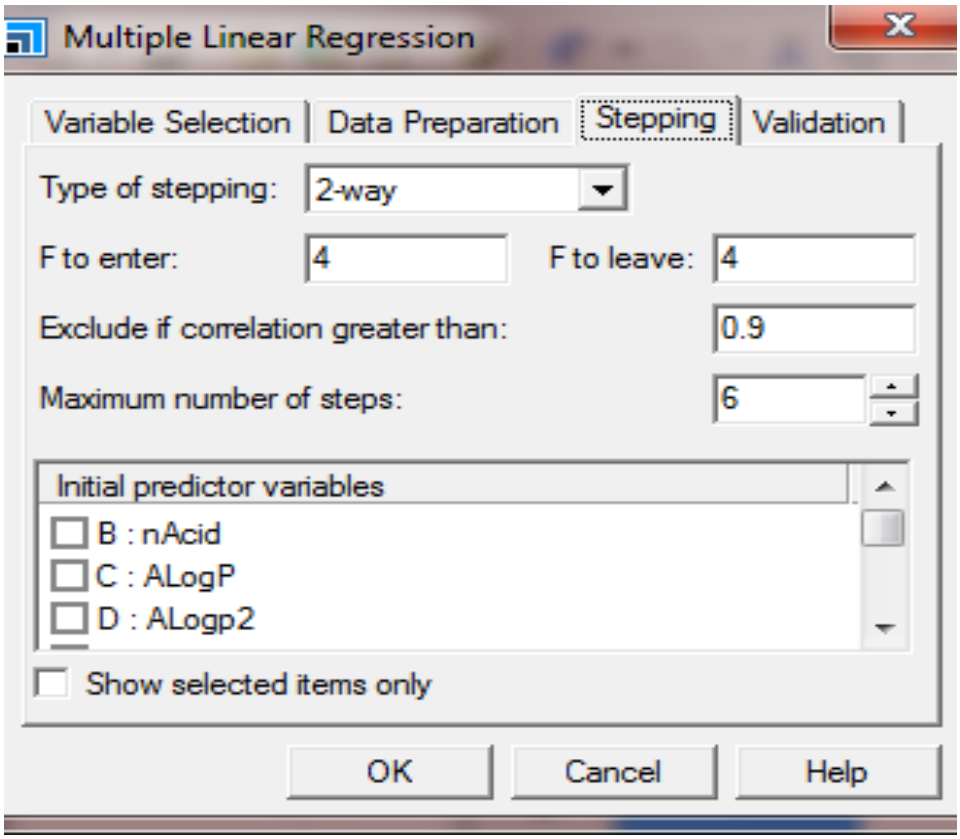



**Fig. 3.30:Multiple Linear Regression Dialogue Box-B**

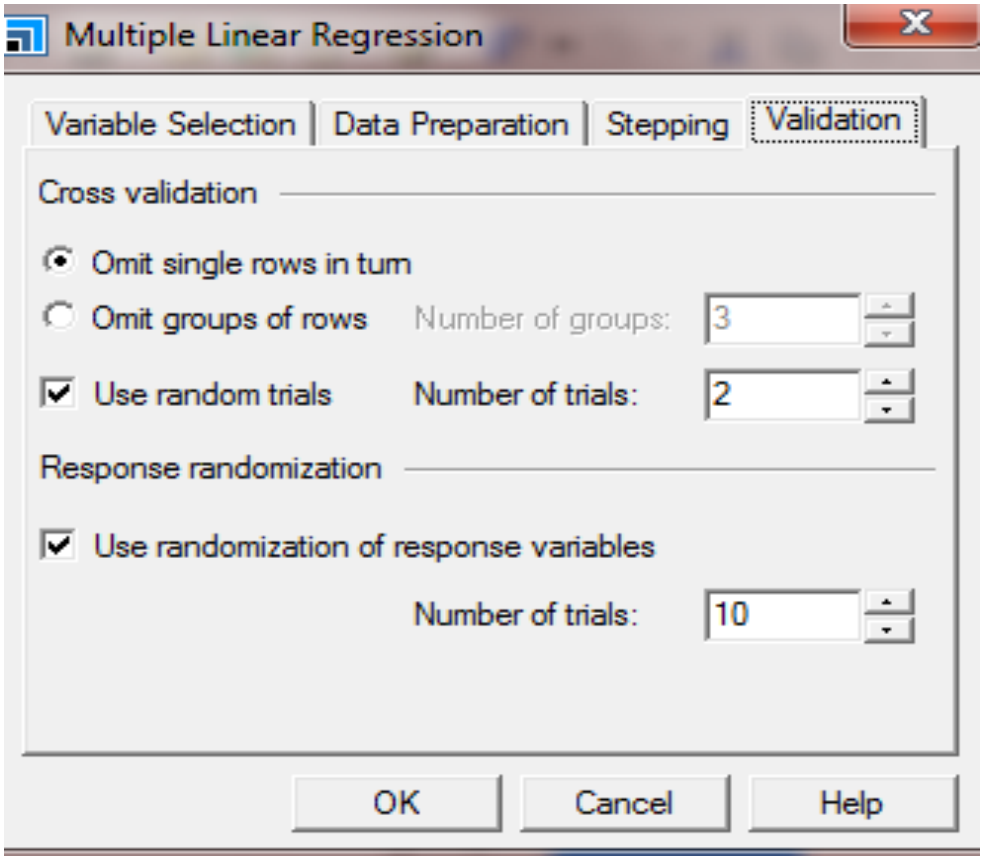



**Fig. 3.31**: **Multiple Linear Regression Dialogue Box-C**

. The **Omit single rows in turn** section was*checked*, **use random trials** was *checked* and the **number of trials** was set at **5**, **use randomization of response** variable was selected and to enable Y-randomization procedure to be performed, the **number of trials**was set at **10**. The **ok**button was *clicked* and the result was displayed on the screen.

### 3.2.8d External Validation

The models were externally validated by testing the previously excluded compounds which form the test set. The value of $R^2_{Pred}$ which gives an indication of the predictive power of a model was calculated using equation (11) (Schüürmann *et al*., 2008)

$$R^2_{Pred} = 1 - \frac{\sum_{i=1}^{N_{ext}} (y_{i,exp} - y_{i,pred})^2}{\sum_{i=1}^{N_{ext}} (y_{i,exp} - \bar{y}_{i,pred})^2} \quad \text{.....( 11)}$$

where $N_{ext}$is the total number of external validation set objects; $y_{i,exp}$ and $y_{i,pred}$are the experimental and predicted values, respectively;$\bar{y}_{i,pred}$is the average response value of the external validation set.

The selection of robust and well predictive QSAR models on the basis of only $R^2$, $Q^2$ and $R^2_{pred}$ might mislead the search for the ideal predictive model, so additional statistical parameters have been suggested by Golbraikh andTropsha (2010) ; a model that must be regarded as truly predictive must satisfy the following conditions.

1. $Q^2 > 0.5$
2. $R^2_{pred} > 0.6$
3. $|R_0^2 - R_0'^2| < 0.3$
4. $\left[\frac{|R_0^2 - R_0'^2|}{R^2} < 0.1\right]$ and 0.85<k<1.15 or [ $(R^2 - R'^2_0)/R^2$)<0.1] and 0.85<k'<1.15

where R is the correlation coefficient between the predicted and observed activities, $R_0^2$ is coefficients of determination (predicted versus observed activities, and observed versus predicted activities $R'_0{}^2$ for regressions through the origin), k and k' slopes of regression lines through the origin. These additional statistical parameters were calculated. This was done with the aid of the program DTC MLRplusValidation(Ambure*et al.*,2015). The inputfile, containing the observed and predicted activity of the test compound, was named and saved in the Xternal-Validation-Metric-Calculator-1.0 program folder. The executable jar file contained in the folder was *double clicked*. This produced the dialog box in Fig.3.32.The name of the input file was *typed* and **ok** was *clicked*, this brought up a prompt message as shown in Fig3.33**. Yes** was *clicked* and the dialog in Fig. 3.34 came up; the average value of the training set activities was imputed and **ok** was *clicked*. The result of the calculation showingtheExternal Validation Parameters and also Golbraikh and Tropsha acceptable model criteria was displayed.

**3.2.8e Y-Randomization**

The final Model was validated by the Y-randomization (randomization of response). This approach is widely used to validate the robustness of QSAR models (Wold and Ericksson, 1995).

It consists of rebuilding models using randomized activities of the training set and subsequent assessment of the model statistics.It is expected that models obtained for the training set with randomized activities should have significantly lower values of pMIC for the training set than the models built using training set with real activities.If this condition is not satisfied, models built for this training set with real activities are not reliable and should be discarded.

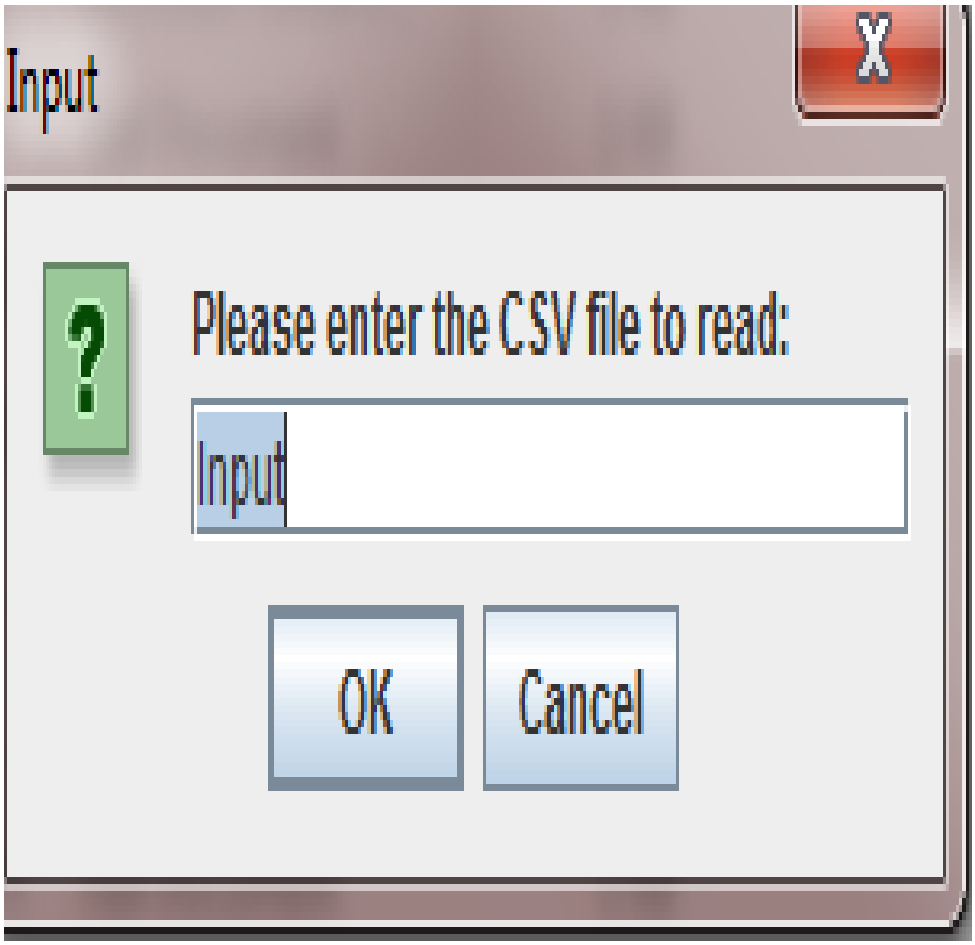



**Fig. 3.32:DTC MLRplusValidation Dialogue Box-A**

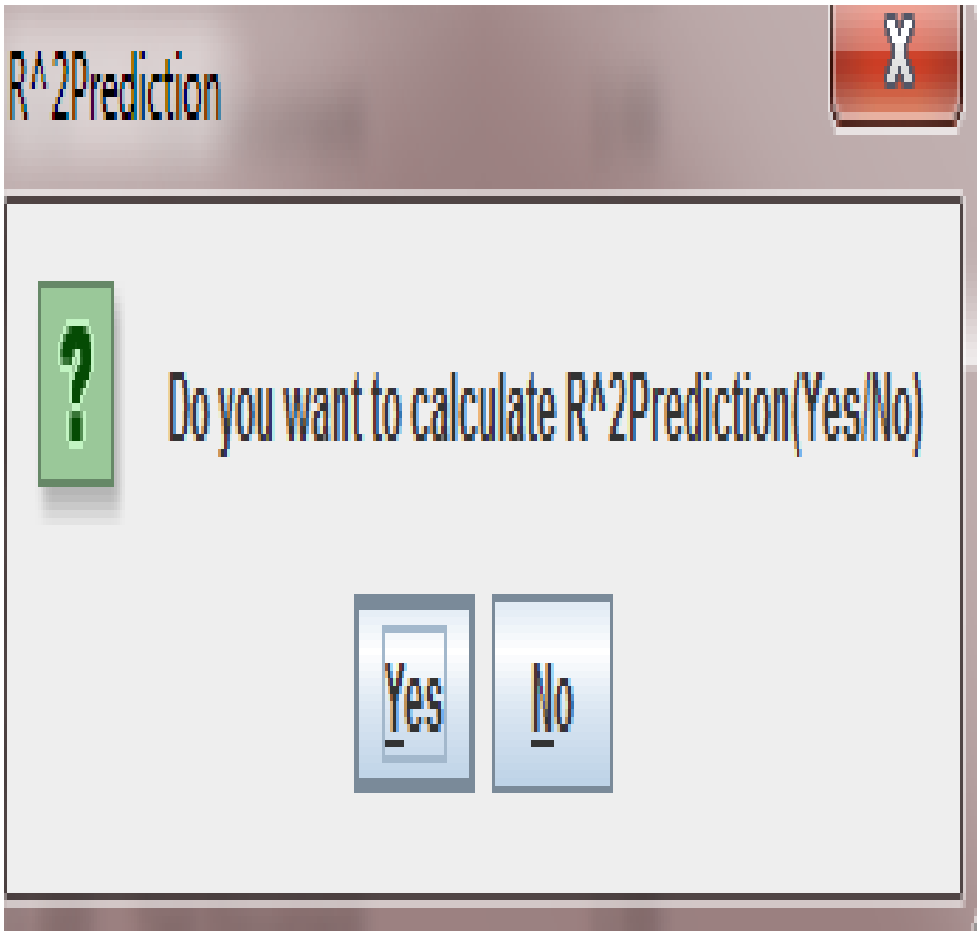
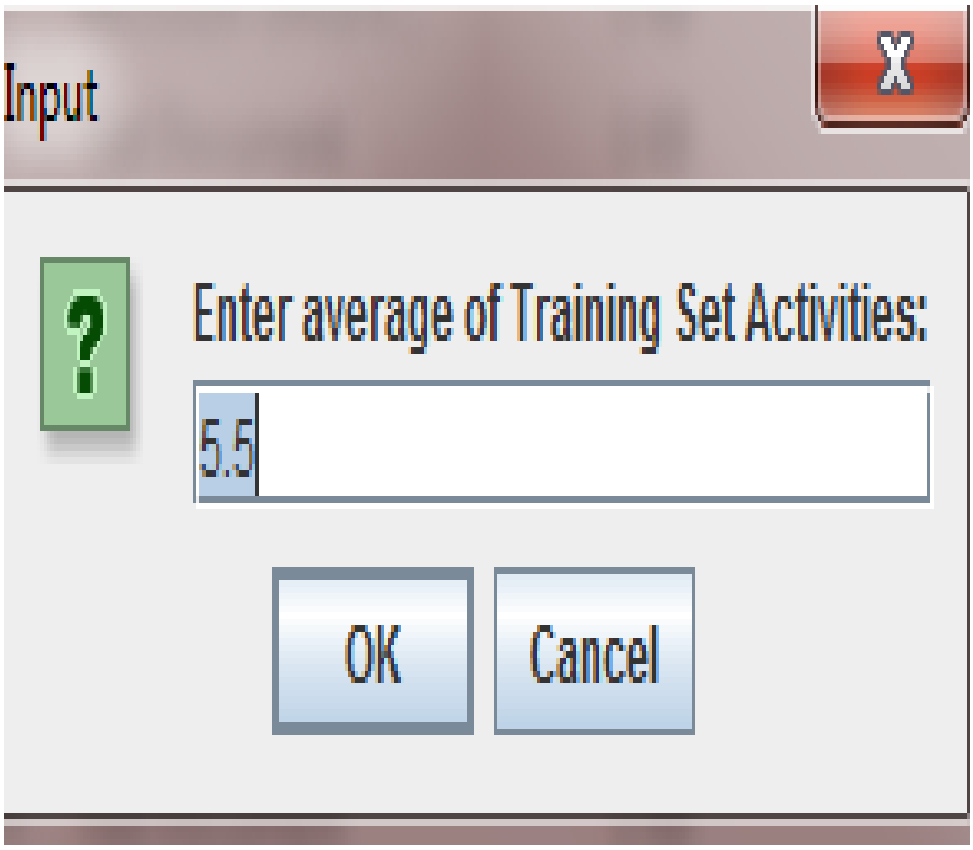



**Fig. 3.33**:**DTC MLRplusValidation Dialogue Box-B**




. **Fig. 3.34**:**DTC MLRplusValidation Dialogue Box-C**

In addition, the parameter $R^2_p$ (Roy,2007) which penalize the model $R^2$ for the difference between squared mean correlation coefficient ($R^2_r$) of the randomized models and square correlation coefficient ($R^2$) of the non-randomized model given (Roy,2007) by eq ( 12 ) was also calculated for the model.

$$R_P^2 = R^2 * \left(1 - \sqrt{|R^2 - R_r^2|}\right) \text{ …….. (12)}$$

A value of$R_P^2$ greater than 0.5 may be taken as an indicator of model acceptability. Y-randomization was done with the aid of Material Studio Software as described in section3.2.8.3.1 above.

**3.2.8f Applicability domain**

Beside these facts, the quality as well as reliability of the QSAR models must be also taken into account (Eriksson *et al*., 2003). Therefore, the OECD principles for validation of the QSAR models for regulatorypurposes clearly stated that a model should be used within the boundaries of its applicability domain (AD) (OECD, 2004). According to the Setubal's Workshop guidance and acceptability criteria for application of QSAR models for chemical management purposes, the AD of a QSAR model is defined as a "physico-chemical, structural, or biological space, knowledge or information on which the training set of the model has been developed, and for which it is applicable to makepredictions for new compounds. The AD of a QSAR should be described in terms of the most relevant parameters, i.e., usually those that are descriptors of the model" (Eriksson*et al*., 2003).The applicability domain of a (Q)SAR model is a theoretical region in chemical space, defined by the modeldescriptors and modeled response, and thus by the nature of the chemicals in the training set, as represented in each model by specific molecular descriptors. Thus, only the predictions for chemicals falling within this domain can be considered reliable and

not model extrapolations (Gramatica, 2007).Without the restriction placed by the applicability domain, QSAR models can predict the activity of any compound even if such a compound is structurally different from those included in the training set. This would lead to unreasonable extrapolation of the model in chemistry space and therefore heighten the chances of inaccurate predictions. Thus for a QSAR model to give reliable outcome, its applicability domain must be defined.One of the recommended hat-based methods for AD investigation in case of linear QSAR models is the widely known leverage approach. In this study, the applicability domain of the final model was calculated by the leverage approach (Eriksson *et al*., 2003). The leverage value of every compound was calculated and plotted against the standardized residuals (William plot) (Gramatica,2007). This method offers a graphical assessment of the leverage values ($h_{ii}$) as a function of the standardized cross validated residuals and it is suitable not only for detection of the structurally-influential outliers, but also for determination of the response outliers (Gramatica, 2007). The leverage is defined as a compound's distance from the centroid of *X*. Mathematically, the leverage ($h_{ii}$) of a given compound in the multidimensional descriptor space, can be calculated as (Eq. 13)

$$h_{ii} = x_i^T (X^T X)^{-1} x_i \quad \ldots\ldots\ldots\ldots\ldots\ldots (13)$$

Where $x_i$ the descriptor row matrix of the compound under consideration and X is is the multidimensional matrix carrying the structural information (calculated molecular descriptors) for each training set compound.The model predictions should be referred as unreliable for those compounds for which $h_{ii}$ diagonal elements are greater than the cut-off leverage value (h*). These compounds are located far from the structural centroid of the model, and therefore could be referred as structurally-influential outliers(Minovski*et al*., 2013). The cut-off leverage value (h*) is usually defined byEq. (14)

$h^* = \frac{3(P+1)}{n}$ ……(14)

Where p is the total number of descriptors used for developing of the QSAR model, while n is the total number of the training set compounds. Moreover, the compounds for which the calculated standardized residual values are greater than three standard deviation units (>±3σ) could be considered as response outliers (Minovski*et al*., 2013).

**3.2.8g Generation of the Williams plot.**

The original descriptor matrix was copied and pasted on an already opened excel worksheet, this was done by*clicking* on the right button on the key pad of the laptop to bring up a **drop down** menu from which the**paste special** menu was selected (see Fig 3.35).This brought up the dialogue box in Fig.3.36. **Transpose** option was *checked*and **ok** was selected to give the transpose of the original matrix. The transposed version ($X^T_{tr}$ = m x n) of the original descriptor matrix ($X_{tr)}$generated was multiplied by the original matrix Xtr by*highlighting* a section of the worksheet such that the number of columns on the highlighted section is equal to the number of columns of the original matrix and the number of rows equal to that of the number of rows of the transposed matrix. In the first cell of the highlighted section **MMULT (array1, array2)** was *typed.* Array 1 is the range of cells covered by the original matrix on the worksheet e.g. E2:G5, while array2 is the range of cells covered by the transpose.These ranges were selected by *highlighting* the sections covered by the respective matrices on the work sheet matrix as shown in Fig 3.37. The **CTRL+ SHIFT + ENTER** was pressed to give the result of the multiplication. The inverse of the result of the multiplication was found by first *highlighting* a section on the worksheet where the result of the operation will be displayed; the number of rows and columns on this section should be the same as that of the matrix whose inverse is being sought.

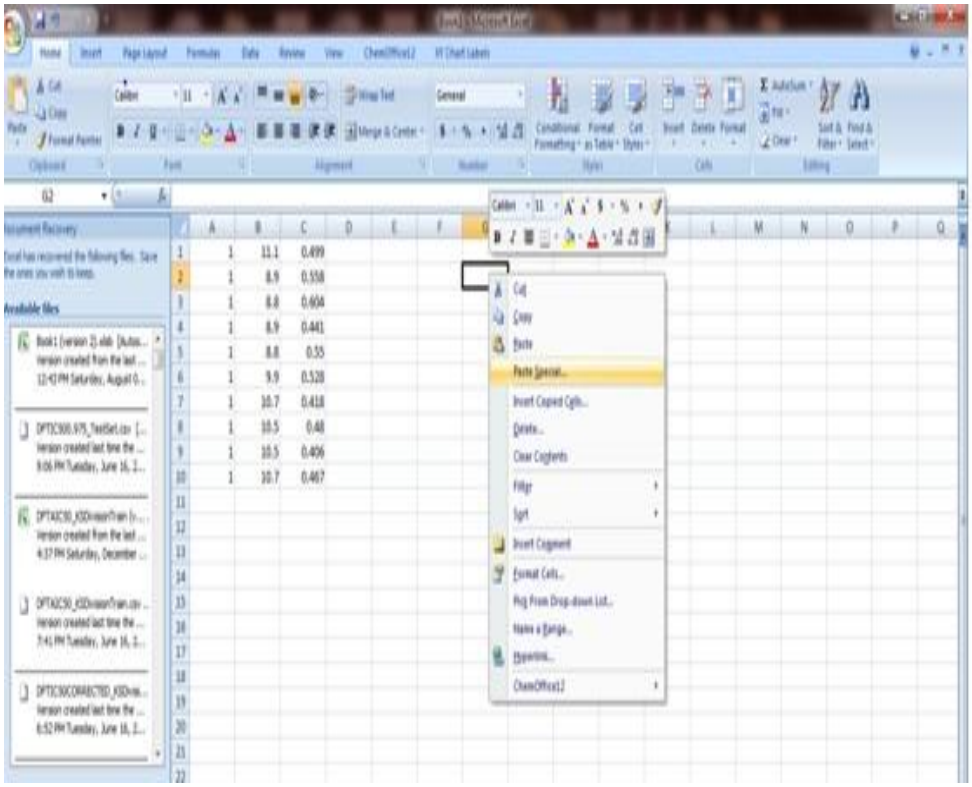


**Fig. 3.35 Excel Worksheet- A**

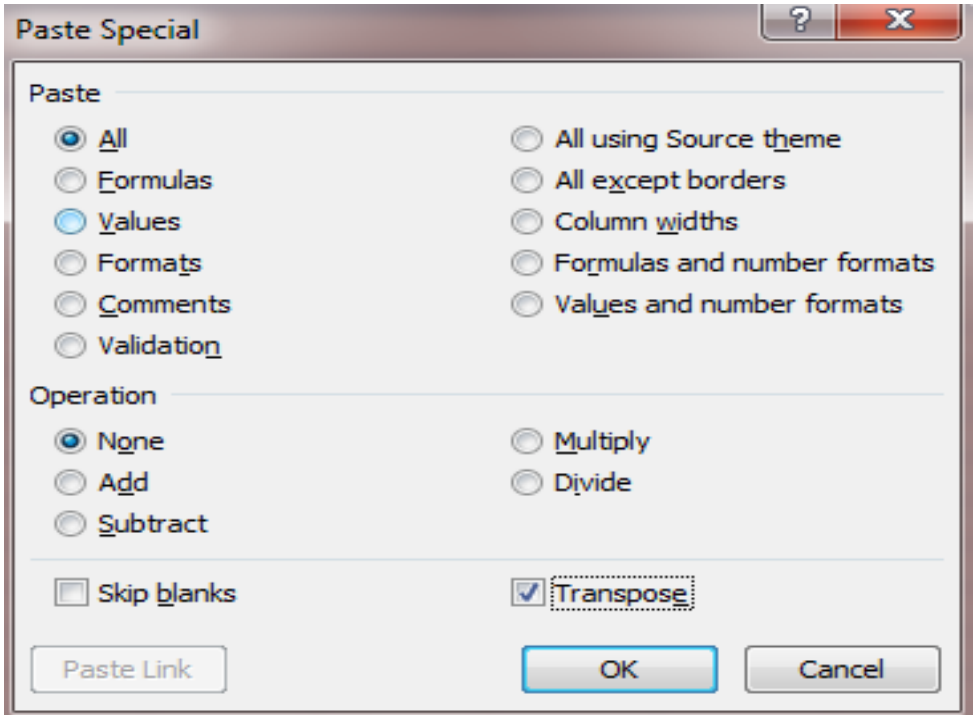



**Fig 3.36 Paste special Dialogue Box**

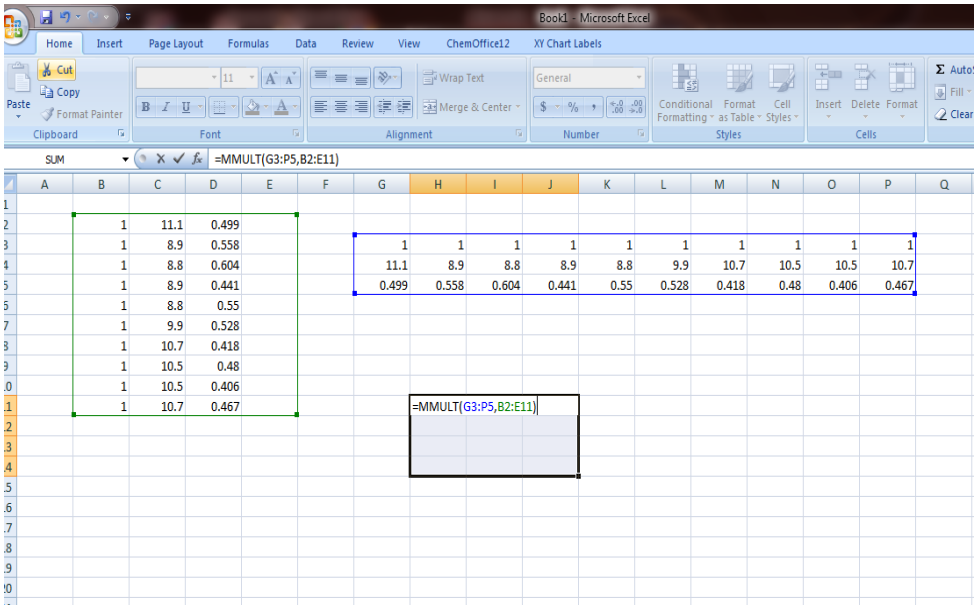



**Fig. 3.37: Excel work sheet-B**

In the first cell of the highlighted section (Fig 3.38) **=Transpose(Range)** was *typed*, 'range' is the range of cells covered by the matrix to be transposed e.g. A2:G5. **CTRL+ SHIFT + ENTER** was *pressed* to give the result of the operation. The original $X_{tr}$, the inverse $(X^{T}_{tr}\ X_{tr}\ )^{-1}$ and the transposed matrix $X^{T}$ were multiplied using the same multiplication steps described above to obtain an n × n symmetric matrix ( hat matrix) whose diagonal elements (leverages) describe the distance of each training set compound to the structural centroid of the model; the leverages were extracted from the hat matrix. The calculation of the hat matrix as well as the extraction of the diagonal elements (leverages) for the test/external validation set compound entering the model, was performed as for the training set described above with slight modification. A transposed version ($X^{T}_{te/ext}$ = m x n) of the original descriptor matrix of test data $X_{te/ext}$ was generated, a clone version of the inverse matrix solved above for the training data ;( $XTtrX_{tr)}^{-1}$ was also generated, the original $X_{te/ext}$, the inverse $(XTtrX_{tr)}^{-1}$ and the transposed matrix $XTte_{/ext\ were}$ multiplied. The leverages from both training and test sets hat matrices were plotted as a function of the standardized cross validated residuals to obtain the Williams plot.

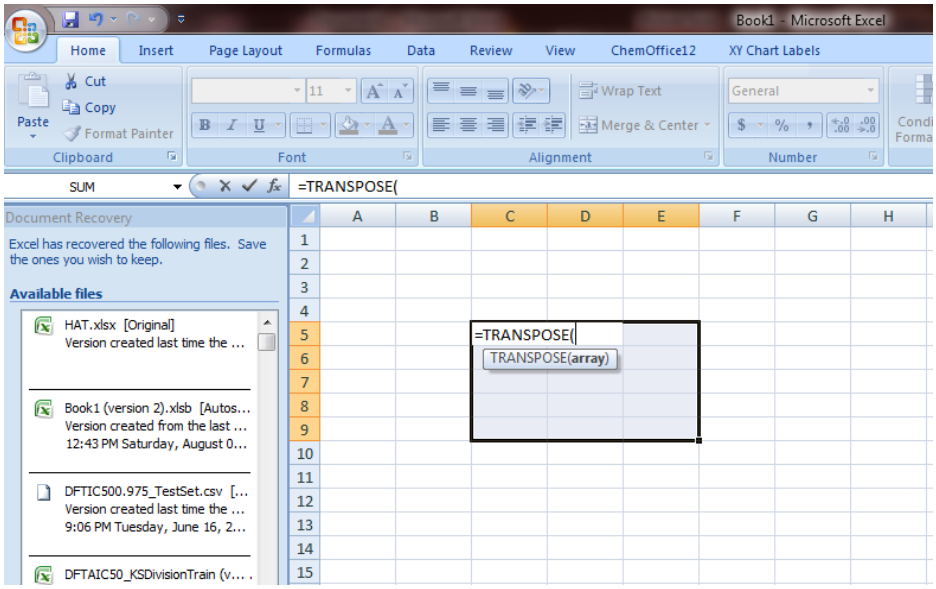



**Fig. 3.38: Excel Worksheet-C**

# CHAPTER FOUR

## 4.0 RESULTS

### 4.1 Result of Optimization and Calculation of Quantum-chemical Descriptors.

The geometry optimization of the molecules took averagely between 12-15 hours to complete. The optimization process ran smoothly except for few of the molecules where the optimization process ran out of cycles and the number of cycles had to be added using the **'Option'** section on the **Calculate** dialog (Fig.3.7). The descriptors evaluated by quantum- chemical calculations used in this study included; the highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO),total energy, total dipole moment, polarizability, ovality, enthalpy, entropy and Gibbs free energy. The values of these descriptors for the individual molecules were extracted after the optimization process was completed, by selecting **Properties** from the **Display** menu to bring up the **Molecule Properties** dialog. The following quantum descriptors were calculated using their various expressions; ionization energy (IP), electron affinity (EA), global hardness ($\eta$), chemical potential (μ), global electrophilicity index ($\omega$) and total softness

### 4.2 Result of the Calculation of 0D, 1D, 2D and 3D Descriptors Using Padel.

The final pool of the calculated 0D, 1D, 2D and 3D theoretical molecular descriptors using Padel Software, 1537 in total, can be separated into sixty-one classes: Acidic group count, ALOGP, APol , Aromatic atoms count, Aromatic bonds count, Atom count, Autocorrelation. Barysz matrix, Basic group count, BCUT, Bond count, BPol,Burden modified eigenvalues, Carbontypes, Chi chain, Chi cluster, Chi path cluster, Chi path, Constitutional, Crippen logP and MR, Detour matrix, Eccentric connectivity index, Atom type electrotopological state, Extended topochemical atom, FMFDescriptor, Fragment complexity, Hbond acceptor count, Hbond donor count, Hybridization

ratio, Information content, Kappa shape indices, Largest chain, Largest Pi system, Longest aliphatic chain, MannholdLogP, McGowan volume, Molecular distance edge, Molecular linear free energy relation, Path counts, Petitjean number, Ringcount, Rotatable bonds count, Rule of five, Topological, Topological charge,Topological distance matrix, Topological polar surface area, Van der Waals volume, Vertex adjacency information (magnitude),Walk counts, Weight, Weighted path, Wiener numbers, XLogP, Zagreb index, 3D autocorrelation, charged partial surface area, gravitational index, moment of inertia, Petitjean shape index, RDF.

### 4.3Result of the Variable Elimination and Selection

After the variable elimination, 307 descriptors survived,they fall under the following categories;ALOGP, APol, Aromatic atoms count,Chi chain, Chi cluster, Chi path cluster, Chi path, Constitutional, Crippen logP and MR, Detour matrix, Eccentric connectivity index, Atom type electrotopological state,3D autocorrelation, Charged partial surface area, Gravitational index, Moment of inertia, Petitjean shape index, RDF,Petitjean number,Barysz matrix, Basic group count, BCUT.

### 4.4Results of8-Methylquinolone Carboxylic against *Mycobacterium tuberculosis H37Rv* (MTB).

The models below represent the three best models generated by the GA-MLRA analysis carried using the BuildQSAR program. Initially several models were generated from which the best three were selected based on their high squared correlation coefficient $R^2$, low standard error $s$ and high Fisher coefficient $F$.

**MODEL 4.1**

pMIC = 5.19424(+/-0.04952) -0.37034(+/-0.05819) MATS4c -0.22646(+/-0.0629) VE3_Dzp +0.63546(+/-0.05634) SCH-3 -0.22182(+/-0.05461) PetitjeanNumber -0.15096(+/-0.04741) RDF95u

**MODEL 4.2**

pMIC = 5.19738(+/-0.0441) -0.24899(+/-0.05778) AATSC5p -0.43449(+/-0.05408) MATS4c -0.35084(+/-0.06424) VE3_Dzp +0.61537(+/-0.0508) SCH-3 -0.25681(+/-0.04934) PetitjeanNumber

**MODEL 4.3**

pMIC = 5.26325(+/-0.05438) -0.30789(+/-0.0572) MATS4c +0.69292(+/-0.06564) SCH-3 +0.26222(+/-0.06631) ZMIC3 -0.29654(+/-0.06421) RDF10u

Table 4.1 shows the statistical parameters as well as the internal and external validation parameters $Q^2_{LOO}$, $R^2_{Pred}$respectively of the three selected models.However, sincethe selection of a robust and well predictive QSAR model on the basis of only $R^2$, $Q^2$ and $R^2_{pred}$ might be misleading,the selected models were subjected to further statistical analysis as suggested by Golbraikh and Tropsha (2000). The result is presented in Table 4.2. Table 4.3 compares the predictive power of the three models.Model 4.2 was chosen as the best model since its $R^2_{pred}$ value is the highest out of the three models, for this model,Table 4.4 shows the experimental and calculated activities for the training sets after the Leave-One-out Internal validation.Fig 4.1 shows the correspondingplot of the experimental activities and the calculated activity of the training set. Table 4.5 shows the experimental and calculated activities of the external validation set and Fig. 4.2 captures the graphic correlation between them. The Y-randomization analysis results are shown in Table 4.6.The (Williams plot) is shown in Fig. 4.3. Table 4.7 shows the correlation matrix of the descriptors in model 4.2.

**Table 4.1 Fitting Parameters for the three selected Models.**

| Model | $R^2$ | $R^2_{adjs}$ | F | s | $Q^2_{LOO}$ | $R^2_{Pred}$ | SPress | SDEP | n | k |
|---|---|---|---|---|---|---|---|---|---|---|
| 4.4.1 | 0.87784 | 0.85008 | 31.61971 | 0.25853 | 0.81987 | 0.78377 | 2.16828 | 0.27828 | 28 | 5 |
| 4.4.2 | 0.90323 | 0.88123 | 41.0676 | 0.2301 | 0.83115 | 0.78708 | 2.03242 | 0.26942 | 28 | 5 |
| 4.4.3 | 0.84924 | 0.82302 | 32.38971 | 0.28089 | 0.77874 | 0.55815 | 2.6633 | 0.30841 | 28 | 4 |

**Table 4.2 Golbraikh and Tropsha acceptable model criteria for the three Models**

| MODEL | $Q^2$ | $R^2_{Pred}$ | $\lvert R_0^2 - R'_0{}^2 \rvert$ | $[(R^2 - R_0^2)/ R^2]$ | $[(R^2 - R_0^2)/ R^2]$ | k | k' |
|---|---|---|---|---|---|---|---|
| 4.1 | 0.81987 | 0.78377 | 0.07916 | 0.00117 | 0.10217 | 0.98827 | 1.00963 |
| 4.2 | 0.83115 | 0.78708 | 0.00866 | 0.01103 | 0.02204 | 0.97919 | 1.01894 |
| 4.3 | 0.77874 | 0.55815 | 0.04133 | 0.15416 | 0.08011 | 0.95061 | 1.04631 |

**Table4.3 Comparison of Observed Activity with Predicted Activityof the three models**

| COMP NO. | OBSERVED ACTIVITY | PREDICTED ACTIVITY | | |
|---|---|---|---|---|
| | | MODEL 4.1 | MODEL 4.2 | MODEL 4.3 |
| 2 | 4.614 | 4.429 | 4.456 | 4.897 |
| 3 | 5.539 | 5.478 | 5.487 | 5.698 |
| 4 | 4.590 | 4.529 | 4.555 | 5.064 |
| 8 | 4.631 | 4.675 | 4.669 | 4.503 |
| 9 | 5.824 | 5.877 | 5.867 | 5.305 |
| 10 | 4.575 | 4.832 | 4.783 | 4.928 |
| 11 | 5.225 | 5.070 | 5.088 | 4.998 |
| 12 | 6.377 | 5.897 | 6.004 | 5.894 |
| 14 | 4.693 | 4.079 | 4.495 | 4.301 |
| 15 | 4.598 | 4.970 | 4.931 | 5.126 |
| 16 | 4.946 | 4.973 | 4.970 | 5.001 |
| 17 | 6.409 | 6.040 | 6.093 | 6.010 |
| 18 | 5.697 | 5.887 | 5.859 | 5.905 |
| 19 | 5.167 | 4.903 | 4.933 | 5.048 |
| 20 | 4.810 | 4.595 | 4.624 | 5.026 |
| 21 | 6.167 | 6.338 | 6.296 | 6.154 |
| 22 | 6.076 | 6.119 | 6.113 | 6.346 |
| 23 | 4.393 | 4.345 | 4.355 | 4.341 |
| 24 | 4.587 | 5.102 | 4.983 | 5.263 |
| 25 | 4.665 | 4.955 | 4.929 | 4.720 |
| 26 | 4.621 | 5.014 | 4.970 | 4.292 |
| 27 | 6.114 | 6.173 | 6.164 | 6.140 |
| 28 | 6.076 | 5.928 | 5.976 | 5.902 |
| 30 | 5.699 | 5.423 | 5.506 | 5.325 |
| 31 | 4.868 | 4.973 | 4.908 | 4.711 |
| 32 | 4.511 | 4.491 | 4.496 | 4.494 |
| 35 | 5.162 | 5.460 | 5.413 | 5.383 |
| 36 | 4.576 | 4.225 | 4.286 | 4.352 |

**Table 4.4Observed Activity versus Predicted Activity of Model 4.2**

| COMP. NO | pMIC OBSERVED | pMIC CALCULATED | RESIDUAL |
|---|---|---|---|
| 2 | 4.614 | 4.456 | 0.158 |
| 3 | 5.539 | 5.487 | 0.052 |
| 4 | 4.590 | 4.555 | 0.035 |
| 8 | 4.631 | 4.669 | -0.038 |
| 9 | 5.824 | 5.867 | -0.043 |
| 10 | 4.575 | 4.783 | -0.208 |
| 11 | 5.225 | 5.088 | 0.136 |
| 12 | 6.377 | 6.004 | 0.373 |
| 14 | 4.693 | 4.495 | 0.197 |
| 15 | 4.598 | 4.931 | -0.333 |
| 16 | 4.946 | 4.970 | -0.024 |
| 17 | 6.409 | 6.093 | 0.316 |
| 18 | 5.697 | 5.859 | -0.162 |
| 19 | 5.167 | 4.933 | 0.234 |
| 20 | 4.810 | 4.624 | 0.186 |
| 21 | 6.167 | 6.296 | -0.129 |
| 22 | 6.076 | 6.113 | -0.037 |
| 23 | 4.393 | 4.355 | 0.039 |
| 24 | 4.587 | 4.983 | -0.397 |
| 25 | 4.665 | 4.929 | -0.264 |
| 26 | 4.621 | 4.970 | -0.348 |
| 27 | 6.114 | 6.164 | -0.050 |
| 28 | 6.076 | 5.976 | 0.100 |
| 30 | 5.699 | 5.506 | 0.193 |
| 31 | 4.868 | 4.908 | -0.040 |
| 32 | 4.511 | 4.496 | 0.015 |
| 35 | 5.162 | 5.413 | -0.251 |
| 36 | 4.576 | 4.286 | 0.290 |

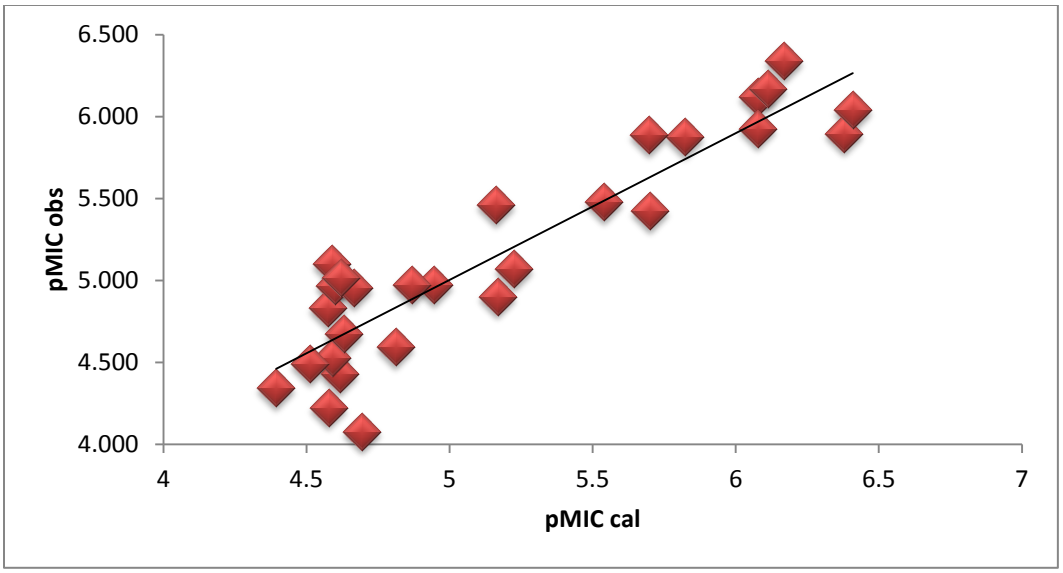



**Fig 4.1: Plot of observed pMIC against predicted pMIC of model4.2**

**Table 4.5Observed Activity versus Predicted Activity(Test Sets) for model 4.2**

| Name | $Y_{obs}$ | $Y_{pres}$ |
|---|---|---|
| 1 | 5.201349 | 5.596334 |
| 6 | 4.852942 | 5.239084 |
| 7 | 5.739929 | 5.378934 |
| 13 | 5.528708 | 5.523666 |
| 29 | 5.534617 | 5.700163 |
| 33 | 5.179799 | 4.907993 |
| 34 | 6.013228 | 6.307102 |
| 38 | 5.534617 | 5.700163 |
| 37 | 4.072322 | 4.271199 |

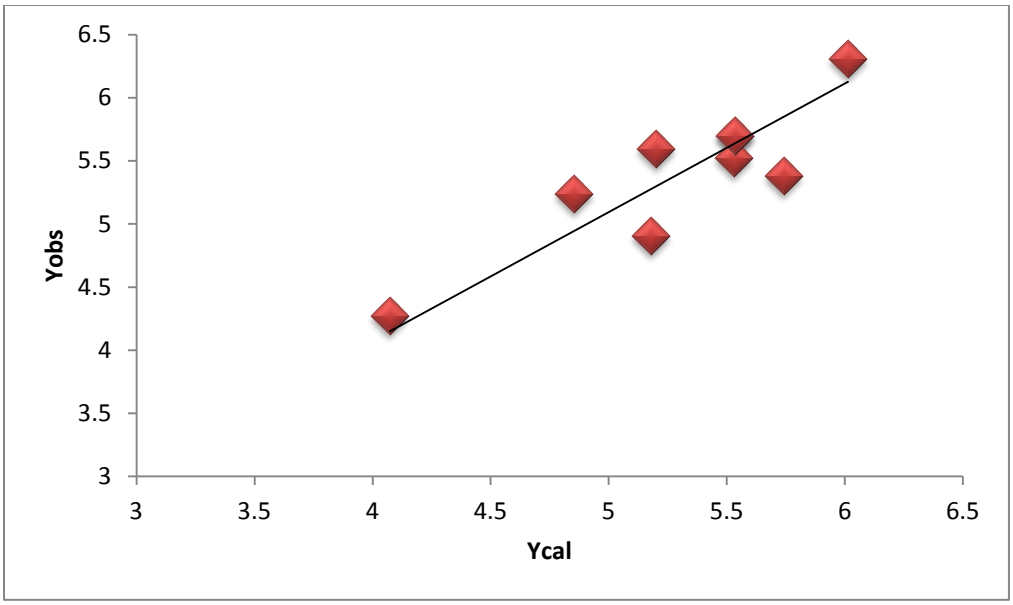



**Fig. 4.2: Plot of observed pMIC against predicted pMIC of model(Test Set) model 4.2**

**Table 4.6Y-Randomization Result for model 4.2**

| Model | R | $R^2$ | $Q^2$ |
|---|---|---|---|
| Original | 0.950383 | 0.903228 | 0.831152 |
| Random 1 | 0.413729 | 0.171172 | -0.42403 |
| Random 2 | 0.533417 | 0.284534 | -0.35129 |
| Random 3 | 0.476735 | 0.227277 | -0.31963 |
| Random 4 | 0.460269 | 0.211848 | -0.35088 |
| Random 5 | 0.511072 | 0.261195 | -0.67450 |
| Random 6 | 0.322599 | 0.104070 | -1.03309 |
| Random 7 | 0.243157 | 0.059125 | -1.24968 |
| Random 8 | 0.344399 | 0.118611 | -0.54445 |
| Random 9 | 0.642345 | 0.412607 | -0.0884 |
| Random 10 | 0.297694 | 0.088622 | -0.94256 |

The average R is0.4245; the average $R^2$ is 0.1939; the average $Q^2$ is -0.59785; the $R^2_p$ calculated according to equation(13) is 0.808101.

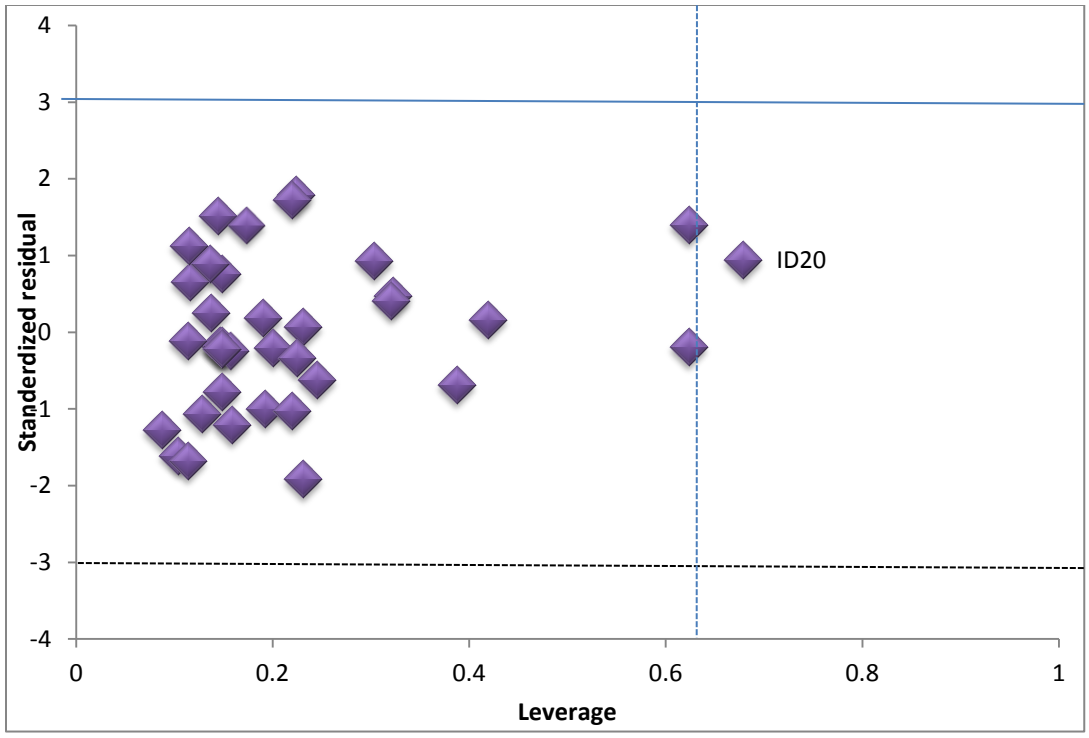



**Fig.4.3: The Williams Plot for Model 4.2**

**Table4.7Correlation matrix among the descriptors in models 4.2**

| | *pMIC* | *AATSC5p* | *MATS4c* | *VE3_Dzp* | *SCH-3* | *PetitjeanNumber* |
|---|---|---|---|---|---|---|
| pMIC | 1 | | | | | |
| AATSC5p | -0.37458 | 1 | | | | |
| MATS4c | -0.09013 | -0.15454 | 1 | | | |
| VE3_Dzp | 0.04369 | -0.42183 | -0.35456 | 1 | | |
| SCH-3 | 0.75332 | -0.38069 | 0.32207 | 0.10729 | 1 | |
| PetitjeanNumber | -0.09199 | -0.08622 | -0.17130 | -0.06721 | 0.10686 | 1 |

## 4.5 Results of 8 -Methylqinolone carboxylic acid against *Mycobacterium Smegmatis ATCC 14468 (MC2)*

According to result of GA-MLR using BuildQSAR software ,initially several models were generated, from which the three best were selected based on their high squared correlation coefficient $R^2$, low standard error *s* and high Fisher coefficient *F* .

**MODEL 4.4**

pMIC = 4.77374(+/-0.03903) -0.18609(+/-0.04924) AATS4i +0.50382(+/-0.05235) SCH-3 -0.44712(+/-0.06573) AVP-1 -0.22376(+/-0.05623) maxHCsats -0.18403(+/-0.04374) PSA

**MODEL 4.5**

pMIC = 4.81625(+/-0.04215) +0.64834(+/-0.0532) SCH-3 +0.21394(+/-0.05229) MOMI-Z -0.40533 (+/-0.06216) RDF95u +0.29364(+/-0.05782) RDF85m

**MODEL 4.6**

pMIC = 4.7754(+/-0.053) +0.60548(+/-0.05689) SCH-3 -0.34713(+/-0.07536) AVP-1 -0.21081(+/-0.05526) nHBAcc

Table 4.8 shows the statistical parameters as well as the internal and external validation parameters $Q^2_{LOO}$, $R^2_{Pred}$respectively of the three selected models.

From the 3 QSAR models, model 4.4 was adjudged the best, based on:

1. It has the highest value of $R^2_{adj}$ (0.899)
2. It has the smallest value of SE (Standard Error of Estimation) of 0.1890.
3. If the value of F exceed value of $F_{table}$ or comparison of $F_{calc}/F_{table}$ more than 1. All the models have value of $F_{calc}/F_{table}$ more than 1, but model 4.5.1 has the biggest value (47.28523).
4. It has the smallest value of PRESS (Predictive Residual Sum of Square) of 1.38083.

**Table 4.8 Fitting Parameters for models 4.4, 4.5 and 4.6**

| Model | $R^2$ | $R^2_{adjs}$ | F | s | $Q^2_{LOO}$ | $R^2_{Pred}$ | SPress | SDEP | n | k |
|---|---|---|---|---|---|---|---|---|---|---|
| 4.4 | 0.91842 | 0.8990 | 47.28523 | 0.18902 | 0.84987 | 0.79748 | 1.38083 | 0.22615 | 27 | 5 |
| 4.5 | 0.88902 | 0.86884 | 44.05681 | 0.21541 | 0.83554 | 0.74605 | 1.51266 | 0.23669 | 27 | 4 |
| 4.6 | 0.83358 | 0.81188 | 38.40288 | 0.25797 | 0.75867 | 0.68725 | 2.21975 | 0.28673 | 27 | 3 |

However since the selection of a robust and well predictive QSAR model on the basis of only $R^2$, $Q^2$ and $R^2_{pred}$ might be misleading, the selected models were subjected to further statistical analysis as suggested by Golbraikh and Tropsha (2000) and the result is presented in Table 4.9. Table 4.10 compares the predictive power of the three models. For model 4.4, Table 4.11 shows the experimental and calculated activities of the training after the Leave-One-out internal validation set while Fig 4.4 shows the corresponding plot of the experimental activities with the calculated ones. Table 4.12 shows the experimental and calculated activities of the external validation set and Fig. 4.5 captures the graphic correlation between them. The Y-randomization analysis results are shown in Table 4.13 and the applicability domain which is the scatter plot defined by standardized residuals as a function of the leverage (Williams plot) is shown in Fig. 4.6. Table 4.14 shows the correlation matrix of the descriptors in model 4.4. Table 4.15 gives the names of the various descriptors in the three models.

### 4.6 Results of a mixture of 8-Methylquinolone Carboxylic acids and Nitroquinolones against *Mycobacterium tuberculosis H37Rv* (MTB)

According to result of GA-MLR using BuildQSAR software, initially several models were generated, from which the three best were selected base on their high squared correlation coefficient $R^2$, low standard error *s* and high Fisher coefficient *F*.

**MODEL 4.7**

pMIC = 5.57745(+/-0.03979) -0.48195(+/-0.0702) ATSC4p -0.32928(+/-0.04833) GATS5i -0.18795(+/-0.04622) maxHBint6 -0.51413(+/-0.08589) MDEC-23 +0.7931(+/-0.07134) RDF65u -0.77285(+/-0.06301) RDF90u

**Table 4.9 Golbraikh and Tropsha acceptable model criteria Models 4.5, 4.5 and 4.5**

| MODEL | $Q^2$ | $R^2_{pred}$ | $\lvert R_0^2 - R'^2_0 \rvert$ | $[(R^2 - R_0^2)/ R^2]$ | $[(R^2 - R_0^2)/ R^2]$ | k | k’ |
|---|---|---|---|---|---|---|---|
| 4.4 | 0.84987 | 0.79343 | 0.06192 | 0.07814 | 0.00049 | 1.03178 | 0.96500 |
| 4.5 | 0.83554 | 0.74605 | 0.09886 | 0.13293 | 0.00042 | 0.99028 | 1.00398 |
| 4.6 | 0.75867 | 0.68725 | 0.16811 | 0.24557 | 0.00096 | 1.01783 | 0.97474 |

**Table 4.10 Comparison of Observed Activity with Predicted Activity between Models 4.4, 4.5 and 4.6**

| COMP NO. | OBSERVED ACTIVITY | Predicted Activity | | |
|---|---|---|---|---|
| | | MODEL 4.5.1 | MODEL 4.5.2 | MODEL 4.5.3 |
| 2 | 4.360 | 4.060 | 4.169 | 4.101 |
| 3 | 4.635 | 4.781 | 4.905 | 4.722 |
| 4 | 4.289 | 4.728 | 4.500 | 4.847 |
| 5 | 4.603 | 4.618 | 4.471 | 4.492 |
| 7 | 5.440 | 5.503 | 5.328 | 5.292 |
| 9 | 5.219 | 5.052 | 4.899 | 5.244 |
| 10 | 3.973 | 4.106 | 4.314 | 4.091 |
| 11 | 4.323 | 4.540 | 4.285 | 4.502 |
| 12 | 5.764 | 5.527 | 5.861 | 5.371 |
| 13 | 5.227 | 5.342 | 5.023 | 5.213 |
| 14 | 4.391 | 4.533 | 4.556 | 4.414 |
| 15 | 4.297 | 4.168 | 4.182 | 4.195 |
| 16 | 4.344 | 4.334 | 4.207 | 4.330 |
| 17 | 6.092 | 5.960 | 5.721 | 5.803 |
| 18 | 5.395 | 5.492 | 5.551 | 5.571 |
| 19 | 4.565 | 4.400 | 4.550 | 4.233 |
| 20 | 4.208 | 4.128 | 4.216 | 4.274 |
| 22 | 5.777 | 5.639 | 5.701 | 5.350 |
| 23 | 4.634 | 4.670 | 4.614 | 4.658 |
| 25 | 4.364 | 4.480 | 4.231 | 4.722 |
| 26 | 4.320 | 4.262 | 4.353 | 4.330 |
| 27 | 5.511 | 5.605 | 5.588 | 5.600 |
| 28 | 5.472 | 5.570 | 5.215 | 5.493 |
| 31 | 4.266 | 4.453 | 4.481 | 4.255 |
| 35 | 5.162 | 4.903 | 5.253 | 4.933 |
| 38 | 4.932 | 4.816 | 5.371 | 5.475 |
| 37 | 4.330 | 4.226 | 4.350 | 4.382 |

**Table 4.11 Observed Activity versus Predicted Activity of Model 4.4**

| COMP. NO | pMIC OBSERVED | pMIC CALCULATED | RESIDUAL |
|---|---|---|---|
| 2 | 4.360 | 4.206 | 0.154 |
| 3 | 4.635 | 4.804 | -0.169 |
| 4 | 4.289 | 4.504 | -0.215 |
| 5 | 4.603 | 4.422 | 0.181 |
| 7 | 5.440 | 5.372 | 0.068 |
| 9 | 5.219 | 5.279 | -0.060 |
| 10 | 3.973 | 4.110 | -0.137 |
| 11 | 4.323 | 4.359 | -0.037 |
| 12 | 5.764 | 5.478 | 0.286 |
| 13 | 5.227 | 5.299 | -0.072 |
| 14 | 4.391 | 4.423 | -0.032 |
| 15 | 4.297 | 4.152 | 0.145 |
| 16 | 4.344 | 4.454 | -0.110 |
| 17 | 6.092 | 5.900 | 0.192 |
| 18 | 5.395 | 5.670 | -0.275 |
| 19 | 4.565 | 4.466 | 0.099 |
| 20 | 4.208 | 4.279 | -0.072 |
| 22 | 5.777 | 5.661 | 0.117 |
| 23 | 4.634 | 4.614 | 0.020 |
| 25 | 4.364 | 4.505 | -0.141 |
| 26 | 4.320 | 4.380 | -0.059 |
| 27 | 5.511 | 5.719 | -0.208 |
| 28 | 5.472 | 5.585 | -0.112 |
| 31 | 4.266 | 4.261 | 0.005 |
| 35 | 5.162 | 4.780 | 0.382 |
| 38 | 4.932 | 4.866 | 0.067 |
| 37 | 4.330 | 4.347 | -0.017 |

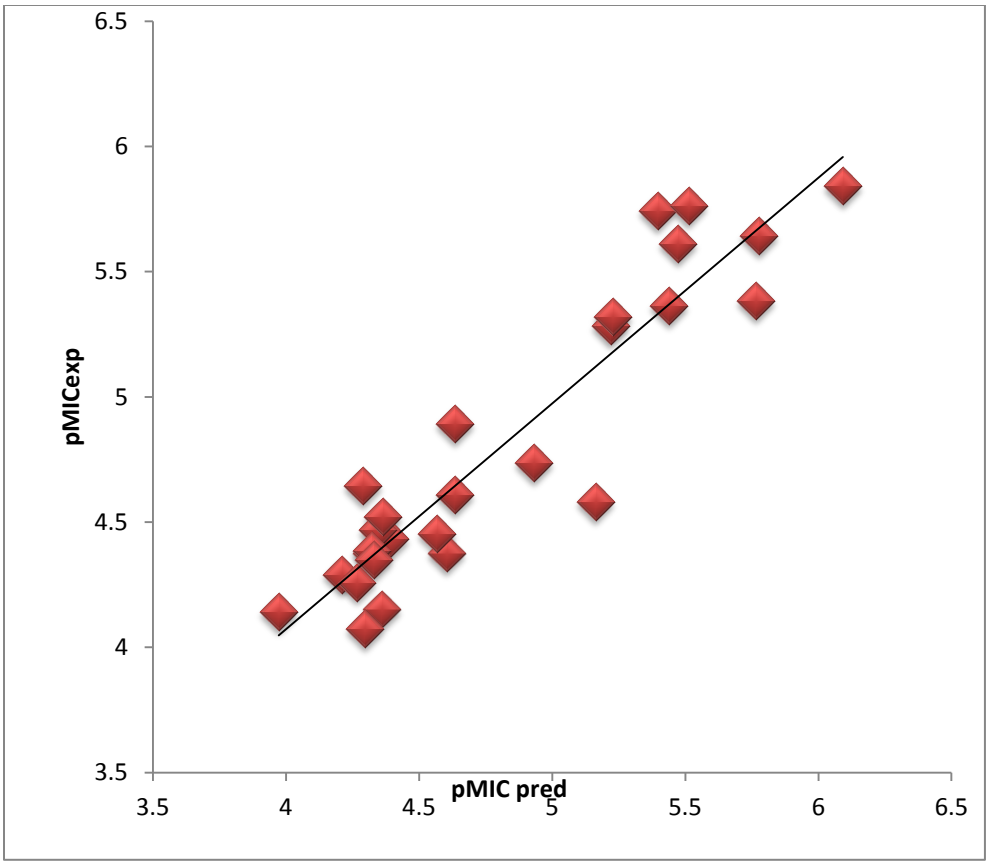



**Fig. 4.4Plot of observed pMIC against predicted pMIC of model for model 4.4**

**Table 4.12 Observed Activity versus Predicted Activity for Test Sets**

| Name | $Y_{obs}$ | $Y_{pred}$ |
|---|---|---|
| 1 | 4.90136 | 5.41943 |
| 6 | 3.95001 | 3.97352 |
| 30 | 5.39686 | 5.01111 |
| 32 | 4.21000 | 3.68403 |
| 33 | 4.57840 | 4.35840 |
| 34 | 5.71220 | 5.38170 |
| 36 | 3.97388 | 3.68265 |

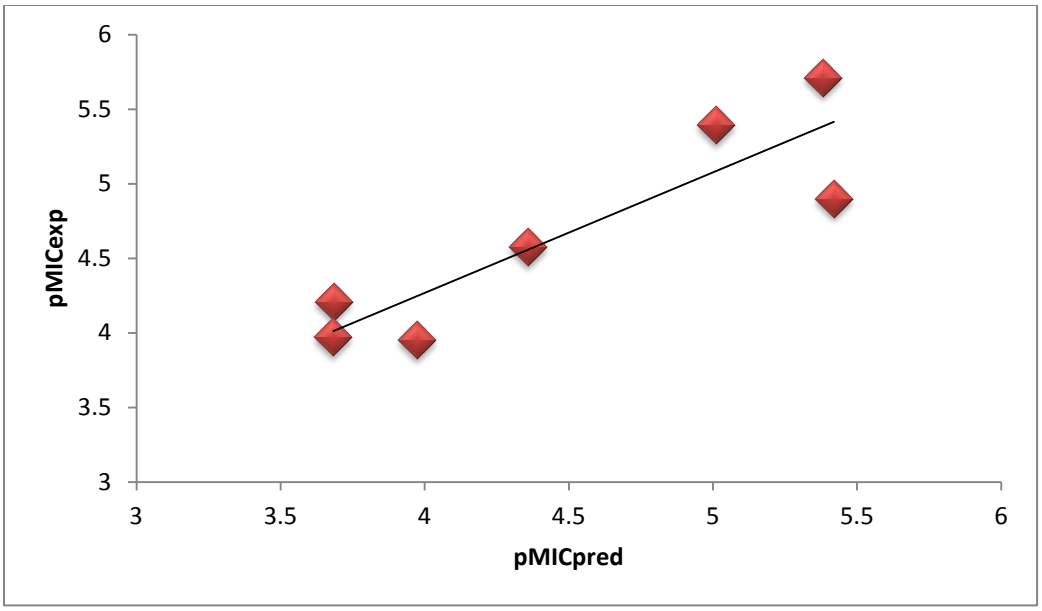



**Fig 4.5 Plot of observed pMIC against predicted pMIC (Test Sets) for model 4.4**

**Table 4.13 Y-Randomization Result for model 4.4**

| Model | R | $R^2$ | $Q^2$ |
|---|---|---|---|
| Original | 0.958344 | 0.918423 | 0.849874 |
| Random 1 | 0.391732 | 0.153454 | -0.461720 |
| Random 2 | 0.419436 | 0.175926 | -0.296950 |
| Random 3 | 0.366119 | 0.134043 | -0.472730 |
| Random 4 | 0.341172 | 0.116399 | -0.991030 |
| Random 5 | 0.200082 | 0.040033 | -0.499230 |
| Random 6 | 0.249337 | 0.062169 | -0.574450 |
| Random 7 | 0.487747 | 0.237897 | -0.225640 |
| Random 8 | 0.694441 | 0.482249 | 0.164201 |
| Random 9 | 0.461186 | 0.212693 | -0.158400 |
| Random 10 | 0.491820 | 0.241887 | -0.516370 |

Random Models Parameters
Average r 0.410307: Average $r^2$: 0.185675 Average $Q^2$: -0.40323 $R^2p$: 0.82999

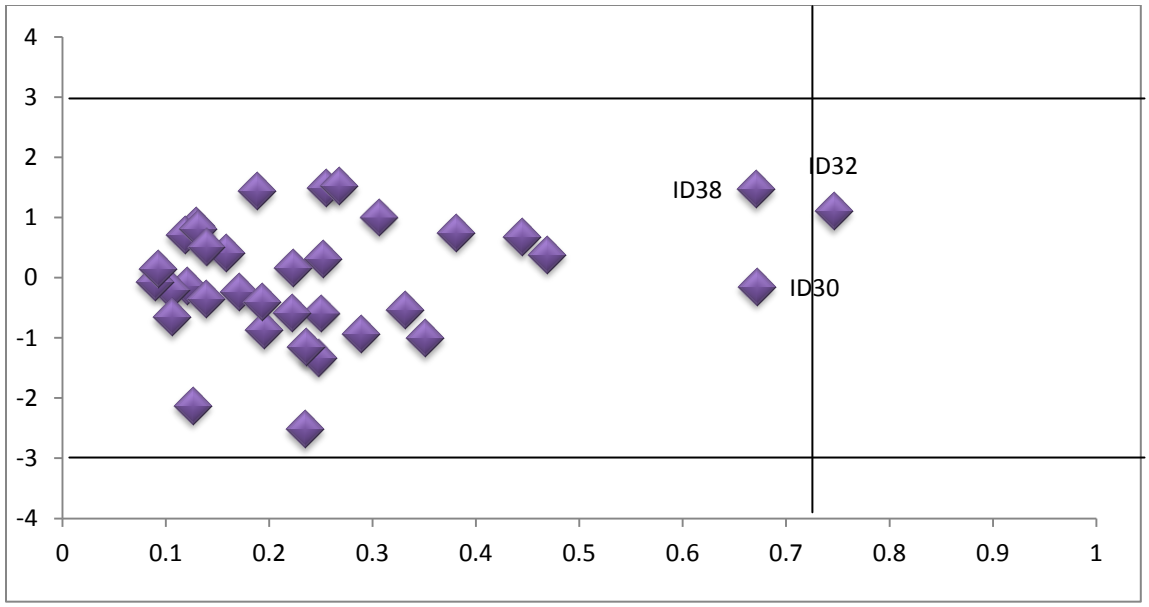



**Fig. 4.6 Williams plot for model 4.4**

**Table4.14Correlation matrix among the descriptors**

| | *MIC* | *AATS4i* | *SCH-3* | *AVP-1* | *maxHCsats* | *PSA* |
|---|---|---|---|---|---|---|
| pMIC | | | | | | |
| AATS4i | .58122 | 1 | | | | |
| SCH-3 | 72544 | -0.68865 | 1 | | | |
| AVP-1 | .0748 | -0.03294 | 0.163049 | 1 | | |
| maxHCsats | 069948 | -0.22347 | 0.128462 | -0.61026 | 1 | |
| PSA | .05221 | 0.026457 | -0.00254 | -0.61682 | 0.469197 | 1 |

**Table 4.15Name of the various descriptors in the three models.**

| Descriptor Symbol | Descriptor Name |
|---|---|
| **SCH-3** | Simple chain, order 3 |
| **AVP-1** | Valence path, order 1 |
| **maxHCsats** , | Maximum atom-type H E-State: H bonded to B, Si, P, Ge, As, Se, Sn or Pb |
| **MOMI-Z** | Moment of inertia along Z axis |
| **RDF95u** | Radial distribution function - 095 / unweighted |
| **RDF85m** | Radial distribution function - 085 / weighted by relative mass |
| **AATS4i** | Average Broto-Moreau autocorrelation - lag 4 / weighted by first ionization potential |
| **PSA** | polar surface area |
| **nHBAcc** | Number of hydrogen bond acceptors. |

**MODEL 4.8**

pMIC = 5.52134(+/-0.04091) +0.28728(+/-0.05095) ATSC6c -0.42489(+/-0.04799) ATSC4p -0.17324(+/-0.04737) MLFER_S -0.24485(+/-0.05745) TDB10i +0.44196(+/-0.0665) RDF65u -0.67208(+/-0.05575) RDF90u

**MODEL 4.9**

pMIC = 5.55875(+/-0.04419) -0.19178(+/-0.04036) ATSC4i -0.2106(+/-0.04527) AATSC7e -0.22168(+/-0.05241) AATSC7i -0.24813(+/-0.05122) nHBint3 +0.65739(+/-0.0667) RDF65u -0.77008(+/-0.06636) RDF90u

Model 4.7 was adjudged the best model based on:

1. It has the highest value of $R^2_{adj}$ (0.88073)

2. It has the smallest value of SE (Standard Error of Estimation) of 0.22292.

3. If the value of F exceed value of $F_{table}$ or comparison of $F_{calc}/F_{table}$ more than 1. All the models have value of $F_{calc}/F_{table}$ more than 1, but model 4.6.1 the biggest value of 42.84356

4. It has the smallest value of PRESS (Predictive Residual Sum of Square) of 2.46008.

However sincethe selection of a robust and well predictive QSAR model on the basis of only $R^2$, $Q^2$ and $R^2_{pred}$ might be misleading, the selected models were subjected to further statistical analysis as suggested by Golbraikh and Tropsha (2000).Table 4.16 shows the statistical parameters as well as the internal and external validation parameters $Q^2_{LOO}$, $R^2_{Pred}$respectively of the three selected models. Table 4.17 shows the result of the Golbraikh and Tropsha statistics.Table 4.18 compares the predictive power of the three models. Table 4.19 shows the experimental and calculated activities of the training set after the LOO internal validation, while Fig 4.7 shows the corresponding plot of the experimental activities with the calculated ones. Table 4.20 shows the experimental and calculated activities of the test set and Fig. 4.8 captures the graphic correlation between them. The Y-randomization analysis results are shown in Table 4.21. Table 4.22 shows the correlation matrix; Table 4.23 gives the names of the descriptors.

**Table 4.16 Fitting Parameters For models 4.7,4.8 and 4.9**

| Model | $R^2$ | $R^2_{adjs}$ | F | s | $Q^2_{LOO}$ | $R^2_{pred}$ | SPress | SDEP | n | k |
|---|---|---|---|---|---|---|---|---|---|---|
| 4.7 | 0.90178 | 0.88073 | 42.84356 | 0.22292 | 0.82634 | 0.77462 | 2.46008 | 0.26512 | 35 | 6 |
| 4.8 | 0.89924 | 0.87765 | 41.6503 | 0.22578 | 0.84034 | 0.69179 | 2.26171 | 0.25421 | 35 | 6 |
| 4.9 | 0.89577 | 0.87344 | 40.10782 | 0.22963 | 0.816 | 0.75632 | 2.60661 | 0.2729 | 35 | 6 |

**Table 4.17Golbraikh and Tropsha acceptable model criteria For Models 4.7 4.8 and 4.9**

| MODEL | $Q^2$ | $R^2_{pred}$ | $\|R_0^2 - R'_0{}^2\|$ | $[(R^2 - R_0^2)/ R^2]$ | $[(R^2 - R_0^2)/ R^2]$ | k | k' |
|---|---|---|---|---|---|---|---|
| 4.7 | 0.82634 | 0.75632 | 0.07293 | 0.09648 | 0.00005 | 0.9679 | 1.03025 |
| 4.8 | 0.84034 | 0.69179 | 0.20735 | 0.00488 | 0.30462 | 0.97024 | 1.02781 |
| 4.9 | 0.8160 | 0.77462 | 0.02903 | 0.0423 | 0.00483 | 0.95514 | 1.04456 |

,

**Table 4.18 Comparison of Observed Activity with Predicted Activity between Models 4.7, 4.8 and 4.9**

| COMP NO. | OBSERVED ACTIVITY | PREDICTED ACTIVITY | | |
|---|---|---|---|---|
| | | MODEL 4.7 | MODEL 4.8 | MODEL 4.9 |
| 3 | 5.539 | 5.755 | 5.606 | 5.504 |
| 4 | 4.590 | 4.832 | 4.745 | 4.988 |
| 8 | 4.631 | 5.192 | 4.662 | 4.895 |
| 9 | 5.824 | 5.413 | 5.393 | 5.266 |
| 14 | 4.693 | 4.750 | 4.845 | 5.032 |
| 15 | 4.598 | 4.800 | 4.742 | 4.665 |
| 16 | 4.946 | 4.961 | 4.987 | 5.082 |
| 17 | 6.409 | 6.253 | 6.197 | 6.120 |
| 18 | 5.697 | 5.639 | 5.445 | 5.679 |
| 19 | 5.167 | 5.286 | 5.115 | 5.182 |
| 20 | 4.810 | 4.773 | 5.323 | 4.802 |
| 21 | 6.167 | 5.908 | 6.305 | 6.335 |
| 22 | 6.076 | 6.007 | 5.794 | 5.815 |
| 23 | 4.393 | 4.290 | 3.998 | 3.876 |
| 25 | 4.665 | 5.145 | 5.124 | 4.994 |
| 26 | 4.621 | 4.486 | 4.646 | 4.645 |
| 27 | 6.114 | 6.266 | 6.255 | 6.301 |
| 28 | 6.076 | 5.788 | 5.777 | 5.811 |
| 30 | 5.699 | 5.901 | 5.896 | 5.829 |
| 35 | 5.162 | 5.005 | 5.165 | 4.999 |
| 36 | 4.576 | 4.201 | 4.359 | 4.193 |
| 49 | 6.004 | 6.136 | 5.957 | 6.229 |
| 56 | 6.046 | 6.252 | 6.057 | 6.236 |
| 58 | 5.451 | 5.161 | 5.719 | 5.790 |
| 57 | 5.177 | 5.075 | 5.029 | 5.089 |
| 52 | 5.507 | 5.379 | 5.768 | 5.483 |
| 45 | 5.157 | 5.626 | 5.747 | 5.518 |
| 43 | 6.076 | 6.227 | 5.871 | 6.048 |
| 46 | 5.582 | 5.617 | 5.322 | 5.593 |
| 39 | 5.479 | 5.696 | 5.182 | 5.575 |
| 44 | 6.114 | 5.994 | 5.820 | 5.964 |
| 59 | 7.000 | 6.430 | 6.902 | 6.319 |
| 51 | 5.539 | 5.024 | 5.504 | 5.208 |
| 47 | 5.007 | 5.271 | 5.377 | 5.355 |
| 45 | 4.922 | 4.982 | 4.769 | 5.027 |

**Table 4.19 Observed Activity versus Predicted Activity of Model 4.7**

| COMP. NO | pMIC OBSERVED | pMIC CALCULATED | RESIDUAL |
|---|---|---|---|
| 3 | 5.539 | 5.595 | -0.056 |
| 4 | 4.590 | 4.726 | -0.136 |
| 8 | 4.631 | 4.649 | -0.018 |
| 9 | 5.824 | 5.566 | 0.258 |
| 14 | 4.693 | 4.767 | -0.074 |
| 15 | 4.598 | 4.726 | -0.128 |
| 16 | 4.946 | 4.984 | -0.038 |
| 17 | 6.409 | 6.232 | 0.177 |
| 18 | 5.697 | 5.529 | 0.168 |
| 19 | 5.167 | 5.119 | 0.048 |
| 20 | 4.810 | 5.202 | -0.393 |
| 21 | 6.167 | 6.268 | -0.101 |
| 22 | 6.076 | 5.826 | 0.249 |
| 23 | 4.393 | 4.122 | 0.271 |
| 25 | 4.665 | 5.035 | -0.370 |
| 26 | 4.621 | 4.644 | -0.022 |
| 27 | 6.114 | 6.232 | -0.118 |
| 28 | 6.076 | 5.855 | 0.221 |
| 30 | 5.699 | 5.852 | -0.153 |
| 35 | 5.162 | 5.165 | -0.003 |
| 36 | 4.576 | 4.397 | 0.178 |
| 49 | 6.004 | 5.965 | 0.040 |
| 56 | 6.046 | 6.055 | -0.010 |
| 58 | 5.451 | 5.681 | -0.230 |
| 57 | 5.177 | 5.047 | 0.130 |
| 52 | 5.507 | 5.720 | -0.213 |
| 45 | 5.157 | 5.662 | -0.505 |
| 43 | 6.076 | 5.884 | 0.192 |
| 46 | 5.582 | 5.381 | 0.201 |
| 39 | 5.479 | 5.204 | 0.275 |
| 44 | 6.114 | 5.843 | 0.270 |
| 59 | 7.000 | 6.953 | 0.047 |
| 51 | 5.539 | 5.509 | 0.030 |
| 47 | 5.007 | 5.300 | -0.293 |
| 45 | 4.922 | 4.815 | 0.106 |

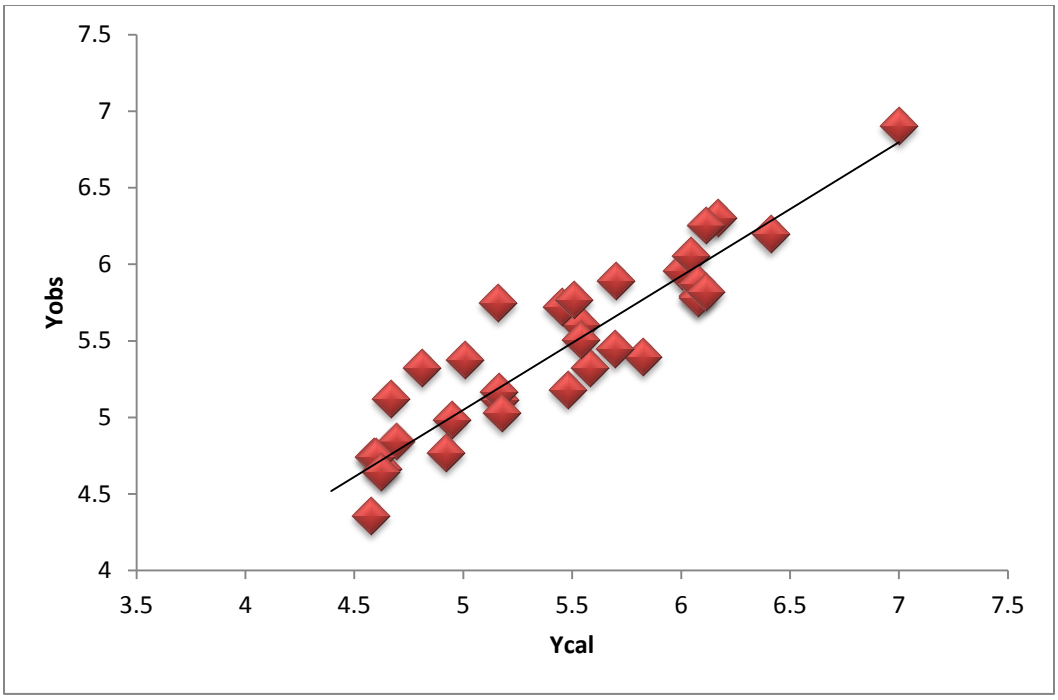



**Fig 4.7: Plotof observed pMIC against predicted pMIC of modelfor model 4.7**

**Table 4.20 Observed Activity versus Predicted Activityfor Test Sets for model 4.7**

| Name | $Y_{obs}$ | $Y_{pred}$ | $(Residual)^2$ |
|---|---|---|---|
| 2 | 4.614036 | 5.013721 | 0.159748 |
| 10 | 4.574629 | 4.781742 | 0.042896 |
| 29 | 5.534617 | 5.747088 | 0.045144 |
| 36 | 5.534617 | 5.747088 | 0.045144 |
| 53 | 5.111259 | 5.629055 | 0.268113 |
| 54 | 5.395774 | 5.351096 | 0.001996 |
| 55 | 5.149354 | 5.685104 | 0.287029 |
| 42 | 5.161151 | 5.222114 | 0.003716 |
| 40 | 5.512862 | 5.14186 | 0.137643 |
| 50 | 6.356547 | 6.124436 | 0.053876 |
| 48 | 4.666956 | 4.95964 | 0.085664 |

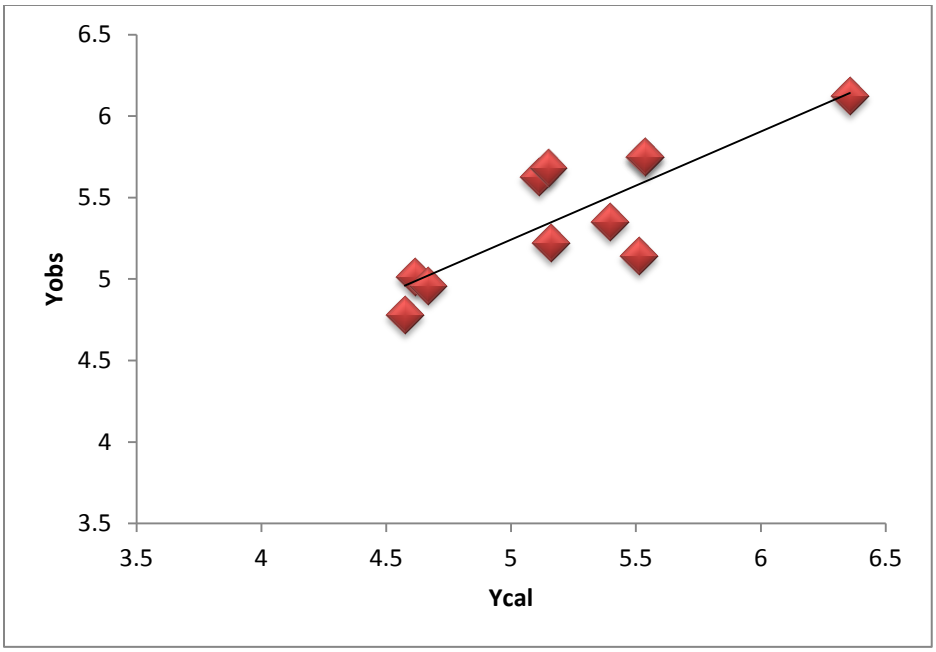



**Fig. 4.8:Plot of observed pMIC against predicted pMIC (Test Sets) for model 4.7**

**Table 4.21 Y-Randomization Result for model 4.7**

| Model | R | $R^2$ | $Q^2$ |
|---|---|---|---|
| Original | 0.948285 | 0.899245 | 0.840345 |
| Random 1 | 0.401913 | 0.161534 | -0.331800 |
| Random 2 | 0.583979 | 0.341032 | -0.028780 |
| Random 3 | 0.325849 | 0.106178 | -0.464030 |
| Random 4 | 0.543584 | 0.295483 | -0.372120 |
| Random 5 | 0.543898 | 0.295825 | -0.140910 |
| Random 6 | 0.536894 | 0.288255 | -0.005290 |
| Random 7 | 0.458818 | 0.210514 | -0.399420 |
| Random 8 | 0.39125 | 0.153076 | -0.416570 |
| Random 9 | 0.292111 | 0.085329 | -0.478010 |
| Random 10 | 0.300356 | 0.090214 | -0.584850 |

Random Models Parameters
Average R:, 0.437865
Average $R^2$ :, 0.202744
Average $Q^2$ :, -0.32218

cRp^2 :, 0.79764

**Table 4.22: Correlation matrix for descriptors in models 4.7, 4.8 and 4.9**

| | *pMIC* | *ATSC4p* | *GATS5i* | *nHBint3* | *maxHBint6* | *MDEC-23* | *RDF65u* | *RDF90u* |
|---|---|---|---|---|---|---|---|---|
| pMIC | 1 | | | | | | | |
| ATSC4p | -0.12615 | 1 | | | | | | |
| nHBint3 | -0.18023 | 0.355843 | 0.132688 | 1 | | | | |
| maxHBint6 | -0.14142 | -0.09484 | -0.15248 | 0.379231 | 1 | | | |
| MDEC-23 | -0.21182 | -0.65826 | -0.38273 | 0.183825 | 0.142275 | 1 | | |
| RDF65u | -0.06533 | 0.088965 | -0.02152 | 0.511373 | -0.06185 | 0.5364 | 1 | |
| RDF90u | -0.57389 | -0.07744 | -0.32916 | 0.145533 | -0.1501 | 0.562108 | 0.669464 | 1 |

**Table 4.23 Names of the descriptors.**

| Descriptor Symbol | Descriptor Name |
|---|---|
| **ATSC4p** | Centered Broto-Moreau autocorrelation - lag 4 / weighted by polarizabilities |
| **ATSC4i** | Centered Broto-Moreau autocorrelation - lag 4 / weighted by first ionization potential |
| **AATSC7e** | Average centered Broto-Moreau autocorrelation - lag 7 / weighted by Sanderson electronegativities |
| **AATSC7i** | Average centered Broto-Moreau autocorrelation - lag 7 / weighted by first ionization potential |
| **ATSC6e** | Centered Broto-Moreau autocorrelation - lag 6 / weighted by Sanderson electronegativities |
| **GATS5i** | Moran autocorrelation - lag 5 / weighted by first ionization potential |
| **nHBint3** | Count of E-State descriptors of strength for potential Hydrogen Bonds of path length 3 |
| **maxHBint6** | Maximum E-State descriptors of strength for potential Hydrogen Bonds of path length 6 |
| **MDEC-23** | Molecular distance edge between all secondary and tertiary carbons |
| **MLFER_S** | Combined dipolarity /polariability |
| **TDB10i** | 3D topological distance based autocorrelation - lag 10 / weighted by first ionization potential |
| **RDF65u** | Radial distribution function - 065 / unweighted |
| **RDF90u** | Radial distribution function - 090 / unweighted |

### 4.7 Result of cytotoxicity ($IC_{50}$) study of 8-Methylquinolones

The models below represent the three best models generated by the GA-MLRA analysis carried out in the BuildQSAR program. Initially several models were generated, from which the three best were selected based on their high squared correlation coefficient $R^2$, low standard error *s* and high Fisher coefficient *F*. Table 4.24 shows the statistical parameters as well as the internal and external validation parameters $Q^2_{LOO}$, $R^2_{Pred}$respectively of the three selected models.The regression models are all significant at the F-value using the Fisher F-statistics.

**MODEL 4.10**

pIC50 = 3.88274(+/-0.00174) +0.04578(+/-0.00208) **AMR** -0.01476(+/-0.0016) **nsssCH**

**MODEL 4.11**

pIC50 = 3.88318(+/-0.00227) +0.02185(+/-0.0021) **AATS6p** +0.03635(+/0.00253)**SpDiam_Dzv**

**MODEL 4.12**

pIC50 = 3.88186(+/-0.00589) +0.04216(+/-0.00692) **AMR**

However sincethe selection of a robust and well predictive QSAR model on the basis of only $R^2$, $Q^2$ and $R^2_{pred}$ might be misleading,the selected models were subjected to further statistical analysis as suggested by Golbraikh and Tropsha (2000). The result is presented in Table 4.25,the predictive power of the three modelsis compares in Table 4.26.

Model 4.11 was chosen as the best model since its $R^2_{pred}$ value is the highest out of the three models, for this model Table 4.27 shows the experimental and calculated activities for the training sets after the Leave-One-out Internal validation while Fig 4.9 shows the corresponding plot of the experimental activities and the calculated activity of the training set. Table 4.28shows the experimental and calculated activities of the external validation set and Fig. 4.10 captures the graphic correlation between them.The Y-randomization analysis results are shown in Table 4.29.Table 4.30 is correlation matrix for the descriptors in the three models.

**Table 4.24 Fitting Parameters of Models4.10 4.11 and 4.12**

| S/N | $R^2$ | $R^2_{adjs}$ | F | s | $Q^2_{LOO}$ | $R^2_{pred}$ | SPress | SDEP | n | k |
|---|---|---|---|---|---|---|---|---|---|---|
| 4.10 | 0.98647 | 0.98260 | 255.16918 | 0.00547 | 0.9754 | 0.95729 | 0.00038 | 0.00617 | 10 | 2 |
| 4.11 | 0.97684 | 0.97023 | 147.63356 | 0.00716 | 0.91459 | 0.97829 | 0.00132 | 0.01150 | 10 | 2 |
| 4.12 | 0.82253 | 0.80034 | 37.07696 | 0.01854 | 0.73041 | 0.97652 | 0.00418 | 0.02043 | 10 | 1 |

**Table 4.25 Golbraikh and Tropsha acceptable model criteria for models 4.10, 4.11 and 4.12**

| MODEL | $Q^2$ | $R^2_{EXT}$ | $\lvert R_0^2 - R'_0{}^2 \rvert$ | $[(R^2 - R_0^2)/ R^2]$ | $[(R^2 - R_0^2)/ R^2]$ | k | k' |
|---|---|---|---|---|---|---|---|
| 4.10 | 0.97540 | 0.95729 | 0.20397 | 0.11205 | 0.32512 | 0.99834 | 1.00161 |
| 4.11 | 0.91459 | 0.97829 | 0.18643 | 0.11740 | 0.30796 | 0.99800 | 1.00196 |
| 4.12 | 0.73041 | 0.97652 | 0.11300 | 0.08608 | 0.20180 | 0.99902 | 1.00095 |

**Table 4.26 Comparison of Observed Activity with Predicted Activity of Models 4.10,4.11 and 4.12**

| COMP NO | OBSERVED ACTIVITY | PREDICTED ACTIVITY | | |
|---|---|---|---|---|
| | | MODEL**4.10** | MODEL **4.11** | MODEL **4.12** |
| 10 | 3.904 | 3.939 | 3.939 | 3.941 |
| 12 | 3.837 | 3.902 | 3.894 | 3.874 |
| 15 | 3.919 | 3.839 | 3.830 | 3.832 |
| 18 | 3.861 | 3.910 | 3.920 | 3.915 |
| 24 | 3.885 | 3.857 | 3.859 | 3.849 |
| 31 | 3.873 | 3.930 | 3.931 | 3.933 |
| 36 | 3.909 | 3.893 | 3.884 | 3.883 |
| 39 | 3.869 | 3.907 | 3.904 | 3.896 |
| 43 | 3.795 | 3.866 | 3.877 | 3.890 |
| 53 | 3.932 | 3.811 | 3.817 | 3.839 |

**Table 4.27 Observed Activity versusPredicted Activity for Model 4.11**

| COMP. NO | pMIC OBSERVED | pMIC CALCULATED | RESIDUAL |
|---|---|---|---|
| 4 | 3.936 | 3.941 | -0.005 |
| 10 | 3.904 | 3.874 | 0.030 |
| 12 | 3.837 | 3.832 | 0.005 |
| 15 | 3.919 | 3.915 | 0.004 |
| 18 | 3.861 | 3.849 | 0.011 |
| 19 | 3.926 | 3.933 | -0.007 |
| 24 | 3.885 | 3.883 | 0.002 |
| 36 | 3.909 | 3.896 | 0.013 |
| 39 | 3.869 | 3.890 | -0.021 |
| 48 | 3.808 | 3.839 | -0.032 |

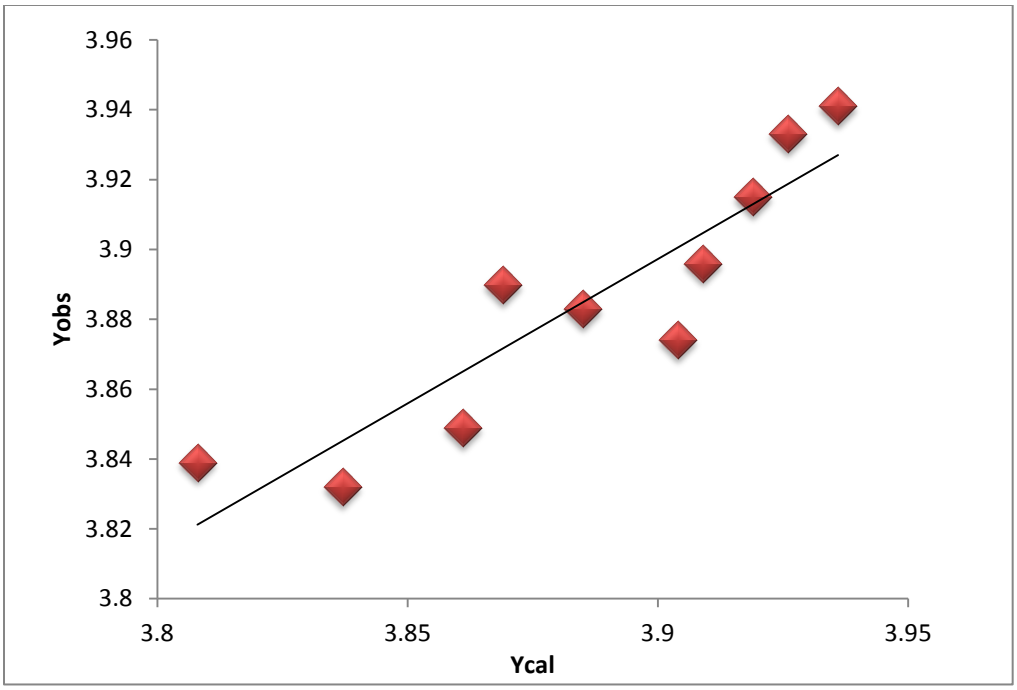



**Fig 4.9:Plot of observed pMIC against predicted pMIC of model for model 4.11**

**Table 4.28Observed Activity versus Predicted Activity (Test Sets) for models 4.11**

| Name | $Y_{obs}$ | $Y_{pred}$ | $(Residual)^2$ |
|---|---|---|---|
| 25 | 3.792258 | 3.824194 | 0.00102 |
| 28 | 3.962295 | 3.946223 | 0.000258 |
| 31 | 3.872928 | 3.878256 | 0.000028 |
| 43 | 3.794552 | 3.836508 | 0.00176 |
| 53 | 3.932074 | 3.909112 | 0.000527 |

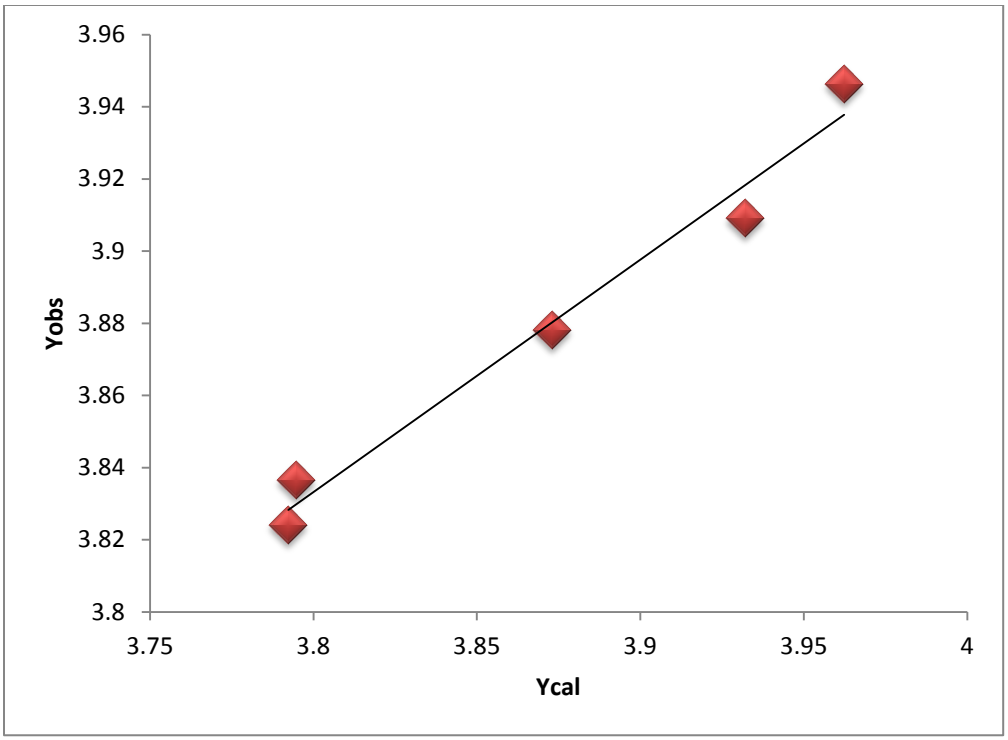



**Fig 4.10Plot of observed pMIC against predicted pMIC (Test Sets) for model 4.11**

**Table 4.29**: **Y-Randomization Result for model 4.11**

| Model | R | $R^2$ | $Q^2$ |
|---|---|---|---|
| Original | 0.988353 | 0.976842 | 0.914586 |
| Random 1 | 0.5928 | 0.351412 | -0.72999 |
| Random 2 | 0.462329 | 0.213748 | -0.25672 |
| Random 3 | 0.577644 | 0.333673 | -0.7044 |
| Random 4 | 0.235494 | 0.055457 | -4.83648 |
| Random 5 | 0.309699 | 0.095914 | -0.44923 |
| Random 6 | 0.497381 | 0.247388 | -3.2135 |
| Random 7 | 0.763864 | 0.583488 | -0.57565 |
| Random 8 | 0.43163 | 0.186305 | -0.68229 |
| Random 9 | 0.44201 | 0.195373 | -0.66982 |
| Random 10 | 0.205131 | 0.042079 | -2.71424 |

Random Models Parameters

Average R =0.451798

Average $R^2$ = 0.230484

Average $Q^2$ = -1.48323

$R^2p$ = 0.868807

**Table 4.30 correlation matrix for model 4.11**

| | *AMR* | *AATS6p* | *nsssCH* | *SpDiam_Dzv* |
|---|---|---|---|---|
| AMR | 1 | | | |
| AATS6p | 0.485888 | 1 | | |
| nsssCH | 0.189033 | -0.05112 | 1 | |
| SpDiam_Dzv | 0.789894 | -0.07114 | -0.06124 | 1 |

# CHAPTER FIVE

## 5.0 DISCUSSION.

### 5.1 Interpretation and evaluation of model 4.2

**SCH-3** known as *Simple chain, order 3* was found to be strongly correlated to the anti-tuberculosis activity. This descriptor has been reported elsewhere (Manish *et al.*,2007) to be important in determining the anti-tuberculosis activity of quinolones. The sign of the correlation is an indication of whether the two variables are positively (more *X* means more *Y*) or negatively (more *X* means less *Y*) related. As can be seen from model 4.1, the pMIC and SCH-3 have a good positive correlation and meaning that the descriptor SCH-3 strongly influences the anti-tuberculosis activity of 8-methylquinolones against *mycobacterium tuberculosis.* The sign of the coefficient of this descriptor is positive implying that groups having more branching are conducive for the activity and this is consistent with the conclusion of Srivastava *et al.,* (2010) in a similar study where the positive sign of this descriptor implies that it contributes positively to the potency of these compounds against the organism and hence an increase in the value of the descriptor will lead to an increase in the potency of the compounds.

**AATSC5p** which stands for *Average centered Broto-Moreau autocorrelation - lag 5 / weighted by polarizabilities* and **MATS4c** which stands for the *Moran autocorrelation - lag 4 / weighted by charges*are 2D Autocorrelation Descriptors. The symbol for each of the autocorrelation descriptors is followed by two indices *d* and *w* where *d* stands for the lag and *w* stands for the weight. The lag is defined as the topological distance *d* between pairs of atoms. The topological distance between a pair of atoms (*i, j*) is given in the *ijth* entry in the Topological Level Matrix. The lag can have any value from the set {0, 1, 2, 3, 4, 5, 6, 7, 8}.The weight can be *m* (relative atomic mass), *p* (polarizability), *e* (Sanderson electronegativity) and *v* (Van der Waals volume). *Relative mass* is

defined as the ratio of atomic mass of an atom to that of carbon. Similarly, the other three weights *p*, *e* and *v* are scaled by the corresponding values for Carbon. The physical interpretation of 2D Autocorrelation descriptors is complicated by the fact that it essentially involves the condensation of topological and atomic information into a single number. However, these descriptors have been widely used in medicinal chemistry. In an earlier study on fluoroquinolonesMinovski*et al*., (2011) foundMATS4m and other topological and electro-topological parameters, to be important in modeling the anti-tuberculosis activity of these compounds.

**VE3_Dzp**: is the Logarithmic coefficient sum of the last eigenvector from Barysz matrix / weighted by polarizabilities, itis based on the Barysz matrix, and is therefore concerned with the presence of heteroatoms and multiple bonds in the molecule. **PetitjeanNumber**is the petitjean Number descriptor. By simple inspection of the descriptors we can see that the descriptors **AATSC5p**,**MATS4c and VE3_Dzp** are weighted bypolarizabilities and charges. If we now base our assumptions on the propertiesaccording to which the descriptors were calculated, wecan say that clearly the factors that influence the developmentof biological activity of the active compounds analyzed herefor the inhibition of growth of *M. tuberculosis* are principallysteric and (in a certain way) electronic in nature; this conclusion is consistent with that of Alejandro *et al*.,(2009). In a similar study,Foroumadi*et al.* (2007) made a similar conclusion where they foundMean atomic polarizability (Mp) which is a constitutional descriptor and Maximum Positive Charge to be responsible forthe anti-tubercular activity for 2-(nitroaryl)-5-(nitrobenzylsulfinyl and sulfonyl)-1,3,4-thiadiazole derivatives leading them to conclude that electronic distribution is among the most important factors in determining the anti-tubercular activity of this group of compounds.

The high correlation coefficient R (0.91) indicates the susceptibility of descriptors (AATSC5p, MATS4c, VE3_Dzp, SCH-3 and SCH-3) to form model 4.2. Squared correlation coefficient ($R^2$) of

0.83 explains 83% variance in biological activity of the tested compounds. It also indicates the statistical significance >99.9% with F-values (41.0676).Cross-validated square correlation coefficient ($Q^2$) by LOO technique was 0.74 which showed a good internal predictive ability of the model.The model was also validated by applying the Y-randomization test. Ten random shuffles of the Y vector were performed and the obtained results are in good agreement with the suggested limits (Eriksson *et al*., 2003).The low $R^2$ and $Q^2_{LOO}$ values of the random models shown in Table 4.6 and the value of $R_P^2 = 0.808101$ ($R_P^2 \geq 0.5$) indicates that there is no chance correlation or structural dependency in the proposed model.Consequently model 4.2 can be considered as a perfect model with both high statistical significant and excellent predictive ability.The applicability domain (AD) of model 4.2 (Williams plot) shown in Fig 4.3 was generated using thewell-known leverage approach. Training set objects (28 compounds with experimental activity values) used in the model development are presented as solid dots, whereas the externalvalidation set objects (9 compounds) presented as solid rectangles labeled with the corresponding number (ID signature). The analysis of the AD for the training set objects shows that only one compound labeled with (ID20) signature can be identified as a typical*X*-outlier ($h > h^* = 0.643$).However the pMIC prediction of this compound was fairly good as can be seen in Table 4.4. None of the external validation set is a typical *X*-outliers with $h > h^*$. Also no compound within the training as well as the external validation set is a typical Y-outlier since the cut off value for standard deviation is $\pm 3.0\sigma$.

**5.2Interpretation and evaluation of model 4.4**

The polar surface area (PSA) is defined as the surface sum over polar atoms, (usually oxygen and nitrogen), including also attached hydrogens. PSA is a commonly used medicinal chemistry metric for the optimization of cell permeability,although it has been reported in anotherstudy to be negatively correlated to the anti-tuberculosis activity (Madhusudana *et al*., 2014).

**SCH-3** known as *Simple chain, order 3* was also found to be strongly correlated to the anti-tuberculosis activity , this descriptor has been reported by Manish *et al.*, (2007) to be important in determining the anti-tuberculosis activity of quinolones against *M. Smegmatis* .It is informative to note that the models for the two organisms( *M .tuberculosis* and *M .smegmatis* ) have **SCH-3** as a common descriptor. This is a pointer to the fact that 8-methylquinolones likely inhibit the growth of these two organisms through the same mechanism. The sign of the coefficient of this descriptor is positive implying thatit contributes positively to the potency of these compounds against the organism;hence an increase in the value of the descriptor will lead to an increase in the potency of the compounds. **AVP-1**known asthe *Valence path, order 1* is a valence degree property and is related to the valence electronsand therefore can differentiate between atoms belongingto different groups of the Periodic Table. Thus, in some way, valence degree characterizes properties related to electronegativity (Alejandro *et al*.,2009).**AATS4i** is *Average Broto-Moreau autocorrelation - lag 4 / weighted by first ionization potential*. **maxHCsat** is*Maximum atom-type H E-State: H bonded to B, Si, P, Ge, As, Se, Sn or Pb.* This descriptoris negatively correlated to the activity.As was the case with the activity of 8-methylquinolones against *M.tuberculosis* described above, from inspection of the descriptors it could be seen that the steric factor and electronic factors are more dominating towards the activity of this group of compounds against *M .smegmatis*.

The usefulness of QSAR models is not just their ability to reproduce known data, verified by their fitting power ($R^2$), but mainly is their potential for predictive application. For this reason, the internal consistency of the training set was confirmed by using leave-one-out (LOO) cross-validation method to ensure the robustness of the model.The high calculated$Q^2_{LOO}$ value, 0.8498 suggests a good internal validation and indicates the robustness and stability of the constructed model.The leverage values can be calculated for every compound and plotted vs. standardized

residuals, and it allows a graphical detection of both the outliers and the influential chemicals in a model.Fig. 4.4, shows the Williams plot, the applicability domain is established inside a squared area within ±3 bound for residuals and a leverage threshold h*,$h^* = \frac{3(P+1)}{n}$ where p is the number of model parameters and n is the number of compounds) (OECD, 2004). It demonstrates that all the compounds of the training set and test set are inside of this square area.The analysis of AD for the training set and sets objects shows that only one compound labeled with (ID38) signature can be identified as a typical *X*-outlier ($h > h^* = 0.667$ ). Also, two compounds from the external validation set (ID30 and ID32) are typical *X*-outliers with $h > h^*$

No compound in both training and test set is a typical Y-outlier with the cut off value for standard deviation set at $\pm 3.0\sigma$. In order to assess the robustness of the model, the Y-randomization test was applied in this study, Y-randomization test confirms whether the model is obtained by chance correlation, and is a true structure–activity relationship to validate the adequacy of the training set molecules. The steps followed during the randomization test are

(I) repeatedly scrambling the activity data in the training set molecules,

(II) using the randomized data to generate QSAR equations, and

(III) Comparing the resulting scores with the score of the original QSAR equation generated with non-randomized data. If the activity prediction of the random model is comparable to that of the original equation, the set of observations is not sufficient to support the model. The new QSAR models (after several repetitions) would be expected to have low $R^2$ and $Q^2_{LOO}$ values. If the opposite happens, then an acceptable QSAR model cannot be obtained for the specific modeling method and data. The results of Table 4.13 indicate that an acceptable model was obtained, and the model developed is statistically significant and robust.

Often times in the statistic model the independent variables are highly correlated. This is not a consequence of the statistical nature of the method employed but is due to the inherent properties of the descriptors used. This disturbing factor if overlooked can lead to instability in the model. The correlation coefficient between independent variables should not exceed $r = 0.7$ (Alejandro, 2009) where *r*, the correlation coefficient, is traditionally considered as a cutoff for the analysis of collinearity between variables. The correlation matrix, (Table4.14), show that the variablesin model are not highly correlated and thus they give specific structural information.

### 5.3 Interpretation and evaluation of model 4.7

The **RDF65u and RDF90u** are among the large number of 3*D* molecular descriptors.They are called Radial Distribution Function descriptors (RDF) and have demonstrated their potential as useful tools for modeling different physico-chemical and biological properties. These descriptors are based on the distances distribution in the geometrical representation of a molecule and constitute a radial distribution function code. Formally, the radial distribution function of an assembly of *N* atoms can be interpreted as the probability distribution of finding an atom in a spherical volume of radius *R*. The general code of the radial distribution function is represented by

$$RDF(R,A) = f \sum_{i=1} \sum_{j>i} A_i \, A_j \, e^{-B(R-R_{ij})^2} \ldots\ldots\ldots (15)$$

where $f$ is the scaling factor and $N$ is the number of atoms. $A \in \{u, m, v, e, p\}$, where $A$ is the characteristic atomic property of the atoms $i$ and $j$ (unweighted, atomic masses, van der Waals volumes, electronegativities and polarizabilities). The exponential term contains the distance $R_{ij}$ between the atoms $i$ and $j$ and the smoothing parameter $B$, which defines the probability distribution of the individual distances. RDF is calculated for a number of discrete points as $10 \leq R \leq 155$ with the defined interval of 5. RDF descriptors are independent of the number of atoms. They are unique regarding the three-dimensional arrangement of the atoms, and they are invariant against translation

and rotation of the entire molecule. Besides information about interatomic distances in the entire molecule, RDF descriptors provide further valuable information about bond distances, ring types, planar and non-planar systems and atom types. (Suleiman *et al*., 2014)

**ATSC6c**and **ATSC4p** are among the 2D Autocorrelation descriptors that have been described in section 5.1and as was stated in that section, the physical interpretation of 2D Autocorrelation descriptors is complicated by the fact that it essentially involves the condensation of topological and atomic information into a single number.**MLFER_S** is a descriptor for "polarity/polarisability" (Platt *et al*., 1999) and is negatively correlated with pMIC. Model 4.6.1 indicates that **RDF65u and ATSC6c** demonstrate positive contributions while**, ATSC4p, MLFER_S, and RDF90u** demonstrates negative contribution towards the anti-*tuberculosis activities*, respectively. Based on the model 4.7in order to design a new ligand with high potency, substitutients with high amount of **RDF65u** and **ATSC6c** low amounts of **ATSC4p, MLFER_S, TDB10i, and RDF90u**should be considered.

A quantitative structure activityrelationship (QSAR) studies were performedNeeraja*et al*., (2011) to explore the anti-tuberculosis compound from the derivatives of natural products .Theoretical results found were in accord with the *in vitro* experimental data with reported growth inhibitory activity towards *Mycobacteriumtuberculosis* or related organisms. QSAR studies indicate that dipole energy and heat of formation correlate well withanti-tubercular activity. A study was carried by Ponnurengam*et al.,* (2007) on chalcones, chalcone-like compounds, flavones and flavanones to understand the relationship between anti-tuberculosis activity and structural features, the best molecular descriptors were identified to be Jurs descriptors (Jurs charged partial surface area), hydrogen bond donor, principal moment of inertia, molecular energy, dipole moment, molecular area, absorption, distribution, and Chi indices (Kier and Hall chi connectivity indices). They

concluded that the compounds reported were more lipophilic in nature and hence, as expected, exhibit good activity since *M. tuberculosis* has a high concentration of lipid layer.

The high correlation coefficient R (0.91) indicates that the model is statistically significant, Squared correlation coefficient ($R^2$) of 0.90178 explains 90% variance in biological activity of the tested compounds. It also indicates the statistical significance >99.9% with F-values (42.84356). Cross-validated square correlation coefficient ($Q^2$) by LOO technique was 0.82634 which showed a good internal predictive ability of the model. The model was also validated by applying the Y-randomization test. Several random shuffles of the Y vector were performed and the obtained results are in good agreement with the suggested limits (Eriksson *et al*., 2003).The low $R^2$ and $Q^2_{LOO}$ values of the random models shown in Table 4.21and the value of $R_P^2$ = 0.808101 ($R_P^2 \geq 0.5$) indicates proposed model is robust.

**5.4Interpretation and evaluation of model 4.11**

**AATS6p**is theAverage Broto-Moreau autocorrelation - lag 6 / weighted by polarizabilities.This descriptor showed positive contribution to the cytotoxic activity of these compounds. **SpDiam_Dzv**known asSpectral diameter from Barysz matrix / weighted by van der Waals volumes is a type of Barysz Distance Matrix *Dz*: This is a weighted distance matrix accounting for the presence of heteroatoms and multiple bonds which also shows positive contribution to the cytotoxic activity of 8-methylquinolones. The *ijth* entry in the matrix *Dz* is defined as (Barysz *et al*., 1983):

$(D^z)_{ij} = 1- (Z_c /Z_i )$, if i=j ……..(16)

$(D^z)_{ij} = \sum_{b-1}^{d_{ij}} Z_c^2 / (\prod_b^*. Z_{b(1)} Z_{(2)} )$……..(17)

Where$Z_C$is the atomic number of carbon,

$Z_i$is the atomic number of the *ijth* atom,

$\prod_b^*$ is the conventional bond order of the bond *b*, that is 1,2,3 and 1.5 for single, double, triple and aromatic bonds respectively;$Z_{b(1)}$ and $Z_{b(1)}$ are the atomic numbers of the atoms on the bond *b*, and the sum runs over all the bonds *b* in the shortest path from the atom *i* to the atom *j*.

The model 4.11 has good capacity of explaining the observed values of biological activity. It explains 91.7% of the variance in the anti-tubercular activity against HR37v strain, which can also be confirmed by the low standard deviation. The model possesses excellent adjustment level: high correlation coefficient ($R$ = 0.958) and low standard deviation ($s$ = 0.131). The model also has good predictive capacity ($Q^2_{LOO}$ = 0.843 and SPRESS= 0.181).As can be seen from Table 4.28, the calculated values for the $pIC_{50}$ are in good agreement with those of the experimental. Also, in order to assess the robustness of the model, the Y-randomization test was applied in thisstudy. The dependent variable vector (pIC50) was randomly shuffled and a new QSAR modeldeveloped using the original independent variable matrix. The new QSAR models (after several repetitions) would be expected to have low $R^2$ and $Q^2_{LOO}$ values (Table 4.29). If the opposite happens then an acceptable QSAR model cannot be obtain for the specific modeling methodand data. Table 4.30 gives the correlation matrix of the descriptors.

# CHAPTER SIX

## 6.0 SUMMARY, CONCLUSION AND RECOMMENDATIONS

### 6.1 SUMMARRY

This study was focused on developing valid models capable of predicting the*in vitro* antimycobacterial activity (MIC) of some quinolones against *Mycobacterium tuberculosis H37Rv (MTB) and Smegmatis ATCC 14468 (MC2)* as well as their cytotoxicity. These models could subsequently be used to design new potential anti-tuberculosis drug candidates with higher potency and less toxicity. For this purpose a QSAR approach was utilized. One of the most important steps in QSAR modeling is to define the number of independent variables in the model equation obtained. In this way the over- parameterization of the mathematical model as well as the chance correlation between the molecular descriptors is avoided (Topliss and Edwards, 1979). Since the modeling procedure started with a pool of several theoretical molecular descriptors, a possibility exists to encounter a chance correlation in a case where the number of examined variables is higher than the number of observations.

The GA was applied to select from the pool of calculated descriptors, the best combinations of those most relevant for obtaining models with the highest predictive power for the activities of interest. Eventually, the MLRA was employed to build different QSAR models. The real usefulness of QSAR models is not just their ability to reproduce known data, verified by their fitting power ($R^2$), but is mainly their potential for predictive application. For this reason the model calculations were performed by maximizing the explained variance in prediction, verified by the leave-one-out cross validation. Initially several models were generated, some of these models with high squared correlation coefficient $R^2$, low standard error $s$ and high Fisher coefficient $F$ were selected for

further analysis; this analysis included internal validation using the Leave One Out technique (LOO). Those models that showed good internal stability were externally validated by testing their predictive performances using the test set. Since the selection of a robust and well predictive QSAR model on the basis of only $R^2$, $Q^2$ and $R^2_{pred}$ might be misleading, the selected models were subjected to further statistical analysis as suggested by Golbraikh and Tropsha (2000).

## 6.2 CONCLUSION

This study focused on developing valid models that would predict the activity of derivatives of 1-(cyclopropyl/2, 4-difluorophenyl/tert-butyl)-1, 4-dihydro-8-methyl-6-nitro-4-oxo-7 (substituted secondary amino)quinoline-3-carboxylic acids which could subsequently be used to design new potential drug candidates, with better activity.The results of the study shows **SCH-3 (**known as Simple chain, order 3)to be strongly correlated to the anti-tuberculosis activity,suggesting that substituent groups atthe C-7 and N-1 positions having more branching are conducive for the anti-tubercular activity. The following descriptors were also found to be to important and in modeling the anti-tubercular activity of these class of compounds;**AATSC, MATS4c, VE3_Dzp,PetitjeanNumber, PSA, AVP-1, maxHCsats, AATS4i,RDF65u, ATSC6c, ATSC4p,MLFER_S, TDB10i, and RDF90u**.This result implies that steric factor is more dominating towards the anti-tubercular activity a was earlier suggested by other workers (Alejandro *et al.,* 2009).The most important descriptors for modeling the cytototoxicity of these group of compounds was found to be**AATS6p** and **SpDiam_Dzv.** In other to design potent anti-tubercular drugs within this class of compounds:

i) those substituent with more branching at the C-7 and N-1 positions should be given priority

ii) Steric factors should be considered first before other factors such as electronic etc.

iii) It is also recommended that the **AATS6p** Average Broto-Moreau autocorrelation - lag 6 / weighted by polarizabilities and**SpDiam_Dzv** Spectral diameter from Barysz matrix / weighted by van der Waals volumes which accounts for the presence of heteroatoms and multiple bondsshould be given priority in designing anti-tuberculosis drug with minimal cytotoxic effect.

Finally, the predictive power of the models which was tested through the extrapolation of the model over the external previously excluded validation data set suggests that they could be used in predicting the anti-tubercular potency and cytotoxicity of this class of compounds prior to their synthesis.

**6.30 RECOMMENDATIONS**

This study was aimed at building QSAR models that could predict the potency of some quninolone compounds against the causative agent of TB, and also to predict the cytotoxicity of these compounds. However it is also important to predict the binding-conformation of these compounds to the target binding site.

Properties related to absorption,distribution, metabolism, excretion have becomeone ofthe most important issues in to assess the effects or risk of compounds on human body. In the light of this, the following are recommended asfurther studies.

1. Molecular docking, which is one of the most frequently used methods in drug design, due to its ability to predict binding conformation of small molecules ligands to appropriate target binding site should be carried out on this class of compounds

2. Other properties of the compounds such as absorption,distribution, metabolism, and excretion should be modeled.

.

## REFERENCES


Ahmed, K., Shaik, A. M., Shaheer, M., Ahmad, A. S., and Maddamsetty, V. R. (2008). Efforts Towards the Development of New Anti-tubercular Agents: Potential for Thiolactomycin Based Compounds,*Journal of Pharmaceutical Sci*ence 11 (2), 56s-80s

Alangaden, G.J., and Lerner S.A.(1997).The clinical use of fluoroquinolones for the treatmentofmycobacterial diseases. *Clinical Infectious Disease*, 25, 1213-1220.

Alejandro S.P., Marcus T.S., América G L., Vicente P. E., Enrique M.P., and Eugenio.U (2009). Design of novel antituberculosis compounds using graph-theoretical and sub-structural approaches.*Molecular Diversity*,13, 445–458

Ambure, p., Aher,J.,Gajewicz, A.,Puzyn,T., and Roy, K.(2015). "NanoBRIDGES" sofwares:open access tools to perform QSAR and nano-QSAR modeling. *Chemometrics and intelligent Laboratory Systems*, 147, 1-3.

Anil,K.,Eric, A., Nacer, L., Jerome, G.,andKoen, A. (2011). The challenge of drugdiscovery for tuberculosis. *Nature, 469(483).*

Alexander,T. (2010). Best Practices for QSAR Model Development, Validation,andExploitation *Molecular Informatics*, 29, 476 – 488.

Barry, C.E., Lee R.E., and Mdluli, K. (1998). Mycolic acids: structure, biosynthesis and physiological functions, *Progress in Lipid Research*, 37: 143–179.

Barry, C. E., and Blanchard, J. S. (2010). The chemical biology of new drugs in the development for tuberculosis. *Current Opinion in Chemical Biology*, 14 (4):456-466.

Bartz, Q.R., Ehrlich J., Mold J.D., Penner, M.A., and Smith, R.M. (1951). Viomycin, a new tuberculostaticantibiotic. *American Review of Tuberculosis*, 63, 4–6.

Barysz, M., Trinajstic N., and Knop J.V. (1983). On the similarity of the Chemical Structures.*International Journal of Quantum Chemistry*. 24,(17) 441-451

Berk, R.A. (2003). The formalities of multiple regression, in Regression Analysis: *A Constructive Critique*, R.A. Berk, ed., London: SAGE Publications Ltd, pp. 103–110.

Bienenstock, J.,and Shaldon S. (1963). Thiacetazone in tuberculosis. *Lancet*, 2, 817–818

Blumberg, H.M., Burman, W.J., Chaisson, R.E., Daley, C.L., Etkind, S.C., Friedman, L.N., Fujiwara, P., Grzemska, M., Vernon, A.A. (2006). Treatment of tuberculosis**.** *AmericanJournal of Critical Care Medicine*.167(4), 603-62.

Boteva. A.A., and Krasnykh, O. P. (2009). The methods of synthesis, modification andbiological activity of 4-quinolones, *Chemistry of Heterocyclic Compounds*, 45(7), 757-785

Bowen, D.A., and Collins D.M. (1952). Development of drug resistance to isoniazid in case of pulmonary tuberculosis. *Tubercule* 33, 276-278

Capon, A.W. (1954). Streptomycin and PAS vs. streptomycin, PAS and isoniazid in thetreatment of pulmonary tuberculosis. *Canadian Medical Association Journal*, 70, 62–67.

Chattaraj, P.K., and Roy, D.R. (2007). Update 1: of electrophilicity index. *Chemical Reviews*.

Chu, D.T.W., and Fernandes, P.B. (1991). Recent developments of the field of quinoloneantibacterial agents. *Advances in Drug Research. Test*, 21: 42–144.

Crumplin, G.C.,andSmith J.T.(1975). Nalixalic acid:an antibacterial paradox,*Antimicrobial Agents and Chemotherapy.* 8,251-261

Cuthbert, J., and Bruce L.G. (1964). Treatment of pulmonary tuberculosis by capreomycin and PAS: a small preliminary trial. *Tubercle* 45, 205–210.

Daffe. M. and Draper P. (1998). The envelope layers of mycobacteria with reference to their pathogenicity. *Advances in Microbial Physiology,* 39: 131–203.

Davis, L., (1991). Handbook of Genetic Algorithms. Van Nostrand R, New York

De-oliveira, D.B., Gaudio, A. C. (2001). BuildQsar: A New Computer Program for QSAR Analysis.*Quantitative Structure-Activity Relation,* 19(6), 599 – 601,

Devillers, J. (1996). *Genetic Algorithms in Molecular Modeling*. Academic Press Ltd.,London. *J.* Devillers, Ed. Academic Press, London, pp. 35-66

Dimasi, J.A., Hansen R,W Grbowski H.G (2003). The price of innovation:new estimates of drug development costs, *Journal of Halth Economics*. 22: 151-185

Dooneif, A.S.,Buchberg, A., and Steinbach, M.M.(1950). Paraaminosalicyclicacis (PAS) in preventing acquired resistance to streptomycin by mycobacterium tuberculosis,*Journal of Bacteriology*. 57, 459-464

Doster, B., Murray F.J., Newman, R., and Woolpert S.F. (1973). Ethambutol in the initial treatment of pulmonary tuberculosis. U.S. Public Health Service tuberculosis therapy trials. *American Review ofRespiratory Disease*, 107, 177–190

Dye, C. (1999). Global burden of tuberculosis, estimated incidence, prevalence, and mortality byCountry. *Journal of the American Medical Association*, 282, 677-686

Emami, S., Shafiee,A., and Foroumadi .A. **(**2005).Quinolones: Recent Structural and Clinical Developments. *Iranian Journal of Pharmaceutical Research*, 3, 123-136

Eriksson, L., Jaworska, J., Worth, A.P., Cronin, M.T.D., McDowell, R.M., and Gramatica P.(2003).Methods for Reliability and Uncertainty Assessment and forApplicability Evaluations of Classification- and Regression-Based QSARs.*Environmental Health Perspectives*, 111 (10), 361-372

Fisher, L. (1971). Rifampin–new and potent drug for TB treatment. *Bulletin National Tuberculosis and RespiratoryDisease Association* 57, 11–12.

Flamm, R.K., Vojtko, C., Chu, D.T., Li, Q., Beyer, J., Hensey D., Ramer N., Clement J.J., andTanaka, S.K. (1995). In vitro evaluation of ABT-719, a novel DNA gyrase inhibitor. *Agents and Chemotherapy*, 39, 964-970.

Foroumadi, A., Sakhteman,A., Sharifzadeh Z., Mohammadhosseini, N., Hemmateenejad, B., Moshafi, M.H., Vosooghi, M., Amini, M., and Shafiee (2007). A Synthesis, anti-tuberculosis activity andQSAR study of some novel 2-(nitroaryl)-5- (nitrobenzylsulfinyl and sulfonyl)-1,3,4-thiadiazole derivatives .*DARU* 15(4):218-226.

Filomina M., Christina V., Susana S and Miguel V. (2010). QSAR Based Design of New Anti-Tuberculosis Compounds: Improved Isoniazid Derivatives against Multidrug-Resistant TB. *Current Pharmaceutical Design* 20 (27), 4427-4454

Frieden F, Sterling, T.R., Munsiff, S.S, Watt CJ, Dye C. (2003). Tuberculosis,*Lancet.* 362:887-899.

Gillespie, S. H. (1999). The activity of moxifloxacin and other fluoroquinolones against*Mycobacterium tuberculosis* and other mycobacteria. In "Moxifloxacin in Practice" (D. Adam and R. Finch, eds., Vol. 1, pp. 71–79. Maxim Medical, Oxford.
*Journal of Antimicrobial Chemotherapy.* 44, 393–395

Gillespie SH and and Billigton O(1999). Activity of Moxifloxacin against mycobacteria.*Journal of Antimicrobial therapy*, 44,393-395

Global Tuberculosis Institute (2015).*Transmission and pathogenesis of TB.* Retrieved from http://globalt.njms.rutgers.edu/abouttb/transmission.html

Golbraikh, A.,and Tropsha, A. (2002). Beware of *q*2! *Journal of Molecular Graphics and Modelling*. 20, 69–276

Gootz, T.D., Brighty K.E(1998).; Chemistry and mechanism of action of thequinoloneantibacterials in: *The Quinolones*. 2nd ed. Academic Press, San Diegopp29

.

Graessle, O.E., and Pietrowski, J.J. (1949). The in vitro effect of para-aminosalicylic acid (PAS) in preventing acquired resistance to streptomycin by Mycobacterium tuberculosis. *Journal of Bacteriology, 57, 459*–464.

Gramatica, P. (2007). Principles of QSAR models validation: internal and external. *QSAR and Combinatorial Science,*26(5),694 – 701.

Gupta, R.A., Gupta, A.K., Soni, L.K., and KaskhedikarS.G. (2007). Rationalization of physicochemical characters of oxazolyl thiosemicarbazone analogs towards multi-drug resistant tuberculosis: A QSAR approach. *European Journal of Medicinal Chemistry 42,* 1109 -1116.

Hemant, K., Khairnar, J., Rajesh, J., Oswal, B., and Rishikesh V. A.**(**2012). A Review on: Antituberculosis Agents. *Journal of Pharmaceutical science and Bioscientific Research,* 2(2),40-48.

Hemmateenejad, B.R., Miri M., Jafarpour M., Tabarzad A., and Foroumadic, A. (2006).MultipleLinear Regression and Principal Component Analysis-Based Prediction of the Anti-Tuberculosis Activity of Some 2-aryl-1,3,4-Thiadiazole Derivatives.*QSAR&Combinatorial Science, 25*(1), 56 – 66.

Henvwer, K.E., Ball, D.M., and Boulamwini, J.K., (2008). Quantitative Structure-activity relationshipstudies on nitrofuranyl anti-tubercular agent. *Bioorganic & Medicinal Chemistry*.16, 8042-53

Hudgins, P.C., Patnode R.A., and Cummings M.M., (1955). The effect of cycloserine on growing and resting tubercle bacilli. *American Review of Tuberculosis*.72, 685–686.

Hutton, P.W. and Tonkin I.M. (1960). Ethionamide (1314') with streptomycin in acute Tuberculosis of recent origin in Uganda Africans: a pilot study. *Tubercle* 41:253–256

Iczkowski, R.P., and Margrave, J.L. (1961). Electronegativity. *Journal of American ChemicalSoc*iety83, 3547–3551

Iseman, M.D. (2002). Tuberculosis therapy: past, present and future. *European Respiratory Journal Supplement*36, 87s–94s

Janin,Y.L. (2007). Antituberculosis drugs: ten years of research. *Bioorganic and medicinal chemistry*. 15(7),2479-2513.

Joiner, C.L., Maclean K.S., Carroll J.D., Marsh K., Collard P., and Knox R. (1954). Isoniazid and P. A. S. in chronic pulmonary tuberculosis: a warning. *Lancet* 267:663–666.

Jhakk, A., Samad A., and Kumar, Y. (2009). 3D-QSAR studies of 1,3,4-Oxadiazole derivatives asantimycobacterial agents. *Iranian Journal of Pharmaceutical Research*. 163-7

Kaida, K., and Sugiyama, K. (1959). Clinical experience with PZA-INH therapy: report on study of resected specimens following the above therapy in particular.*British Journal of Disease of the Chest.* 36,378– 388

Khade A.P., Patil A.M., and Jain B.B. (2011).2D-QSAR study of novel Oxazoline benzyl esters

derivatives as anti-tuberculosis agents. *International Journal of Pharma and Bio Science*

4501-9

Khunt, R.C., Khedkar V.M., Chadwa R.S., (2012). Synthesis, antitubercular evalution and 3D-

QSAR study of N-Phenyl-3-(4fluorophenyl)-4-substituted pyrazole derivatives. *Bioorganic & Medicinal Chemistry Letter*. 22, 666-78

Kola, I.,and Landis J.(2004). Can the pharmaceutical industry reduce attrition rates? *Nature Reviews Drug Discovery*.3 (8), 711–715

Kumar D., Raj K.K., and Bailey M., (2013). Anti-mycobacterial activity evaluation, time-kill

kinetics and 3D-QSAr study of C-(3-aminomethyl-cyclohexyl)-methylamine derivative.

*Bioorganic & Medicinal Chemistry Letter*. 23, 1365-69

Kumar D., Kapoor A., and Thangdurai A.(2011). Synthesis, anti-mycrobial evaluation and QSAR studies of 3-ethoxy-4-hydroxybenylidene-(4-nitrobenzylidene) hydrazides. *Chinese Chemistry Letters*. 22, 1293-96

Koopmans T.A. (1933). Ordering of wave functions and energies to the individual electrons of an atom. *Physica.* 1, 104–113

Lalloo, U.G., and Ambaram, A. (2010). New anti-tuberculosis drugs in development. *Current HIV/AIDS Reports*. 7 (3), 143-151

Lee C., Yang W., andParr R.G. (1988). Development of the Colle-Salvetti correlation energy formula into a functional of the electron density. *PhysicalReview*; 37:785–789.

.

Lees,A.W., Tyrrell W.F., Smith J., and Allan G.W., (1970). Ethambutol in the retreatment of chronic pulmonary tuberculosis. *British Journal of Disease of the Chest.* 64:85–89

Leonard, V.S., and Racheal, E B. (2009*).*Challenges, successes and hopes in the development of novel TB therapeutics. *Future Medicinal Chemistry. 1(4),749- 756*

Lescher, G.Y., Froelich E.J., Gruett M.D., Bailey J.H. and Brundage. R.P.(1962).*Journal of Medicinal and Pharmaceutical Chemistry*. 5, 1063

Long-term follow-up of a clinical trial of six-month and four-month regimens of chemotherapy in the treatment of pulmonary tuberculosis. Singapore Tuberculosis Service/British Medical Research Council (1983). *The American Review of Respiratory Disease* 133(5), 779-83

Ma, Z., Lienhardt C., McIlleron H., Nunn A. J., and Wang X., (2010). Global tuberculosis drug development pipeline: the need and the reality. *The Lancet*, 375 (9731), 2100-2109.

Manvar, A., Malde A., and Veraj (2008). Synthesis, anti-tubercular activity and 3D-QSAR study of coumarin-4-acetic acid benzylindene hydrazides. *European Journal of Medicinal Chemistry*. 43, (2) 395-403.

Manish C.B., Denise M., and Subhash C.B (2007). Quantitative structure-activity relationship (QSAR) studiesof quinolone antibacterials against M. fortuitum and M.smegmatis using theoretical molecular descriptors, *Journal of Molecular Modeling*, 13, 111–120.
DOI 10.1007/s00894-006-0133-z

Madhusudana, R. G., and Yellala, V.R.R. (2014). QSAR studies on 1, 3, 5-triazine-bisazomethines as potential anti-tubercular agents,*International Journal of Pharmaceutical and Bio-scienc.* 3(2),184193

Mandeep, S. J and William R. B (2010).Epidemiology and Challenges to the Elimination of Global Tuberculosis, *Clinical infectious disease*. 50 (3), 156-164

Manchester, K. (1984). Tuberculosis and leprosy in antiquity: an interpretation. *Medical History*, 28,162-173

Marshall G. (1948). Streptomycin treatment of pulmonary tuberculosis. *British Medical Journal* 2, 769–782.

Marshall G. (1955). Various combinations of isoniazid with streptomycin or with P.A.S. in the treatment of pulmonary tuberculosis; seventh report to the Medical Research Council by their Tuberculosis Chemotherapy Trials Committee. *British Medical Journal* (1), 435–445.

Mitchison D.A. (2005). The diagnosis and therapy of tuberculosis during the past 100 years. *American Journal of Respiratory and Critical Care Medicine.* 171:699–706.

Minovski,N.,Jezierska-Mazzarello,A., and Vracko, M. (2011). Investigation of 6-fluoroquinolones activity against Mycobacterium tuberculosis using theoreticalmolecular descriptors: a case study. *Central European Journal of Chemistry*. 9, 855doi: 10.2478/s11532-011-0071-1

Murray, J.F. (2004). A century of tuberculosis. *American Journal of Respiratory and Critical Care Medicine* 169:1181–1186

Neeraja, D., Bhartendu, N.M., and Vishwa, M.K., (2011). 2D-QSAR model development an analysis onvariant groups of anti -tuberculosis drugs.*Bioinformation* 7(2), 82-90

Norris S. and Mandell G.L (1988). The quinolones: History and Overview, The quinolones, San Diego, Academic Press Inc, pp 1

Organisation for Economic Co-operation and development. (2004). Report from the Expert Group on (Quantitative) Structure-Activity Relationships [(Q) SARs] on the Principles for the Validation of (Q) SARs. Retrieved from http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono/(2004)24&docclanguage=en

Palker, M.B., Noolvi M.N., and Patel H.M., (2011). 2D-QSAR study of Fluoroquinolone derivatives; an approach to design anti-tubercular agents. *International Journal ofDrug Design and Discovery*.3, 559-74

Parr, R.G., Szentpaly, L.V., and Liu, S. (1999) Electrophilicity index. *Journal of American. Chemical Society*. 121, 1922–1924

Parr,R.G., and Pearson, R.G.(1983). Absolute hardness: companion parameter to absolute electronegativity. *Journal of American Chemical Society*,105, 7512–7516

Patnode, R.A., and Hudgins, P.C. (1958). Effect of kanamycin on Mycobacterium tuberculosis in vitro. *American Review of Tuberculosis*. 78,138–139

Platts, J.A., Butina, D., Abraham, M.H., and Hersey A. (1999). Estimation of molecular linear free energy relation descriptors using a group contribution approach. *Journal of Chemical Information and Computer Sciences*. 39, 835–45

Ponnurengam, M.s., Sethu., K.G.B., and Doble M.(2007). QSAR studies on Chalcones and flavanoids asAnti-tuberculosis Agents using Genetic Function Approximation (GFA) method. *Chemical and Pharmaceutical Bulletin* 55(10), 44-9

Rajasekaran, S., Gopalkrishna R., and Sanjay P.N. (2011). 2D QSAR studies of some novel quinazolinone derivatives as antitubercular agents *JournalComputational Method in MolecularDesign.* 1(3): 69-82.

Ravichandran, V., Abhishek J, Shalini S., Christapher, P. V., and Ram K .A. (2011). Validation of QSAR Models - Strategies and Importance.*International Journal of Drug Design and Discovery* 2(3), 511- 519http://dx.doi.org/10.1128/AAC.39.4.964

Raviglione, M.C., andPio A. (2002). Evolution of WHO Policies for tuberculosis control, 1948-2001. *Lancet.* 359, 775-80

Raviglione, M.C, and Smith I.M. (2007). XDR tuberculosis – implications forglobal public health.

*New England Journal of Medicine*. 356, 656-9.

Roy, K. (2007). On Some Aspects of Validation of Predictive Quantitative Structure-Activity Relationship Models. *Expert Opinion in Drug Discovery.* 2, 1567-1577

Saito, H., Tomioka, H., Sato, K., and Dekio, S. (1994). *In vitro* and *in vivo* antimycobacterialactivities of a new quinolone, DU-6859a. *Antimicrobial Agents in Chemotherapy.*38, 2877–2882

Sakamot, K. (2012). The pathology of Mycobacterium tuberculosis infection.*Veterinary pathology.*49(3), 423-439

Schüürmann, G., Ralf-Uwe, E.J., Wang, C., and Kühne B .J. (2008). External Validation and Prediction Employing the PredictiveSquared Correlation Coefficient – Test set Activity Mean. *Chem. Inf. Model*. 48 (11): 2140

Sentthikumar, P., Dinakaran M., Yogeeswari, P., Sriram D., China A., and Nagaraja, V. (2009). Synthesis and antimycobacterial acticvities of novel 6-nitroquinolone-3-carboxylic acids.*European Journal ofMedicinal Chemistry*. 44(1), 345-58. doi: 10.1016/ejmech.2008.02.031

Sirgel, F. A., Venter, A., and Heilmann, H.-D. (1995). Comparative *in vitro* activity of BAY y 3118, a new quinolone, and ciprofloxacin against *Mycobacterium tuberculosis* and *Mycobacterium avium* complex. *Journal of. Antimicrobial. Chemotherapy.* 35, 349–351

Sofie, D. (2009). Quantum chemistry in QSAR Quantum chemical descriptors use, benefits and drawbacks. Ph.D. Thesis. Ghent University; Holland

Somner, A.R., and Brace A.A., (1962). Ethionamide, pyrazinamide and cycloserine used successfully in the treatment of chronic pulmonary tuberculosis. *Tubercle* 43, 345–360.

Sriram D., Palaniappan S., Murugesan D.,Yogesh C., and Perumal Y.(2009).Synthesis and in-vitro Antimycobacterial Evaluation of 1-(Cyclopropyl/2,4- difluorophenyl/tert-butyl)-1,4-dihydro-8-methyl-6-nitro-4-oxo-7-(substituted secondary amino)quinoline-3-carboxylic acids *Archive der Pharmazie*, 342,100112

Srivastava,A.K., Vinay K.P., Akanchha S., and Avni P. (2010). Molecular modeling of 8-methoxy quinolone analogues by using quantitative structure activity relationship. *Journal of Saudi Chemical Society*, 14, 217–222

Stead, W.W., Kathleen, E., Cave, M.D., and Bates, J.H. (1995). When did Mycobacterium tuberculosis infection first occur in the new world? An important question with publichealth implications. *American Journal of Respiratory and Critical Care Medicine* 151(4),1267-8

Suleiman, M.M., Isaev G.V., Klenina V.O., and OgurtsovV.V. (2014). Synthesis, biological activity evaluation and QSAR studies of novel 3-(aminooxalyl-amino)-and 3-(carbamoyl propionylamino)-2- phenylamino-benzoic acid derivatives. *Journal of Chemistryand Pharmaceutical Research, 6(5), 1219-1235*

Sunduru, N., Sharma M., Chauhan, P.M.S.(2010). Recent advances in the design and synthesis of Heterocycles as anti-tubercular agents. *Future Medicinal Chemistry*, 2(9), 1469-1500.

Swaan, P.W., and Ekins, S. (2005). Reengineering the pharmaceutical industry by crash-testing molecules. *Drug Discovery Today*. 10,1191-1200

Thoren, M., and Hinshaw, H.C. (1952). Therapy of pulmonary tuberculosis with isoniazid alone andin combination with streptomycin and with para-amino-salicylic acid. *Stanford MedicalBulletin*,10, 316–31

Topliss J. G.and Edwards R .P (1979).Chance factors in studies of quantitative structure-Activity relationship.*Journal of. Medicinal Chemistry.* 22 (10), 1238-1244

Todd M., Paul F.H., Douglas Y., and Alexander T. (2012). Does Rational Selection of Training and Test Sets Improve The Outcome of QSAR Modeling?*Journal of Chemical Information and Modeling.* 52(10), 2570-8. Doi:10.1021/ci300338w

Todeschini, R., and Consonni, V. (2000). Handbook of Molecular Descriptors. VCH Verlag GmbH, Weinheim, Garmany,doi: 10.1002/9783527613106

Ventura, C., and Martins, F. (2008).Application of quantitative structure-activity Hydrazide family. *Journal of Medicinal Chemistry*. 51, 612-24.

Vicente E., Duchowicz P.R., and Castro E.A., (2009). QSAR analysis for quinoxaline-2-carboxylate 1,4-di-N-Oxide as anti-mycobacterial agents. *Journal of Molecular Graphics and Modelling* .28;28-36

Virsodia V., Pissurlenkar R.S.S., and Manvar D., (2008). Synthesis, screening for anti-tubercular activity and 3D-QSAR studies of substituted N-Phenyl-6-methyl-2-oxo-4-phenyl-1,2,3,4-tetrahydro-pyrimidine-5-carboxamides. *European Journalof Medicinal Chemistry*. 43, 2103-15

Warren, H., and Sean, O. (2014).The ***Spartan'14*** Tutorial User's Guide. Retrieved from (www.wavefun.com)

Winkler, D.A. (2002). The role of quantitative structure-activity relationships (QSAR) in biomolecular discovery. *Briefings in Bioinformatics* 3(1),73-86

Wold, S., and Ericksson, L. (1995). Statistical validation of QSAR Results in: Van de Waterbeemed H.,(Ed.), Chemometric methods in Molecular design, VCH, Wienheim (Germany) ., pp. 309 – 318

World Health Organization (2011), Global Tuberculosis Report. Retrieved from http://apps.who.int/iris/bitstream/10665/44728/1/9789241364380_eng.pdf

World Health Organization (2013). Global Tuberculosis Report. Retrieved from http://apps.who.int/iris/bitstream/10665/91355/1/9789241564656_eng.pdf

World Health Organization. Tuberculosis Fact sheet. Geneva: (2014).Retrieved from http://apps.who.int/iris/bitstream/10665/137094/1/9789241564809_eng.pdf

World Health Organization. (1991). Forty-fourth World Health Assembly. Geneva: World Health Organization. Retrieved from http://www.who,int/tbpublication/1991/en/

Yap C.W (2011). PaDel –Descriptor: An open source software to calculate moleculardescriptors and fingerprints. *Journal of Computational Chemistry*. 32 (7), 1466-1474

Yew, W. W., Piddock, L. J. V., Li, M. S. K., Lyon, D., Chan, C. Y., and Cheng, A. F. B. (1994). *In vitro* activity of quinolones and macrolides against mycobacteria. *Journal ofAntimicrobial Chemotherapy.* 34, 343–351

Youmans G.P., Raleigh G.W., Youmans A.S. (1947). The tuberculostatic action of para- aminosalicylic acid. *Journal of Bacteriology* 54,409–416.