QUANTUM MODELLING OF SOME POTENT, NON-TOXIC ANTIMALARIAL COMPOUNDS

BY

ZAKARI YA'U IBRAHIM

B.Sc CHEMISTRY (ABU Zaria) 2008

M.Sc/SCI/6099/2011- 2012

A DISSERTATION SUBMITTED TO THE SCHOOL OF POSTGRADUATE STUDIES,

AHMADU BELLO UNIVERSITY, ZARIA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF A MASTER OF SCIENCE DEGREE IN PHYSICAL CHEMISTRY

DEPARTMENT OF CHEMISTRY,

FACULTY OF PHYSICAL SCIENCE,

AHMADU BELLO UNIVERSITY,

ZARIA, NIGERIA.

DECEMBER, 2016

DECLARATION

I declare that the work in this dissertation entitled "**Quantum modelling of some potent, non-toxic antimalarial compounds**" has been carried out by me in the Department of Chemistry of Ahmadu Bello University, Zaria, under the supervision of Prof. A. Uzairu and Dr. S. E. Abechi. The information obtained from literature have been duly acknowledged in the text and a list of references provided. No part of this dissertation was previously presented for the award of another degree and/ or diploma by me or anyone else at this or any other institution.

Zakari Ya'u Ibrahim _____ _____

Signature Date

CERTIFICATION

This dissertation, entitled"**QUANTUM MODELLING OF SOME POTENT, NON- TOXIC ANTIMALARIAL COMPOUNDS**",an original piece undertaken by Zakari Ya'u IBRAHIM, meets the regulations governing the award of the degree of Master of Science in Physical Chemistry of Ahmadu Bello University, Zaria, and is approved for its contribution to knowledge and literary presentation.

Prof. A. Uzairu

Chairman, Supervisory Committee	Signature	Date

Dr. S. E. Abechi

Member, Supervisory Committee	Signature	Date

Prof. A. O. Oyewale

Head of DepartmentSignature	Date	

Prof. K. Bala

Dean, School of Postgraduate Studies	Signature	Date

ACKNOWLEDGEMENTS

All praises, gratitude and adoration are to almighty Allah, the most knowledgeable, the Wisest, the Creator and Sustainer of the Universe for keeping me healthy during this struggle. May the peace and blessing of Allah be upon His Prophet and Messenger, Muhammad (SAW), the entire member of his household and those who follow him on the path of righteousness.

My first gratitude and appreciation goes to my supervisors, Prof. A. Uzairu and Dr. S. E. Abechi, for their enormous supports and encouragement during the course of this work. I am very grateful for all their advice, motivations and supports. I thank them very much.

I can never forget the roles played by my parents, Mr and Mrs Ibrahim D. Muhammed, my siblings for all the support and understanding they gave me.

I would like to express my gratitude to my colleagues, Mr David E. Author, Eric Mudima and Abduljalil Ajala, for their contribution to the accomplishment of this work.

Finally, I must not forget to mention the contribution of my wife, Malama Zainab S. Zakariyya, who has encouraged me all the way and whose encouragement has made sure that I give all it takes to finish that which I have started. To my children, Zakiyya and Amina who have been affected in every way possible by this quest. My love for you all can never be quantified.

DEDICATION

I dedicate this dissertation in loving memory of my son, Ibrahim Dio Zakariyya.

ABSTRACT

The Quantitative structure-activity relationship (QSAR) analyses were carried out for series of 45 naphthylisoquinoline and 22 bisbenzylisoquinoline alkaloids to find out the structural requirements of their antimalarial activities as well as the cytotoxicity using molecular descriptors. All the structures before the calculation of the descriptors were fully optimised in Spartan 14 v.1.1.0 using density functional B3LYP and the standard Pople's basis set of 6-311G*. The combination of the Genetic algorithm (GA) and Multiple Linear Regression (MLR) analyses methods were applied to derive QSAR models from the descriptors. For the selection of the best descriptors, the elimination selection stepwise regression method was utilised. The statistically significant models with parameters; R^2 = 0.7483, Q^2 = 0.6042 for activity of naphthylisoquinoline and R^2 = 0.6256, Q^2 = 0.7865 for the cytotoxicity were obtained. Furthermore, the models for activity and cytotoxicity of bisbenzylisoquinoline give R^2 = 0.8845, Q^2 = 0.7942 and R^2 = 0.8742, Q^2 = 0.7115 statistical parameters, respectively. The activity models generated, revealed descriptors BCUTc-1l (encoding connectivity information and atomic properties of the molecule), $SssCH_2$ (the electro-topological state indices for number of $–CH_2$ group connected with two double single bonds), minHsOH (electrotopological state atom type descriptor), Wlambda2.unity (encoding molecular 3D information regarding molecular size, shape, symmetry), μ, dipole moment and logP, lipophilicity, to be responsible for the activities of naphthylisoquinoline, while from the cytotoxicity models, MDEC-33 descriptor, which stand for molecular distance edge between all tertiary and quaternary carbons, determined the cytotoxicity. In the bisbenzylisoquinoline alkaloids, the descriptors responsible for the activities are BCUTw-1h (encode atomic properties relevant to intermolecular interactions), CrippenlogP, lipophilicity and maxssO (electrotopological state atom type descriptor) while ATSc4 (autocorrelation

descriptor, weighted by charges), VCH-6 (describing the overall topology of the molecules) and SHBint9 (strength for potential hydrogen bonds of path length 9) descriptors causes cytotoxicity. The accuracy of the proposed MLR models were illustrated using the following evaluation techniques: cross validation, validation through an external test sets, and Y-randomisation. Furthermore, the domain of applicability which indicates the area of reliable predictions was defined. The good correlation between experimental and predicted biological activity/cytotoxicity for the four selected models further highlights the reliability of the constructed QSAR models.

TABLE OF CONTENTS

CONTENT PAGE

Cover page i

Title pageii

Declarationiii

Certificationiv

Acknowledgementsv

Dedicationvi

Abstractvii

Table of contents

List of Tables xii

List of Figures xiv

CHAPTER ONE

1.0 INTRODUCTION1

1.1 Background of the Study1

1.2 The Research Problem2

1.3 Justification of the Study2

1.4 Research Questions 3

1.5 Aim and Objectives of the Study4

1.6 Research Hypotheses 5

1.7 Significance of the Study5

1.8 Scope and Limitation of the Study5

CHAPTER TWO

2.0 LITERATURE REVIEW 7

2.1 Malaria7

2.2 Approaches to Antimalarial Drug Discovery 9

2.2.1 Optimisation of therapy with available drugs 9

2.2.2 Development of analogues of existing agents 10

2.2.3 Discovery of natural antimalarial products.10

2.2.4 Investigation of compounds that were originally developed to treat other diseases 10

2.2.5 Evaluation of drug resistance reversers 11

2.2.6 Exploiting compounds active against new targets. 12

2.3 Naphthylisoquinoline Alkaloids12

2.4 Bisbenzylisoquinoline Alkaloids13

2.5 Theoretical Framework of the Study 14

2.6 Molecular Descriptors16

2.7 Quantitative-Structure Activity/Toxicity Relationship. 17

CHAPTER THREE

3.0 MATERIALS AND METHODS21

3.1 Materials21

3.1.1Instrumentation 21

3.2 Methodology 21

3.2.1 Biological data35

3.2.2 Molecular structure sketching 35

3.2.3 Geometry optimisation and calculation of some quantum-chemical descriptors39

3.2.4 Theoretical background for calculation of the quantum descriptors 44

3.2.5 Calculation of 0- 3D descriptors from PaDEL. 48

3.2.6 Correlation matrix 54

3.2.7 Variable elimination and selection 64

3.2.8 Data standardisations 71

3.2.9 Division of data into training and test sets71

3.2.10 Development of a quantitative structure- activity relationship model 74

3.2.11 Descriptor selection and model development 75

3.2.12 Model validation 84

3.2.12.1 Internal validation84

3.2.12.2 External validation *86*

3.2.13 Y-Randomisation 95

3.2.14 Applicability domain 103

3.2.15 Generation of the Williams plot 105

CHAPTER FOUR

4.0 RESULTS 124

4.1 Results of Optimisation and Calculation of Quantum-chemical Descriptors 125

4.2 Results of the Calculation of 0D, 1D, 2D and 3D Descriptors using Padel 125

4.3 Results of the Variable Elimination and Selection126

4.4 Results of Antimalarial Activity (pIC_{50}) of Naphthylisoquinoline Derivatives Against K1 Strain of *Plasmodium falciparum*126

4.4.1 Evaluation of model 4.4.4 135

4.5 Results of Antimalarial Activity (pIC_{50}) of Bisbenzylisoquinoline Derivatives Against K1 Strain of *Plasmodium falciparum*136

4.5.1 Evaluation of model 4.5.1 137

4.6 Results of Cytotoxicity (pIC_{50}) Study of Naphthylisoquinoline Derivatives Against L6 Cell Line. 144

4.6.1 Evaluation of model 4.6.1 145

4.7 Results of Cytotoxicity (pIC_{50}) Study of Bisbenzylisoquinoline Derivatives Against KB Cell Line. 145

4.7.1 Evaluation of model 4.7.3 153

CHAPTER FIVE

5.0 DISCUSSION 161

5.1 Interpretation of the Descriptors in Model 4.4.2 161

5.2 Interpretation of the Descriptors in Model 4.5.1 163

5.3 Interpretation of the Descriptors in Model 4.6.1 164

5.4 Interpretation of the Descriptors in Model 4.7.1 164

CHAPTER SIX

6.0 SUMMARY, CONCLUSION AND RECOMMENDATION 167

6.1 Summary167

6.2 Conclusion 167

6.3 Recommendation168

References 169

List of Tables

Table Page

Table 2.1Representative of the molecular descriptors 18

Table 3.1Series of naphthylisoquinoline alkaloids with their antiplasmodial activity against K1 strain as well as their cytotoxicity against L6 cell line. 22

Table 3.2 Series of Bisbenzylisoquinoline Alkaloids with their antiplasmodial activity against K1 strain as well as their Cytotoxicity against KB cell line 30

Table 3.3 Format of training set compounds and their activities in the data folder of MLR plusValidation 97

Table 4.1 Ravichandran *et al*., parameters for the activity models of naphthylisoquinoline alkaloids.128

Table 4.2 Comparison of Observed Activity to Predicted Activity of naphthylisoquinoline alkaloids models 129

Table 4.3 Y-randomisation result of QSAR analysis of the activity of naphthylisoquinoline derivatives. 132

Table 4.4 Correlation matrix of the descriptors selected for the activity of naphthylisoquinoline derivatives model generation134

Table 4.5 Ravichandran *et al*., parameters for the activities models of bisbenzylisoquinoline alkaloids138

Table 4.6 Comparison of Observed Activity to Predicted Activity of bisbenzylisoquinoline alkaloids models 139

Table 4.7 Y-randomisation result of QSAR analysis of the activity of bisbenzylisoquinoline derivatives. 141

Table 4.8 Correlation matrix of the descriptors selected for the activity of bisbenzylisoquinoline derivatives model generation 143

Table 4.9 Ravichandran *et al*., parameters for the cytotoxicity models of naphthylisoquinoline alkaloids146

Table 4.10 Comparison of Observed Toxicity to Predicted Toxicity of naphthylisoquinoline alkaloids models 147

Table 4.11 Y-randomisation result of QSAR analysis of the cytotoxicity of naphthylisoquinoline derivatives 150

Table 4.12 Correlation matrix of the descriptors selected for the cytotoxicity of bisbenzylisoquinoline derivatives model generation152

Table 4.13 Ravichandran *et al*., parameters for the cytotoxicity models of bisbenzylisoquinoline alkaloids155

Table 4.14 Comparison of Observed to Predicted cytotoxicity of bisbenzylisoquinoline alkaloids models 156

Table 4.15 Y-randomisation result of QSAR analysis of the cytotoxicity of bisbenzylisoquinoline derivatives 158

Table 4.16 Correlation matrix of the descriptors selected for the cytotoxicity of bisbenzylisoquinoline derivatives model generation 160

List of Figures

Figure Page

Figure 3.1: Workflow for themodel development34

Figure 3.2: A chemdraw worksheet36

Figure 3.3: A highlighted molecule on Chemdraw 38

Figure 3.4 Spartan graphic user interface 40

Figure 3.5 File pull down menu on Spartan 41

Figure 3.6 3D display of molecule in Spartan 42

Figure 3.7 Spartan minimising molecule 43

Figure 3.8Spartan setup menu display45

Figure 3.9 Spartan calculations dialogue46

Figure 3.10 File browser 47

Figure 3.11 A folder of PaDEL-Descriptor jar 49

Figure 3.12 PaDEL home page 50

Figure 3.13 PaDEL file browser51

Figure 3.14 PaDEL file output 52

Figure 3.15 PaDEL 1-3D molecular descriptors 53

Figure 3.16 Welcome to material studio dialogue55

Figure 3.17 A folder of PaDEL-Descriptor jar 56

Figure 3.18 Material studio blank sheet 57

Figure 3.19 Study table dialogue58

Figure 3.20 Material studio work sheet 59

Figure 3.21 Material studio study table 60

Figure 3.22 Initial analysis dialoguebox 61

Figure 3.23 Correlation matrix dialoguebox 62

Figure 3.24 Correlation matrix 63

Figure 3.25 QSAR statistical software 65

Figure 3.26 Data pretreatment dialogue 66

Figure 3.27 NEW_v-WSP dialogue67
Figure 3.28 Queries window 68
Figure 3.29 A choose file dialogue69
Figure 3.30 A save dialogue70
Figure 3.31 A V-WSP dialogue72
Figure 3.32Standardisedata dialogue73
Figure 3.33 BuildQsar worksheet 77
Figure 3.34 A file dialogue78
Figure 3.35 Create new data dialogue79
Figure 3.36 BuildQsar raw data 80
Figure 3.37 BuildQSAR variable selection dialogue81
Figure 3.38 BuildQsar tools bar82
Figure 3.39 Genetic search dialogue83
Figure 3.40 Variable selection results dialogue85
Figure 3.41 Training set working sheet 87
Figure 3.42 Statistics menu brought down 88
Figure 3.43 Model building menu brought down 89
Figure 3.44 Multiple linear regression dialogue90
Figure 3.45 Stepping of multiple linear regression dialogue 91
Figure 3.46 Validation of multiple linear regression dialogue 92
Figure 3.47 Model internal validation result dialogue 93
Figure 3.48 MLRplus validation folder 96
Figure 3.49 MLR plus validation dialogue98
Figure 3.50 Training set file dialogue99
Figure 3.51 Test set file dialogue 100
Figure 3.52 Save file dialogue 101
Figure 3.53 Result summary pad 102
Figure 3.54 Excel Worksheet-A 106

Figure 3.55 A descriptor Matrix 107

Figure 3.56 A paste option pull down menu 108

Figure 3.57 Transpose matrix 109

Figure 3.58 Multiplication worksheet 111

Figure 3.59 Multiplication of transpose and the original matrix 112

Figure 3.60 Inverse pull down dialogue 113

Figure 3.61 Inverse worksheet 114

Figure 3.62 Inverse of a descriptor matrix 115

Figure 3.63 Extracted diagonal element of hat matrix 116

Figure 3.64 Highlighted residual and leverage values 118

Figure 3.65 A scatter pull down 119

Figure 3.66 A plot of residual against the leverage values 120

Figure 3.67 A pull down menu from axis 121

Figure 3.68 Format axis dialogue 122

Figure 3.69Generated Williams plot 123

Figure 4.1 Plots of the observed and predicted Log $[IC_{50}]^{-1}$ values for the training and test sets for activities of naphthylisoquinoline alkaloids 131

Figure 4.2 Williams plot for an externally validated for activities of naphthylisoquinoline alkaloids. Cut-off value: h*= 0.618 133

Figure 4.3 Plots of the observed and predicted Log $[IC_{50}]^{-1}$ values for the training and test sets for activities of bisbenzylisoquinoline alkaloids 140

Figure 4.4 Williams plot for an externally validated for activities of bisbenzylisoquinoline alkaloids. Cut-off value h*, = 0.75 142

Figure 4.5 Plots of the observed and predicted Log $[IC_{50}]^{-1}$ values for the training and test sets for cytotoxicity of naphthylisoquinoline alkaloids 149

Figure 4.6 Williams plot for an externally validation for cytotoxicity of naphthylisoquinoline alkaloids. Cut-off value h*= 0.176 151

Figure 4.7 Plots of the observed and predicted Log $[IC_{50}]^{-1}$ values for the training and test sets for cytotoxicity of bisbenzylisoquinoline alkaloids 157

Figure 4.8 Williams Plot for an externally validated for cytotoxicity of bisbenzylisoquinoline alkaloids. Cut-off value: h* = 0.92 159

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background of the Study

Malaria remains one of the most calamitous diseases of all the disease causing parasites. It is the most dangerous and destructive infectious agent in the developing world (Greenwood *et al.,* 2005 and Winter *et al.,* 2006). Malaria is usually believed to be caused by poverty, but it is also a cause of poverty and a major factor militating against economic development. An estimated case of malaria is about 247 million with 3.3 billion people at risk in 2006, causing deaths in approximately millions, with children under the ages of 5 years the most affected. It is widespread in tropical and subtropical regions, including parts of the Americas, Asia, and Africa. About 109 countries were estimated to be endemic for malaria in 2008, out of which 45 are in the WHO African region (WHO, 2008).

Malaria is caused by a parasite called *plasmodium* that is transmitted from one human to another by the bite of infected *Anopheles* mosquitoes. Although four species of the genus *Plasmodium* cause human malaria, *Plasmodium falciparum* is the deadliest. The traditional remedies are no longer effective and the incidence of malaria by*P.falciparum*, the most dangerous species of parasite,*Plasmodium,* continues to grow, while some traditional drugs such as chloroquine and its congeners are losing their activity due to multi drug resistance. New antimalarial drugs are urgently needed. Not only should these drugs be efficacious against *P.falciparum* strains, but to ensure good compliance, they should provide a cure within a reasonable length of time, they should be safe and low cost, and they should be available in an appropriate formulation for oral use. Natural antiplasmodial compounds such as naphthylisoquinoline and bisbenzylisoquinoline alkaloids exist as cited by Mambu and Grellier(2008). These compounds were found to possess manifold activities against the genius

plasmodium. Antimalarial drug discovery is achieved through different approaches, all of which take into account the cost of drug discovery and the cost of drug itself. Quantitative structure-activity relationship (QSAR) methodologies save resources and expedite the process of the development of new molecules and drugs andhence play an important role to minimise trial and error in designing new antimalarial drugs.

1.2 The Research Problem

In 2012, about630,000 people died from malaria (WHO,2013), with pregnant women and children under the age of five being the most vulnerable to the infection (Shetty, 2012). The parasite developed resistance to a number of antimalarial drugs such as chloroquine and its derivatives which is the most widely used treatment drug for malaria. Many countries now rely on therapies that are expensive to slow down the development of resistance.

Various antimalarial drugs currently in use are chemically related or have a similar mode of action, thus increasing the risk of cross-resistance and clinical failure of newly introduced therapies. There is, therefore, the need for the design of a more potent drug for malaria.

1.3 Justification of the Study

During the 1990s, child deaths caused by malaria increased by up to two-folds in some parts of sub-Saharan Africa. The disease also re-emerged in several countries in central Asia, Eastern Europe and South-East Asia. The majority of the cases in 2006 were estimated as Africa (86%), followed by South-East Asia (9%) and3% in Eastern Mediterranean region (Meneguzzi*et al*., 2009).Although malaria is a curable and preventable disease, its prevalence increased in the 1980s and 1990s as the parasites

developed resistance to the most frequently used antimalarial drugs and the vectors became resistant to insecticides.

Increasing resistance of *P. falciparum*against traditional antimalarialdrugs like chloroquine or the antifolates makes the developmentof new, highly efficient, and easily available agents anurgent task.Nowadays, drug development is too expensive to be guided by trial anderror. Quantitative structure-activity relationship (QSAR), molecular modelling, and protein crystallography are important andvaluable tools in computer-assisted drug design. In view of the above fact,QSARhas evolved as excellent optimisation technique which involves the derivation of a mathematical formula, which relates the biological activities of a group of compounds to their physico-chemical properties.

1.4 Research Questions

The questions that this study tried to answer were

i. Are there correlations between the molecular descriptors and the activity/cytotoxicity of naphthylisoquinoline derivatives?
ii. What structural features of the molecules are responsible for the activity/cytotoxicity of naphthylisoquinoline derivatives?
iii. Are there correlations between the molecular descriptors and the activity/cytotoxicity of bisbenzylisoquinoline derivatives?
iv. What structural features of the molecules are responsible for the activity/cytotoxicity of bisbenzylisoquinoline derivatives?
v. Will the generated models be used to predict the activity/cytotoxicity of all the derivatives of naphthylisoquinoline/bisbenzylisoquinoline alkaloids?

1.5 Aim and Objectives of the Study

The aim of this research was to carry out QSAR analysis and develop predictive QSAR models that could explain the antimalarial activities of naphthylisoquinoline and bisbenzylisoquinoline alkaloids using theoretical molecular descriptors.

The aim of the research will be achieved through the following objectives:

i. Collection of naphthylisoquinoline and bisbenzylisoquinoline compounds with their activity and cytotoxicity values from published articles.
ii. Optimisation of these compounds so that equilibrium structures of the compounds with lower energy will be obtained for descriptors calculation.
iii. Determination of molecular descriptors using PaDEL software.
iv. Treatment of the calculated descriptors by correlating them with the experimental activity/cytotoxicity thereby removing the not so informative descriptors and also removing descriptors that are highly correlated with another descriptor (s) but less correlated with the activity/cytotoxicity.
v. Division of the data set into training and test sets. The training set is for constructing the model while the test set is for validation of the model.
vi. Performing a multiple linear regression (MLR) analysis using BuildQSAR modelling software.
vii. Validation of the predictive model using external method of validation such as calculating Y-randomisation and $R^2_{\text{pred.}}$
viii. Checking the applicability domain of the model which serves as a kind of quality control procedure in the prediction of new compounds.

1.6 Research Hypothesis

The working hypotheses of the study are:

a) The null hypothesis (H_o) which states that:

i. There is no correlation between the antimalarial activity/cytotoxicity of naphthylisoquinolines derivatives and their physio-chemical properties.

ii. There is no correlation between the antimalarial activity/cytotoxicity of bisbenzylisoquinoline alkaloids derivatives and their physico-chemical properties.

b) The alternate hypothesis (H_1), which states that:

i. There is positive correlation between the antimalarial activity/cytotoxicity of naphthylisoquinolines derivatives and their physicochemical properties.

ii. There is positive correlation between the antimalarial activity/cytotoxicity ofbisbenzylisoquinoline alkaloids derivatives and their physico-chemical properties.

1.7 Significance of the Study

The study would avail us with the various structural information responsible for the observed antimalarial activity/cytotoxicity of naphthylisoquinoline and bisbenzylisoquinoline derivatives which can be exploited by the synthetic chemist to develop antimalarial drugs with better activity/cytotoxicity.

1.8 Scope and Limitation of the Study

The study was aimed at quantitatively determined the relationship between antimalarial activity/cytotoxicity with physico-chemical parameters of antimalarial compounds. There are concerns over limited availability of data set. Those that are available may be uneven and may contain outliers. Due to financial constraint and other factors, the activities of these compounds will not be determined *in vivo* on *P. falciparum* and also there will be no experimental validation.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Malaria

Malaria remains one of the most devastating diseases, causing as many as 2.7 million deaths annually with an estimated 400 to 900 million new cases each year (Breman, 2001). Areas that are mostly affected by malaria include Africa, South East Asia, India and South America; however, the disease is spreading to new areas, such as central Asia, and Eastern Europe. Local transmission of malaria in the United States, unheard of in the era between World War II and 1980, now accounts for an increasing number of cases (Mundy*et al*., 1996). Clinical cases in the US are now in average of 1,300 per year (Filler *et al*., 2003).

Clinical manifestations includes headache, periodically recurrent fever, chills, myalgia, sudoresis, prostration and the presence of high anaemia in cases of severe malaria in children and pregnant women. All over the world, the majority of deaths occur in children, other high risk groups include pregnant women, refugees, migrant workers and non-immune travellers-over 20 million western tourists at risk annually.

Malaria is caused by protozoans that belong to the genius plasmodium, phylum apicomplexa and family plasmodidae. Four species of malaria parasites are pathogenic to humans: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*and *Plasmodium malariae*. *P.ovale*seems to be limited to sub-Saharan Africa and some Islands of the western pacific, whereas *P.falciparum* is prevalent in endemic malarial countries, such as Brazil. *P.falciparum* is the agent of severe and potentially fatal human malaria (Krettli *et al*, 2001; Brazil, 2006 and Brazil, 2007).Severe *P.falciparum* human malaria can

include neurological symptoms such as delirium and convulsions, metabolic acidosis, multi-organ system failure which can lead to coma and even death if not properly treated (Fidock *et al*., 2004; Brazil, 2006).

The treatment of malaria aims to interrupt the blood schizogony that cause the pathogenesis and clinical symptoms of the infection. Prompt and effective treatment is probably the most cost-effective method of malaria control. Oral treatments prevent progression to a severe state of the disease and the resultant complications. If the drugs are administered effectively, a decrease in overall malaria-related morbidity and mortality may be achieved. However most people living in endemic areas have little or no access to diagnosis and treatment. Furthermore, the treatment is commonly inadequate due to lack of availability of quality-assured, effective drugs (Guerin*et al*., 2002; Fidock *et al.,* 2004). Most of the antimalaria drugs that are currently in used belongs to the classes of aminoguinolines (chloroquine, amodiaquine, primaguine), guimolinomethanol derivatives (guanine, mefloquine, halofantrine), diaminopyrimidines (pyrimethamine), sulfanamides (sulfadoxine, sulfa-diazine), biguamides (proguanil and derivatives), antibiotics (tetracyclines, doxycydine, clindamycin), sesquiterpenes (artemisimin, dihydroartemisinin, arteether, arteneter, artesunate) and naphtoquinones (atoraquone) (Guerin *et al.*, 2002).

The growing resistance of existing antimalaria drugs is a factor militating against eliminating this deadly disease. Fortunately, there are still very few drugs that are active against malaria. (artemisium, atovaguone, and chloroquine analogues), (Fidock, 2010 and Wells *et al.,* 2009). However, the emergence of chloroquine resistant malaria parasite strains has prompted the research for alternative strategies to combat the disease. The clones of *P.falciparum* were used most often for in vitro testing of the antimalarial activity on different strains, among which are chloroquine-sensitive (NF54, NF54/64, 3D7, D6, F32, D10, HB3, FCC1-HN, Ghana), Chloroquine-resistant (FCB1, W2, FCM29, BHz 26/86, Dd2, EN36, ENT30, FCR3, FCR-3/A2), and/or multidrug-resistant (K1, TM91C235) strains to find effective compounds

against resistant malaria. (Najafi*et al.,*2013). At present, drug resistance to malaria parasite is widespread, and no new chemical class of antimalarials has been introduced into clinical practice since 1996.There has recently been an increase in parasite strains with reduced sensitivity to the newest drug (Gamo *et al.,* 2010). Meanwhile, recent advances in genome-based technologies and in vitro screening of whole parasites and a great number of compounds (natural and synthetic) have broadened the range of therapeutic targets and are accelerating the development of a new generation of treatment for both the control and eradication of malaria (Wells *et al*., 2009) .

2.2 Approaches to Antimalarial Drug Discovery

Regardless of the approach employed in drug discovery, it is essential to take into account specific concerns, such as the need to limit the cost of drug discovery and the cost of the drug itself. Several approaches to antimalarial drug discovery are the following:

2.2.1 Optimisation of therapy with available drugs

This involves optimisation of therapy with existing agents, through provision of additive or ideally synergistic antiparasitic activity. In addition to this, the use of combinational therapy slows the progression of resistance to the new agents. Preferably, a combination regimen that will prevent resistance development should include at least two agents against which parasite resistance has not yet developed and which have similar pharmacokinetics, so that low blood levels of a single agent will not be present. For Africa and other areas where there is high drug resistance, effort is ongoing to develop a combinational therapy which include new combination of older agents (chlorproguanil/dapsone) and even newer agents (e.g. artemisinin derivatives).

2.2.2 Development of analogues of existing agents

With this approach, analogues of the existing antimalarial are improved upon. Many antimalarial such as chloroquine, primaquine, etc. were developed through chemical strategies to improve upon quinine (Stocks *et al*., 2001). For instance,4-aminoquinolines that are closely related to chloroquine, appearto offer the antimalarial potency of the parent drug, evenagainst chloroquine-resistant parasites (Kaschula *et al*., 2002 andRaynes *et al*., 1999). Pyronaridine was developed in China and is now undergoing extensive clinical trials in other areas (Ringwald *et al*., 1996).

2.2.3 Discovery of natural antimalarial products

Another approach to antimalarial chemotherapy is the compounds derived from natural products. This approach makes use of the available knowledge of medicinal plants among natives of malarious regions, where the appreciation of the use of plant products to treat febrile illness has grown over many generations. Hence, as a great improvement over random screening, plant products having specific clinical activity can be the starting point for a medicinal chemistry effort. Two most important drugs currently available to treat severe *falciparum* malaria, quinine and derivatives of artemisinin, have source rooted to natural product. In the case of artemisinin, relatively simple chemical modifications of the natural product parent compound have led to aseries of highly potent antimalarials that are playing an increasingly important role in the treatment of malaria (Meshnick, 2001).

2.2.4 Investigation of compounds that were originally developed to treat other diseases

A fourth approach to antimalarial chemotherapy is identifying agents that are developed for treatment of other diseases. The agents may act using different mechanisms against the malarial

parasite. Given the cost of antimalarial drug discovery, the compounds have advantage in the fact that they have been developed for human indicator and as such will be inexpensive to develop as antimalarials. For example, Folate antagonists, tetracyclines and other antibiotics weredeveloped for their antibacterial properties and were laterfound to be active against malaria parasites (Clough andWilson, 2001). Atovaquone was initially identified as anantimalarial, but its development was expedited by thediscovery of its activity against *Pneumocystis*. More recently,its potential as an antimalarial (as a component of thecombination drug Malarone) has been re-explored, and it wasfound to have marked antimalarial synergy with proguanil(Canfield *et al*., 1995). These and many more examples suggest that new antimalarial drugs can be screened. The approach makes available large assay for potential antimalarial.

2.2.5 Evaluation of drug resistance reversers

In drug resistance approach, previous effective agents are combined with compounds that reverse parasites resistance to these agents. Many drugs have been shown to reverse the resistance of *P. falciparum* to chloroquine *in vitro*, most notably the antihypertensive verapamil (Martin *et al*., 1987) and the antidepressant desipramine (Bitonti *et al*., 1988). In general, for the resistance reversal to be effective, unacceptable high concentration of the agents are required, but two or more combination of these agent at pharmacological concentrations may provide clinically relevant resistance reversal, as suggested by studies with Verapamil, desipramine and trifluoperazine (Van Schalkwyk *et al*., 2001).

2.2.6 Exploiting compounds active against new targets

New targets identification and subsequent discovery of compounds that act on these targets is the most innovative approach to chemotherapy. Progress towards the characterisation of the biology of malaria parasites has been stimulated by the development of technology to disrupt *plasmodial*genes. The readily accessible genome sequence facilitates genomic approaches to drug discovery, although the more difficult and risky biological and parasitological validation of putative drug targets remains essential and typically limits progress. Among the classes of drugs that are effective in the treatment of the *P.falciparum* malarial are flavonoid, diterpenes, bisbenzylisoquinoline and naphthylisoquinoliesalkaloids.

2.3 Naphthylisoquinoline Alkaloids

The naphthylisoquinoline alkaloids (NIQs) represent a class of natural products isolated from rare and difficult-to-cultivate tropical plants. They have manifold activities against various tropical diseases.Among them are the antimalarial compoundsdioncophylline C (**1**) and dioncopeltine A (**2**) (François *et al*., 1997), andancistrotanzanine B (**3**), which is active against *Leishmania donovani*, the pathogen causing leishmaniasis (Bringmann *et al*., 2003; 2000).Theseacetogenic isoquinoline alkaloids (Bringmann *et al*., 2000) with their characteristicstructures involving axial chirality and stereogenic centers,occur exclusively in two families of tropical plants, viz. the Dioncophyllaceaeand the Ancistrocladaceae (Bringmann *et al*., 1995, 1998).Some C,C-linkednaphthylisoquinolines, among them e.g. dioncophylline C (1,0.01 mM) and dioncopeltine A (2, 0.02 mM), show interesting *in vitro*and even *in vivo* activities against *P. falciparum*(François *et al*., 1997; 1994 and Wiesner *et al*., 2003).These andother promising bioactivities make the naphthylisoquinolinealkaloids suitable pharmaceutical lead structures for the synthesis of new potent agents (Bringmann *et al*., 2001 and Hallock *et al*., 1994).

dioncophylline C (1) dioncopeltine A (2) ancistrotanzanine B (3)

2.4 Bisbenzylisoquinoline Alkaloids

Bisbenzylisoquinoline (BBIQ) alkaloids areextracted from thewood of a menispermaceous climber, *Triclisia patens*(Dwuma-Badu *et al*., 1975 and Tackie *et al*., 1974). Screening of the related species,*Triclisia dictyophylla,*revealed that this plant also has potent antiplasmodialproperties *in vitro*. Constituent alkaloids, suchas phaeanthine, pycnamine, aromoline, isotetrandrine,obamegine, and gilletine, from both species showed antiplasmodial activity *in vitro* (Marshall, 1991; Partridge *et al*., 1988 and Partridge *et al*., 1989), and it is thought thatthese alkaloids are responsible, at least in part, for theantimalarial activity of these two menispermaceous species.Independent investigations into the antimalarial activity of*Tiliacora triandra* (Menispermaceae) from Thailand haveshown that the BBIQ alkaloids tiliacorine, tiliacorinine, andnortiliacorinine A are the active components of plant which are used in traditional medicine (Pavanand *et al*., 1981). The antimalarialactivity of eight BBIQ alkaloids is the subject of a patent (Van Dyke, 1991).Plants containing BBIQ alkaloids are also used intraditional medicine for the treatment of a number of diseases, including amoebic dysentery, leishmaniasis, bacterial infections, and cancer (Fournet *et al*., 1988; Kuroda *et al*., 1976; Liu and Xiao, 1983; Marini-

Bellolo, 1980 and Oliver-Bever, 1986). The BBIQ alkaloid tetrandrinehas been used clinically in China for the treatment ofamoebic dysentery, lung cancer, and leukemia (Liu and Xiao 1983).

2.5 Theoretical Framework of the Study

Many important properties of matter are dependent on the microscopic properties of the molecules from which it is made up. This is particularly true when bio-molecular properties are considered (David, 2002). Empirical methods for building predictive models of the relationships between molecular structures and useful properties are becoming increasingly important. These have arisen because drug discovery and development have become more complex (David, 2002). As drug leads progress down the development pipeline, the ability to predict physicochemical properties of these leads is becoming increasingly important in reducing the number of expensive, late development failures. Quantitative structure-activity relationship (QSAR) has much to offer in these areas (David, 2002). The QSAR modelling is based on the idea that all the information related to a molecule can be derived from its chemical nature by means of parameters that encode or describe different molecular features, and these parameters, or descriptors, can be correlated to a particular chemical or biochemical activity, i.e.

Activity = F (structure) = F (·) F (·) = g {$F_E[F_R(\cdot)]$}

Here, $F_R$is the molecular representation; $F_E$contains the structural encoding descriptors; and *g* represents the mapping function. The input domain of Fis then the global information that characterises the molecules, and the output domain is typically a set of numbers quantifying the property of interest.

F_R: the molecular representation can be substantiated by means of SMILES notation, H depleted graphs, geometrical representations, and by quantum-mechanical optimised structures.

F_E: encodes the structure i.e. descriptors of the molecules. These descriptors can be produced in different ways, in particular:

(a) Based on molecular properties using experimental quantities as descriptors. In this approach, the feature representation of the molecule is realized through several characteristic experimental properties (solvatochromic parameters). In this case, the limitation lies mainly on the fact that a given set of experimental values is required for each compound. This is the method employed to derive Kamlet-Taft parameters (linear solvation energy relationship, LSER) (Branco*et al*., 2015).

(b) Based on group contributions. This relies on the idea that a molecule acts as a number of fragments (atoms, bonds, moieties) independently contributing to the investigated property, usually by an additive scheme. Many of these methods are of questionable accuracy and have limited applicability due to the oversimplification of the molecular structure representation (Rogers and Seddon, 2003).

(c) Based on structural molecular descriptors. Molecular descriptors such as topological indices, quantum-chemical, geometrical, and electrostatic descriptors and so forth, are used to encode molecular features. This approach can take advantage of full, realistic representation of the molecules if descriptors are derived from quantum chemical geometry optimisations, especially those carried out at *ab initio* level (Branco*et al*., 2015).

The mapping function (g) can be based on linear or non-linear methods, such as multivariate statistics, standard regression analyses, genetic algorithm regression models and artificial neural networks. Linear methods are more easily interpretable, and conclusions on the mechanisms of action can eventually be

derived. Non-linear methods, on the other hand, yield more accurate models, yet difficult to understand in physical terms (Branco*et al*., 2015).

Thus, the goal of a QSAR is to find a function that predicts any molecular activity (IC_{50} in this case), using information solely derived from the chemical structures of the molecules (Bernazzani *et al.,* 2006) thereby offering important structural insight in the design of novel drugs by exploring and harnessing the structural requirements controlling the observed biological activities of potential drug candidates, reducing the requirement for lengthy, costly and hazardous laboratory test aiding speedy drug discovery and development.

The basic procedures for developing this assumed quantitative relationship between the macroscopic and the microscopic properties of matter (activity/cytotoxicity of naphthylisoquinoline and bisbenzylisoquinoline derivatives, in this regard) include the collection of data of naphthylisoquinoline and bisbenzylisoquinoline derivatives whose biological activities as well as their cytotoxicity (IC_{50}) have been experimentally determined from literature, followed by the geometry optimisation of the collected molecules, conversion of the molecular structures into mathematical descriptors that encapsulate the key properties of the molecules relevant to the biological activities being modelled, splitting of the data set of molecules into a training set and a test set, model building using the training set and external validation of the generated QSAR models.

2.6 Molecular Descriptors

It is an assumption that differentstructures of a molecule have different chemical properties and similar molecular structures have similar molecular properties. Every molecular representation describes a different way to look at the molecular structure and its chemical meaning is strongly immersed in the framework of the chemical theories. The final result of a logical and mathematical procedure which

transforms chemical information encoded within a symbolic representation of a molecule into a useful number is referred to as molecular descriptor (Todeschini and Consonni, 2000). Some of the molecular descriptors that were used in this study are shown in Table 2.1**.**

2.7 Quantitative Structure Activity/Toxicity Relationship (QSAR).

Quantitative structure-activity relationship(QSAR) is the mathematical relationship between biological activity and molecular structure. Chemical structures give rise to different properties, and hence specific features areassumed to be responsible for the biological activity. The mathematical relationships between activity/toxicity and structure can be established and then used to predict theactivity of new chemicals. In other words, the biological activity of new chemicalscan be predicted from structure alone without experimental measurement of activity. Moreover, this offers enormous advantages for the pharmaceuticalindustry to get an early idea of biological properties, such as oral absorption, beforesynthesis, providing information for compound development. Many studies which form the basis of modern QSAR exist today some of which includes the works byHammett and Taft involving the Hammett electronic constant and steric effects as well as many other made way for the work carried out by Hansch, who isconsidered the founder of modern day QSAR (Hammett, 1970, Hammett, 1935 and Taft, 1952). The research carried out by Hansch and co-workers produced an equation thatillustrated that the biological activity (Log 1/C) from chemicals could be describedby a series of parameters: hydrophobicity and electronic effects (Hansch *et al.*,1962). From this work, there have been vast developments and improvements forQSAR and these principles are the foundations for new approaches based on the machine learning and statistics.

Table 2.1Representative of the molecular descriptors

SN	Descriptor	Meaning	Class
1	minHBd	Number of Hydrogen Bond Donor	2D ElectrotopologicalStateDescriptor
2	μ	Dipole moment	Electronic
3	nH	Number of hydrogen atoms	AtomCountDescriptor
4	SssCH2	Sum of the atom level E-state values for all the methylene in the molecules	2D ElectrotopologicalStateDescriptor
5	Wlambda2.unity	Directional WHIM, weighted by unit weights	3D WHIMDescriptor
6	MDEC-33	Molecular distance edge between all tertiary carbons	2D MDEDescriptor
7	BCUTp-1l	nhigh lowest polarizability weighted BCUTS	2DBCUTDescriptor
8	BCUTc-1l	nhigh lowest partial charge weighted BCUTS	2DBCUTDescriptor
9	minHsOH	Minimum atom-type H E-State: -OH	2D ElectrotopologicalStateDescriptor
10	Weta2. Unity	Directional WHIM, weighted by unit weights	3D WHIM Descriptor
11	RBF	rotatable bond fraction	Constitutional indices
12	nTB	number of triple bonds	Constitutional indices
13	nBnz	number of benzene-like rings	Ring descriptors
14	ARR	aromatic ratio	Ring descriptors
15	PW5	path/walk 5 - Randic shape index	Topological indices
16	MAXDN	maximal electrotopological negative variation	Topological indices
17	ATSC8s	Centred Broto-Moreau autocorrelation of lag 8 weighted by I-state	2D autocorrelations
18	MATS1m	Moran autocorrelation of lag 1 weighted by mass	2D autocorrelations
19	nArCOOR	number of esters (aromatic)	Functional group counts
20	MLOGP	Moriguchi octanol-water partition coeff. (logP)	Molecular properties

Quantitative structure-activity relationship (QSAR) methodologies save resources and expedite the process of the development of new molecules and drugs. Quantitative structure-activity relationship (QSAR) has a large number of applications other than those of drug design, some of which include, prediction of toxicity to environmental species, selection of compounds with optimal pharmacokinetic properties, prediction of the fate of molecules released into the environment, prediction of a variety of physico-chemical properties of molecules, prediction of toxicity to humans through deliberate, occasional and occupational exposure, identification of hazardous at early stages of product development (Puzyn *et al*., 2010).

Quite a number of QSAR researches related to design of antimalarial drugs have been reported in literature.Ferreira*et al*. (2010) reported in the modelling and multivariate analysis of new antimalarial artemisinins that descriptors such asradial distribution function, RDF060e,hydration energy, HE andthe distance between the O1 atom from the ligandand the iron atom from heme, d(Fe–O1) were the properties responsible for distinguishingcompounds of higher or lower antimalarial activity. In the study of QSAR analysis of benzothiazole derivatives of antimalarial compounds conducted by Hadanu*et al*. (2015), descriptors EHOMO,ELUMO, polarisability,atomic net charges: C4, C5, and C6 were reported to be responsible for the pharmacological activity of the benzothiazole derivatives. Descriptors dipole moment X component, lipole X component, Kier chiV4, index and ipso- Atom E-state index were reported to be responsible for the antimalarial activity of bisbenzofuran derivatives (Paliwal*et al*., 2012). The QSARstudy of some substituted 4-quinolinyl and 9-acridinylhydrazones carried out by Sahu*et al*. (2012)indicated that the T_N_O_3,IdAverage, chiV6chain, Most-vePotential and T_C_N_6 molecular descriptors played an important role in determining antimalarialactivities. AQSAR study of the antimalarial activity of new artemisinin compounds (Santos *et al.,* 2014) was conducted and the molecular features responsible for the activity of the compounds are hydration energy (HE), charge on

oxygen atom of O11(QO11), torsion angle O1−O2−Fe−N2 (D2) and maximum rate of R/Sanderson Electronegativity(RTe+). Patel *et al*. (2014) conducted a QSARresearch on tylophorine derivatives as efficacious in antimalarial drug design and reported electronegativityCount, SaaNcount, SaasCcount, and the chiV6chainas the most significant molecular descriptors responsible for the inhibitory activity to *Plasmodium kinases*.In his study of antimalarial activity of *Plasmodium falciparum* dihydroorotate dehydrogenase (PfDHODH) inhibitors, Hou*et al*. (2016) reported the hydrogen bonding ability, polarisability and ring complexity as predominant factors for inhibitors antimalarial activity. Despite these studies and many more, a systematic QSAR study is yet to be carried out on naphthylisoquinolines andbisbenzylisoquinoline as antimalarial agents. Hence, 45 derivatives of naphthylisoquinolines compound and 20 derivatives ofbisbenzylisoquinoline were used in this research and were subjected to various kinds of QSAR analyses.

As with other technics, model development comes with some limitations such as insufficient data size (e.g., lack of data on weak or nonactive compounds), lack of curation of data, lack of standardised procedures that facilitate sharing of data or models, toxicological modelling which is inherently a complicated problem (Valerio, 2009; Tropsha *et al*., 2003; Benigni and Giuliani, 2003; Yap *et al*., 2006; Tropsha and Golbraikh, 2007 and Yang, 2010).

CHAPTER THREE

3.0 MATERIALS AND METHODS

3.1 Materials

3.1.1 Instrumentation

A core i7 personal computer (CPU) with a processor speed of 2.6GHz, 8GB RAM and 1TB hard drive was used to carry out this study. The softwares used wereSpartan 14 v.1.1.0 for geometric optimisation,PaDel software for calculating molecular descriptors, BuildQSAR software for constructing models, vWSP program for variable reduction, DIC MRL plus software for model validation, material studio software for correlation matrix calculation andChemBioDraw v 12.0.1V for drawing the molecular structures, and Microsoft Excel for statistical analysis.

3.2Methodology

This research is a theoretical study that uses DFT/B3LYP functional method in QSAR modelling of naphthylisoquinoline and bisbenzylisoquinoline derivatives. The structure of naphthylisoquinolines as well as their activities and cytotoxicity are presented in Table 3. 1. Those of bisbenzylisoquinoline derivatives are presented in Table 3. 2. The descriptors of all the naphthylisoquinolines derivatives as well as those of bisbenzylisoquinoline derivatives were used for the computational analysis. Forty- five (45) naphthylisoquinoline derivatives and twenty(20)bisbenzylisoquinoline derivatives as antimalarial agents are applied to statistical calculation as dependent variable for antiplasmodial activities.The summary of these protocols is given in Figure 3.1.

Table 3.1 Series of Naphthylisoquinoline alkaloids with their antiplasmodial activity against K1 strain as well as their cytotoxicity against L6 cell line.

Comp. No.	Structure	Antiplasmodial Activity IC_{50} (µM)	pIC_{50}[a]	Cytotoxic Activity IC_{50} (µM)	pIC_{50}[b]
1		0.71	6.15	-	-
2		1.24	5.91	100.00	4.00
3		2.40	5.62	43.79	4.36
4		1.70	5.77	34.34	4.46
5		3.49	5.46	94.18	4.03

6				
	0.14	6.86	100.00	4.00
7				
	0.12	6.92	27.00	4.57
8				
	1.70	5.77	100.00	4.00
9				
	-	-	98.20	4.01
10				
	0.22	6.65	37.51	4.43

11				
	0.26	6.58	40.23	4.40
12				
	18.87	4.72	100.00	4.00
13				
	1.88	5.73	58.84	4.23
14				
	0.82	6.09	52.22	4.28
15				
	2.49	5.60	100.00	4.00
16				
	0.44	6.36	100.00	4.00

17				
	10.83	4.97	100.00	4.00
18				
	9.91	5.00	100.00	4.00
19				
	6.44	5.19	56.70	4.25
20				
	0.53	6.27	100.00	4.00
21				
	3.00	5.52	100.00	4.00

22				
	0.46	6.34	96.76	4.01
23				
	7.96	5.10	38.36	4.42
24				
	0.86	6.07	50.76	4.29
25				
	6.91	5.16	34.87	4.46
26				
	0.76	6.12	57.07	4.24

27				
	1.18	5.93	71.98	4.14
28				
	0.48	6.32	72.54	4.14
29				
	17.40	4.76	386.03	3.41
30				
	1.21	5.92	184.86	3.73
31				
	14.73	4.83	127.95	3.89
32				
	22.80	4.64	410.72	3.39
33				
	15.63	4.81	406.98	3.39

34				
	2.09	5.68	300.95	3.52
35				
	1.09	5.96	217.17	3.66
36				
	15.95	4.80	90.05	4.05
37				
	1.33	5.87	95.67	4.02
38				
	2.34	5.63	151.00	3.82
39				
	0.96	6.02	23.13	4.64
40				
	4.48	5.35	289.21	3.54
41				
	1.75	5.76	267.86	3.57

42	10.87	4.96	209.73	3.68
43	26.14	4.58	470.88	3.33
44	1.08	5.97	264.65	3.58
45	11.34	4.95	287.39	3.54

NB: a: pIC_{50} for activity

b: pIC_{50} for cytotoxicity

Table 3.2 Series of Bisbenzylisoquinoline alkaloids with their antiplasmodial activity against K1 strain as well as their cytotoxicity against KB cell line.

Comp. No.	Structure	Antiplasmodial Activity IC_{50} (µM)	pIC_{50}[a]	Cytotoxic Activity IC_{50} (µM)	pIC_{50}[b]
1		0.63	6.20	108.00	3.97
2		6.32	5.20	411.00	3.39
3		0.96	6.02	46.40	4.33
4		1.36	5.87	105.00	3.98
5		3.46	5.46	82.20	4.09

6	1.06	5.97	74.40	4.13
7	0.09	7.05	41.20	4.39
8	1.46	5.84	43.60	4.36
9	0.57	6.24	NT	-
10	0.83	6.08	31.90	4.5
11	1.43	5.84	104.00	3.98

12	0.45	6.35	17.80	4.75
13	0.74	6.13	55.40	4.26
14	3.92	5.41	54.80	4.26
15	22.00	4.66	421.00	3.38
16	42.10	4.38	448.00	3.35
17	1.16	5.94	NT	-

18				
	2.06	5.69	18.80	4.73
19				
	1.81	5.74	38.30	4.42
20				
	2.07	5.68	294.00	3.53

NT = Not toxic

NB: a: pIC_{50} for activity

b: pIC_{50} for cytotoxicity

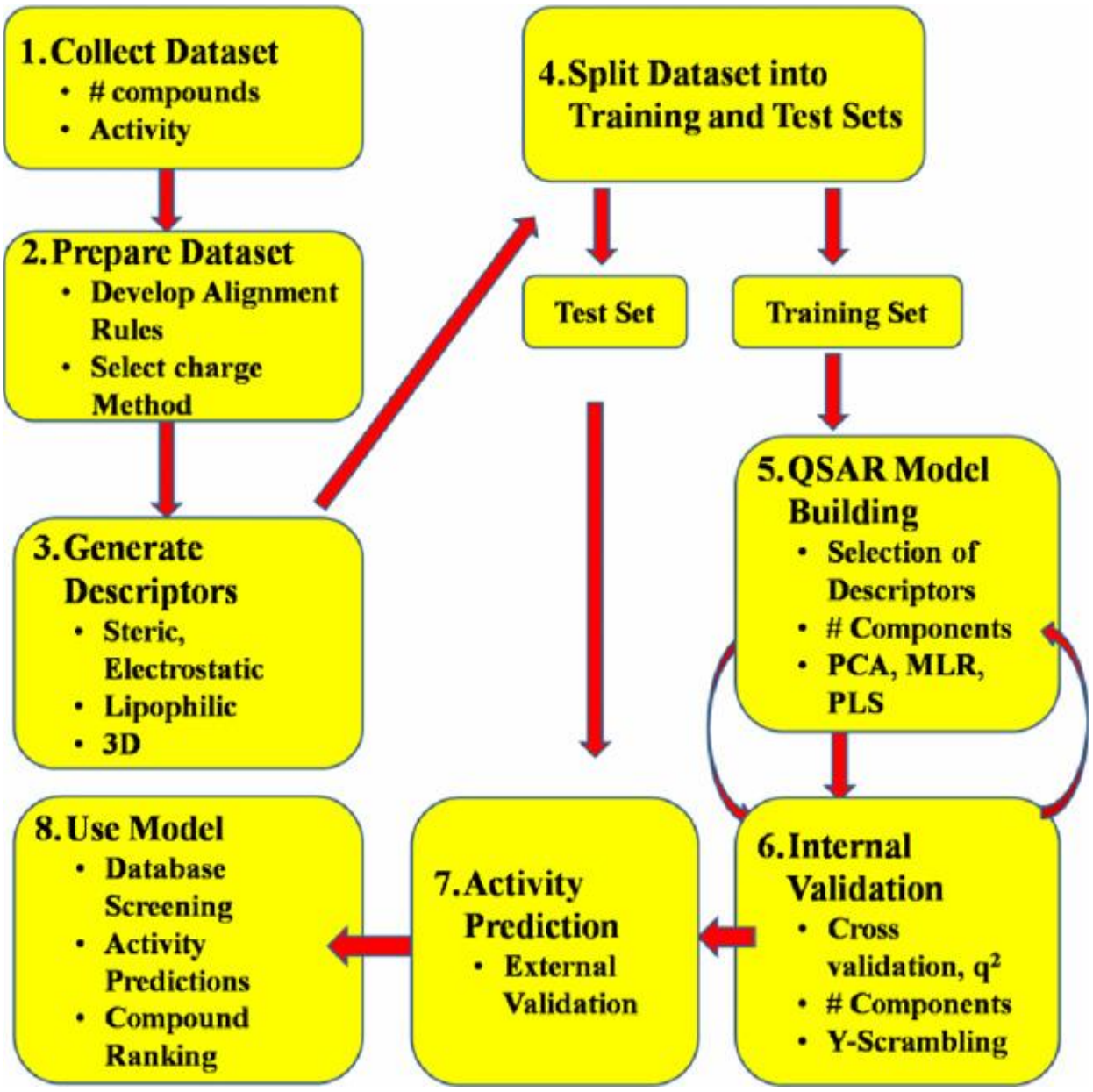


Figure 3.1Workflow for the model development

3.2.1 Biological data

The biological activities/ cytotoxicity used in this research were obtained from the *in vitro* tests for50% inhibitory concentration (IC_{50}) against a multidrug-resistant, K1 strain of *Plasmodium falciparum* as well as the cytotoxicity against L6 and KB cell lines.The data set of 45 compounds of naphthylisoquinoline with their antimalarial activities/cytotoxicity were taken from published journals (Brigmann *et al*., 2004a; 2004b) and presented in Table 3. 1. The 20 compounds of bisbenzylisoquinoline derivatives with their antiplasmodial activity/cytotoxicity were also obtained from, Marshal *et al*. (1994) and presented in Table 3.2.

3.2.2 Molecular structure sketching

The molecular geometry of the structures were drawn using the Chemdraw software. The Chemdraw software package was launched by *double clicking* the Chemdraw icon on the desktop of a core i7 personal computer (CPU) with a processor speed of 2.6GHz, 8GB RAM and 1TB hard drive.This displayed a work sheet containing a sketch pad seen on the left hand side of the software while on the right hand side is a screen which shows 3D of the sketched molecules as shown in Figure 3.2. Using these sketching tools, the 2D structures of all the compounds were build according to the following procedure:

In drawing the structure of the representative of the series as seen in *Structure 1*, the compound was drawn firstly by selecting the benzene ring, "A" ( ) and then the worksheet was clicked to append it. Then cyclohexene ring ( ) was selected, with cyclohexene like cursor at C_7 on the benzene ring, "A" the cursor key was dragged down to the right and released to append the molecule by the side of the benzene as seen in *Structure 2*.

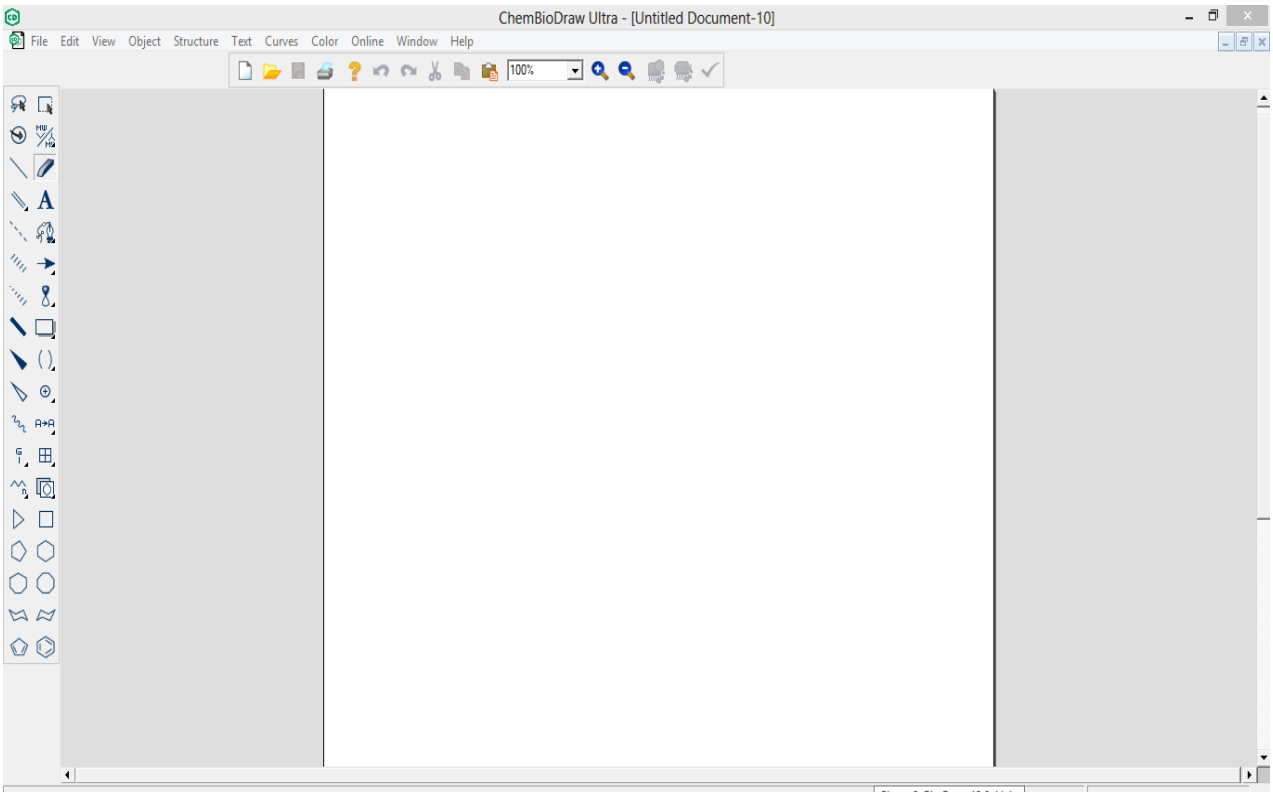


Figure 3.2 A chemdraw worksheet

Structure 1 *Structure 2* *Structure 3* *Structure 4*

Nitrogen atom ( ) was thereafter inserted into the second ring at position 4 by selecting the 'A' sign from the menu bar of Figure 3.2 which create a rectangular box in which N was typed. After which double bond ( ) was inserted in between positions 3 and 4 by selecting the double bond and clicking on the single bond between positions 3 and 4. The wedged ( ) and hashed wedged ( ) bonds were then attached at C_5 through selecting the bond from the menu bar of Figure 3.2 and clicking on C_5 to attach. The methyl (CH_3) groups were thereafter attached to these bonds, methoxy group (OCH_3) was attached at position C_1 through selection of a solid bond ( ) and the icon A from the menu then by clicking position C_1, a rectangle is introduced in which OCH_3 was typed appeared. This lead to the *structure 3*.

Another benzene ring, "B", ( ) was drawn at position 8 through selection of the solid bond ( ) from the menu of Figure 3.2 and then selecting the benzene ring also from the menu then clicking on the tail end of the solid bond attached at position 8. A methoxy group (OCH_3) was attached using the solid bond to the carbon atom at position 8' as described above. Yet again another benzene ring, "C" ( ) was selected from the sketch pad and through the $C_{2'}$ and $C_{7'}$dragged down to the right and released, resulting to the overall structure as shown in *structure 4*. The whole structure was highlighted as shown in Figure 3.3.

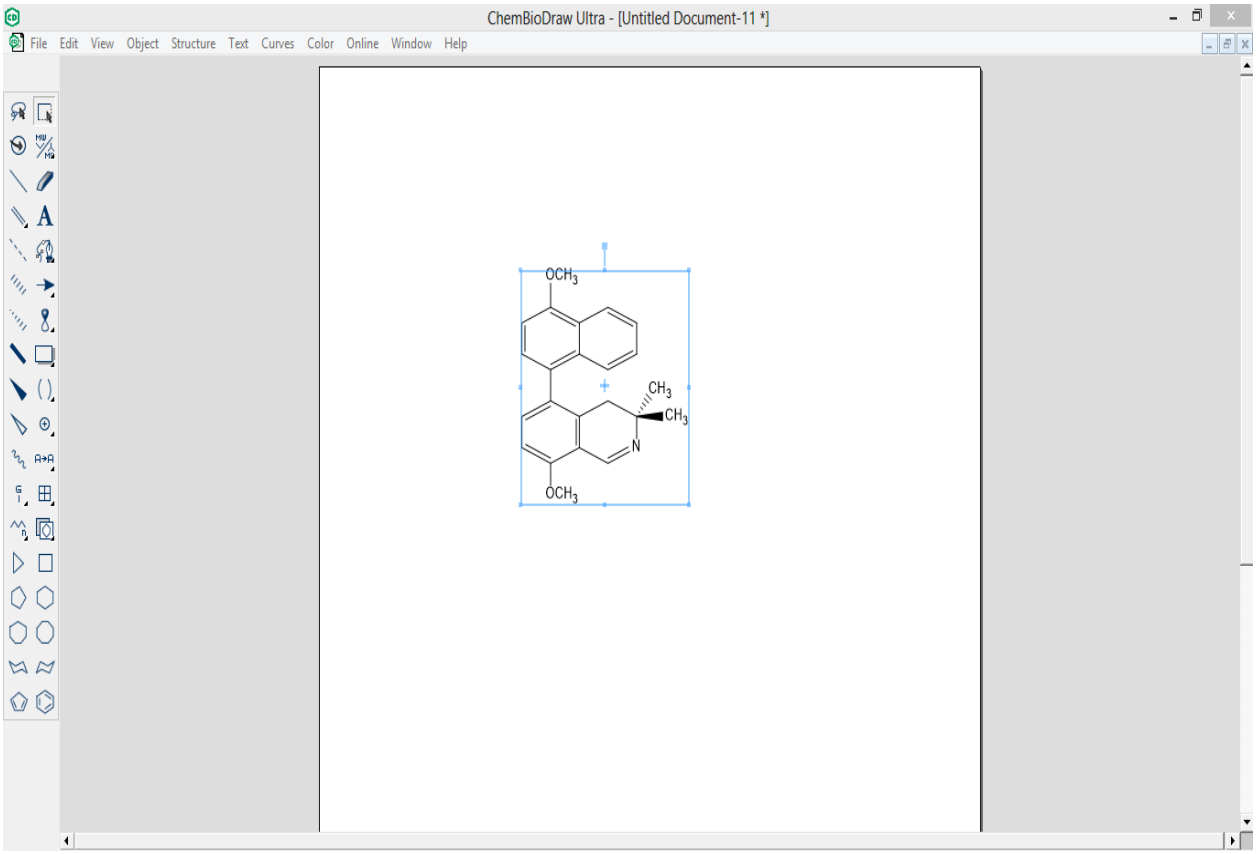
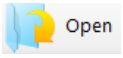
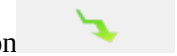


Figure 3.3 A highlighted molecule on chemdraw

From the menu bar, the menu *File* was selected which display several option from which *Save As* ....., was selected to save the selected structure as an *sdf* file in a new folder. These detailed procedures were followed in sketching other structures until all other molecules have been drawn.

3.2.3 Geometry optimisation and calculation of some quantum-chemical descriptors

Optimisation is the process of finding the equilibrium or lowest energy geometry of molecules.The molecular geometry of all the structures wereoptimised in Spartan 14 v.1.1.0 using B3LYP/6-311G* i.e. the compound was optimised using 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. The choice of DFT/B3LYP functional was due to the fact that the method leads to satisfactory results when molecular geometrics and energies are considered (Molfetta *et al*., 2007 and Lameira *et al*., 2006). The compound to be optimised was retrieved from the folder containing the 2D structures that was built earlier in section 3.2.2. To do this, first of all, Spartan 14 v.1.1.0 software was launched to have the Spartan interface/workspace (Figure 3.4). Secondly, the *File* command at the top of the menu bar of Spartan 14 v.1.1.0 interface was clicked to open a sub-commands (Figure 3.5). Thirdly, from the sub-command, a command 

was clicked to open the folder containing the saved 2D structures. The folder was double clicked to open it and then the structure was selected,displaying the 2D structure in 3D on the screen(Figure 3.6), from whichicon  was selected from the menu bar, which results in Figure 3.7containing a building tools.Towards the bottom right hand side of the figure, icon  was *clicked* to minimise the energy of the molecule.

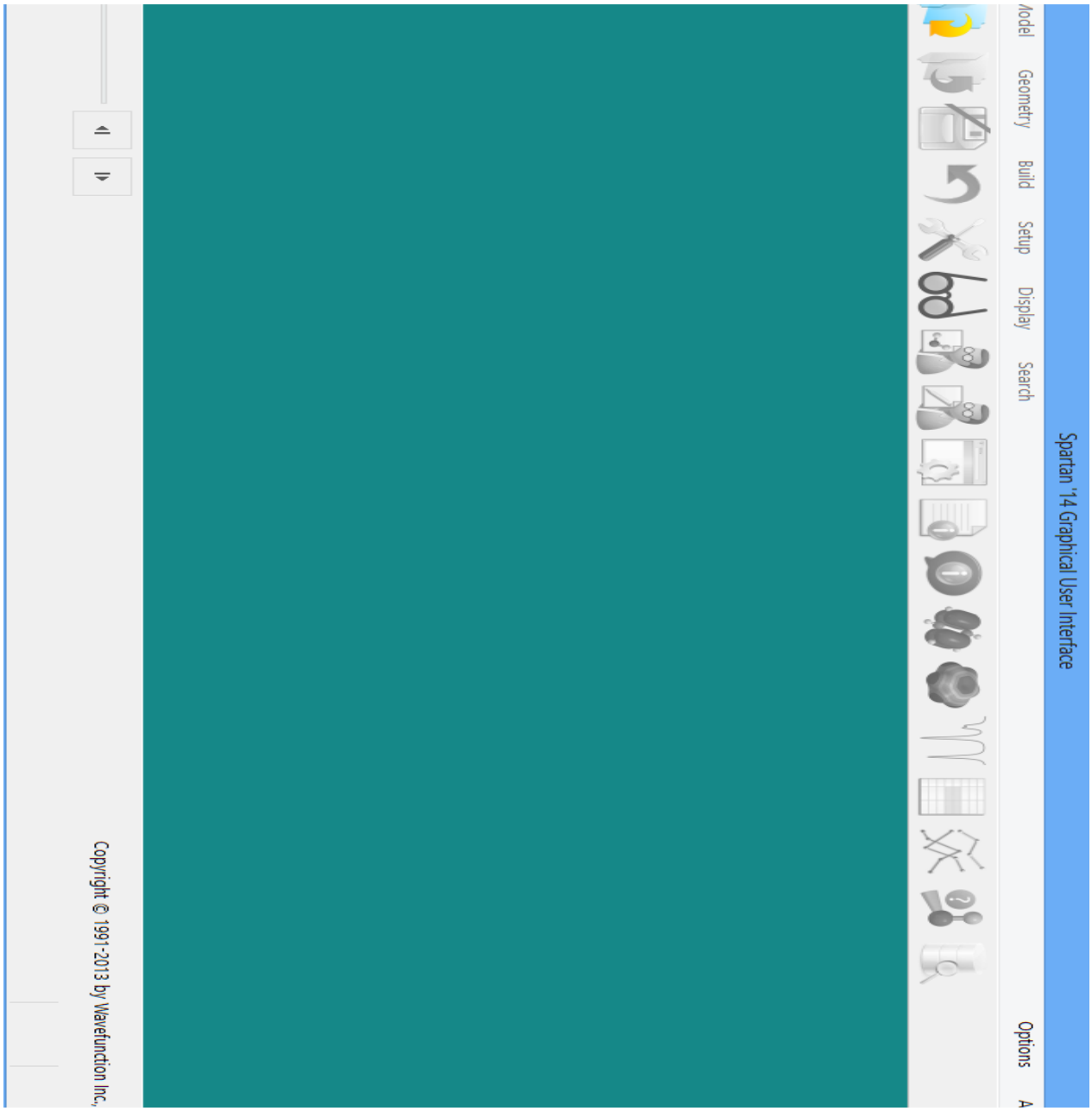


Figure 3.4 Spartan graphic user interface

Figure 3.5 File pull down menu on spartan

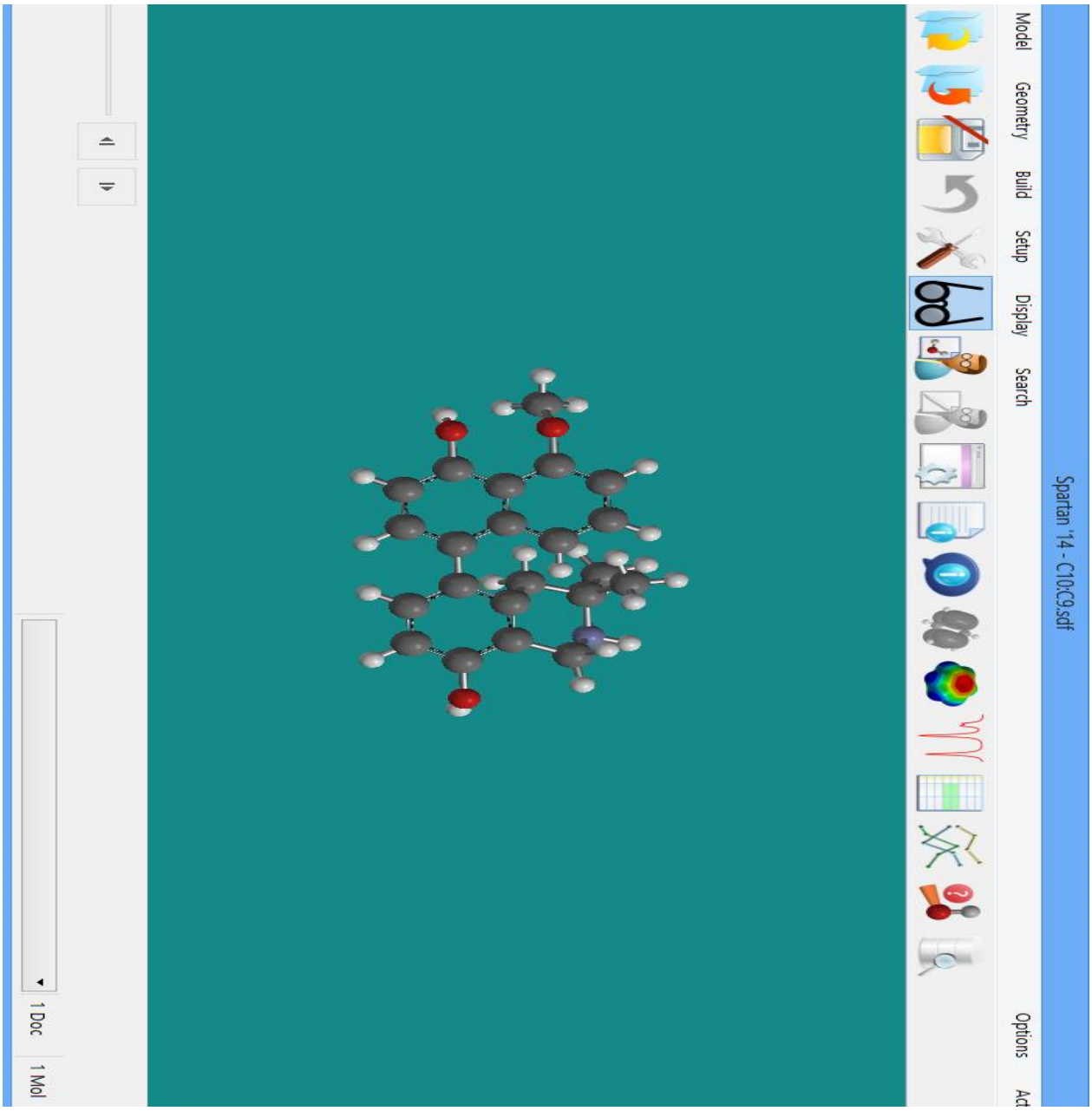


Figure 3.6 3D display of molecule in Spartan

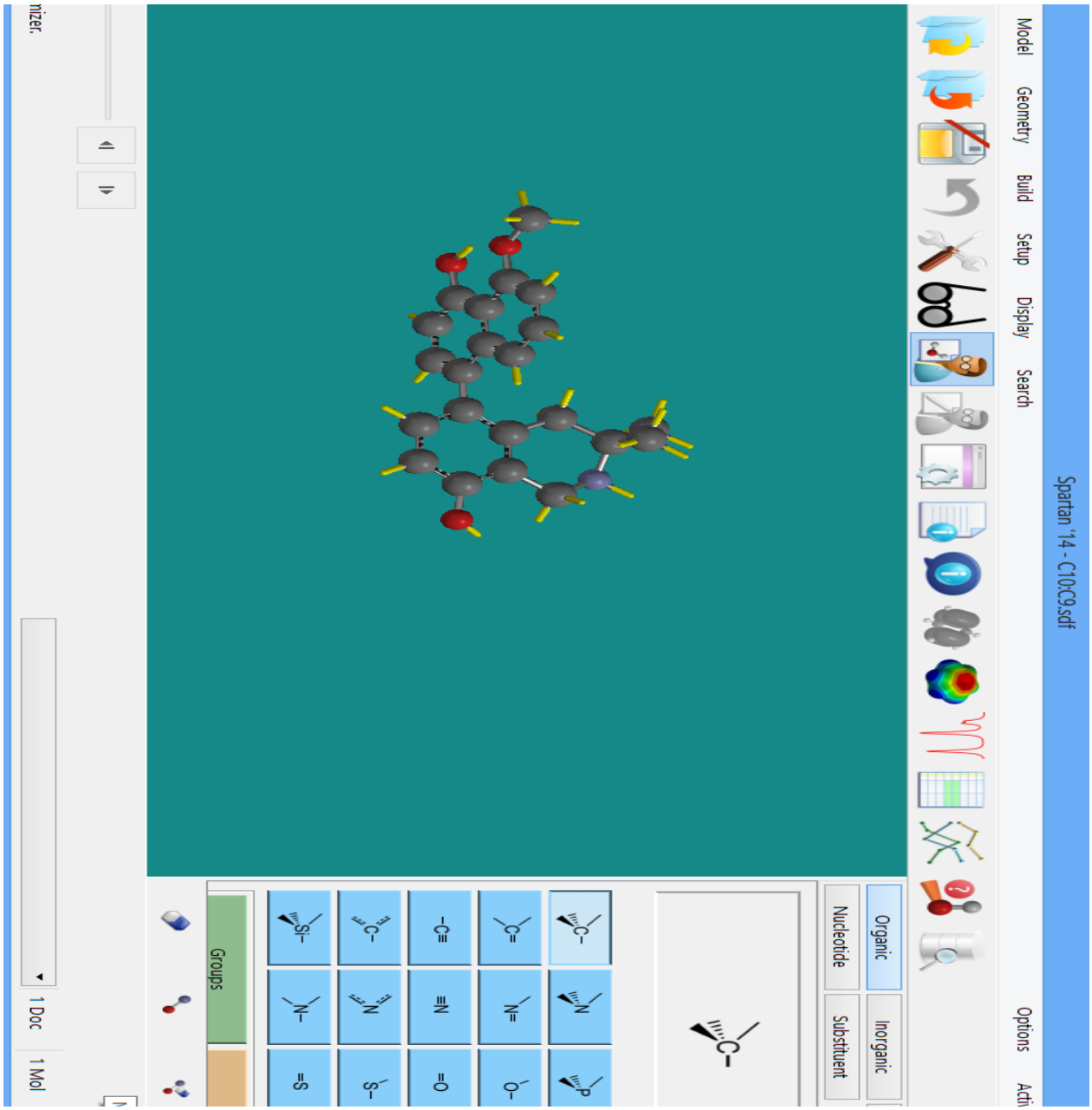


Figure 3.7 Spartan minimising molecule

It took few seconds to minimise the molecules, after which icon  was selected to remove the building tools and return the minimised molecule. Next, from the *Menu Bar*,the *Setup*menu was *clicked*,followed by selecting *calculations*(see Figure 3.8) which resulted in the display of the

*calculations dialogue*box (Figure 3.9), 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 optimisation 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. 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 polarisability and logP. 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 dialogue was *clicked*; *a file browser* appeared (Figure 3.10). In the section *File name*the name of the compound was typed and*Save*was clicked. A notification appeared indicating that the calculation has been submitted. This procedure was applied to optimise all other compounds in this study.

3.2.4 Theoretical background for calculation of the quantum descriptors

Quantum chemical descriptors were calculated theoretically. According to Koopman's theorem (Pearson, 1988), the energies of the HOMO and LUMO orbitals of the molecules are related to the ionisation potential, I, and the electron affinity, A, respectively, by the following relations:

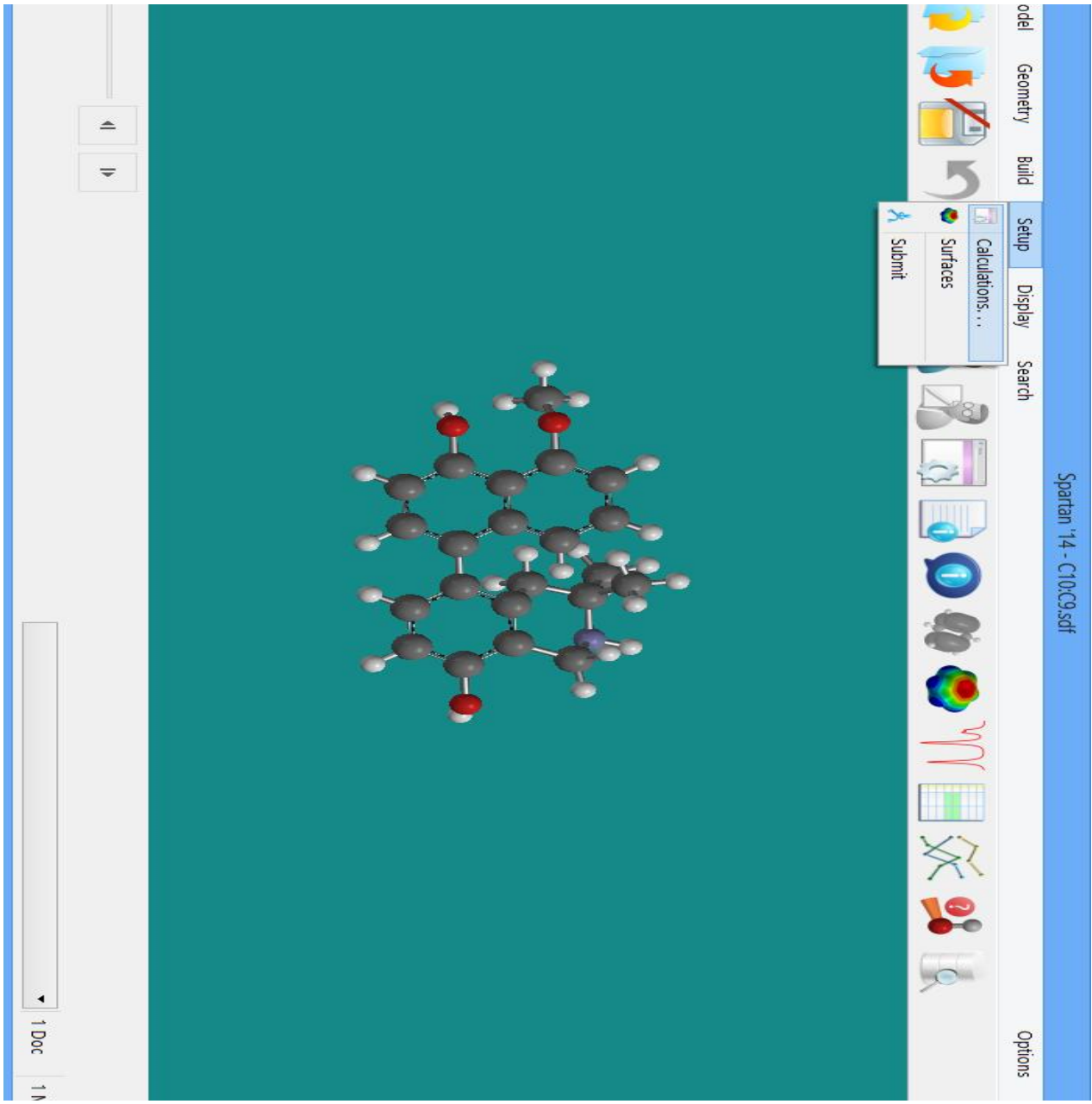


Figure 3.8 Spartan setup menu display

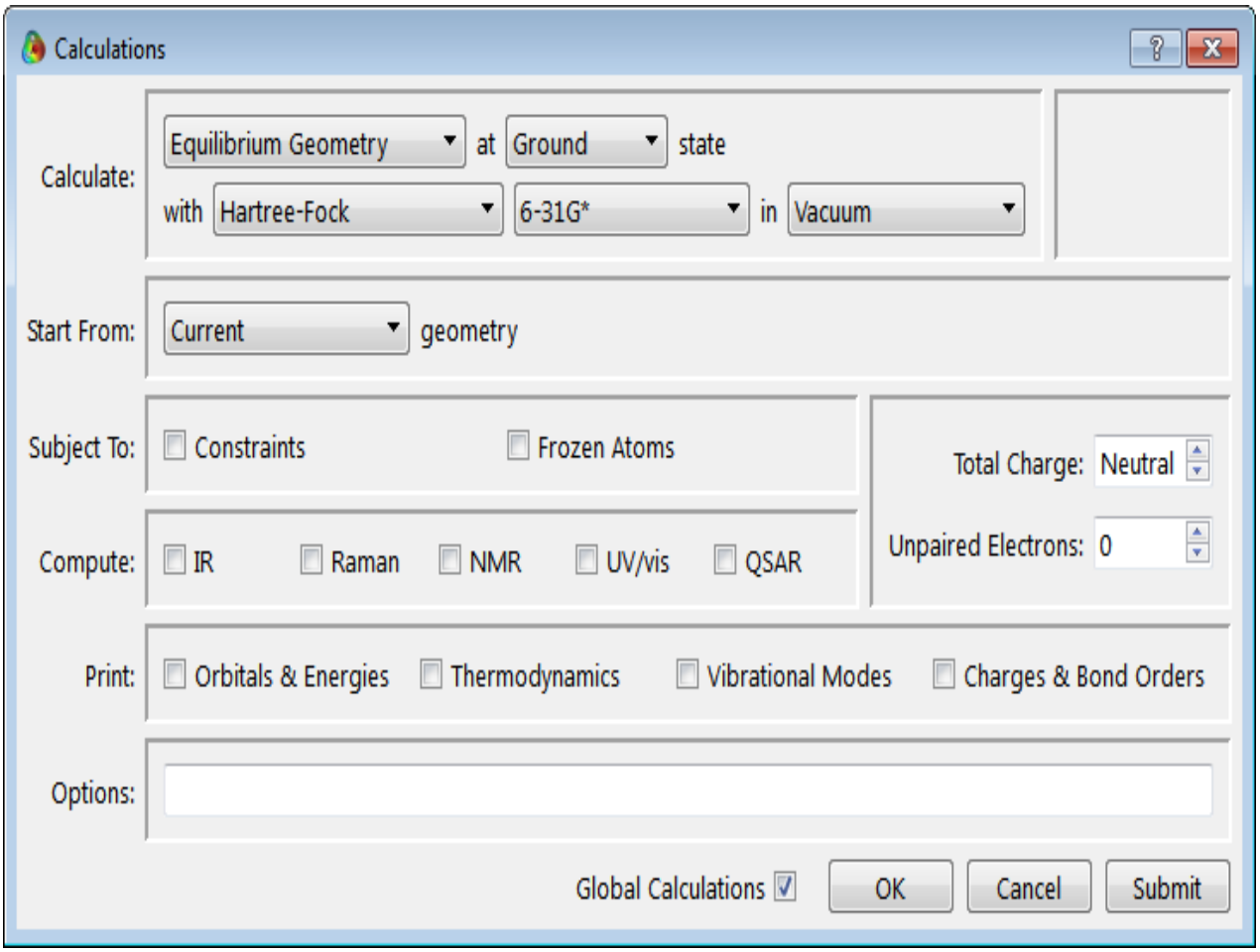


Figure 3.9 Spartan calculations dialogue

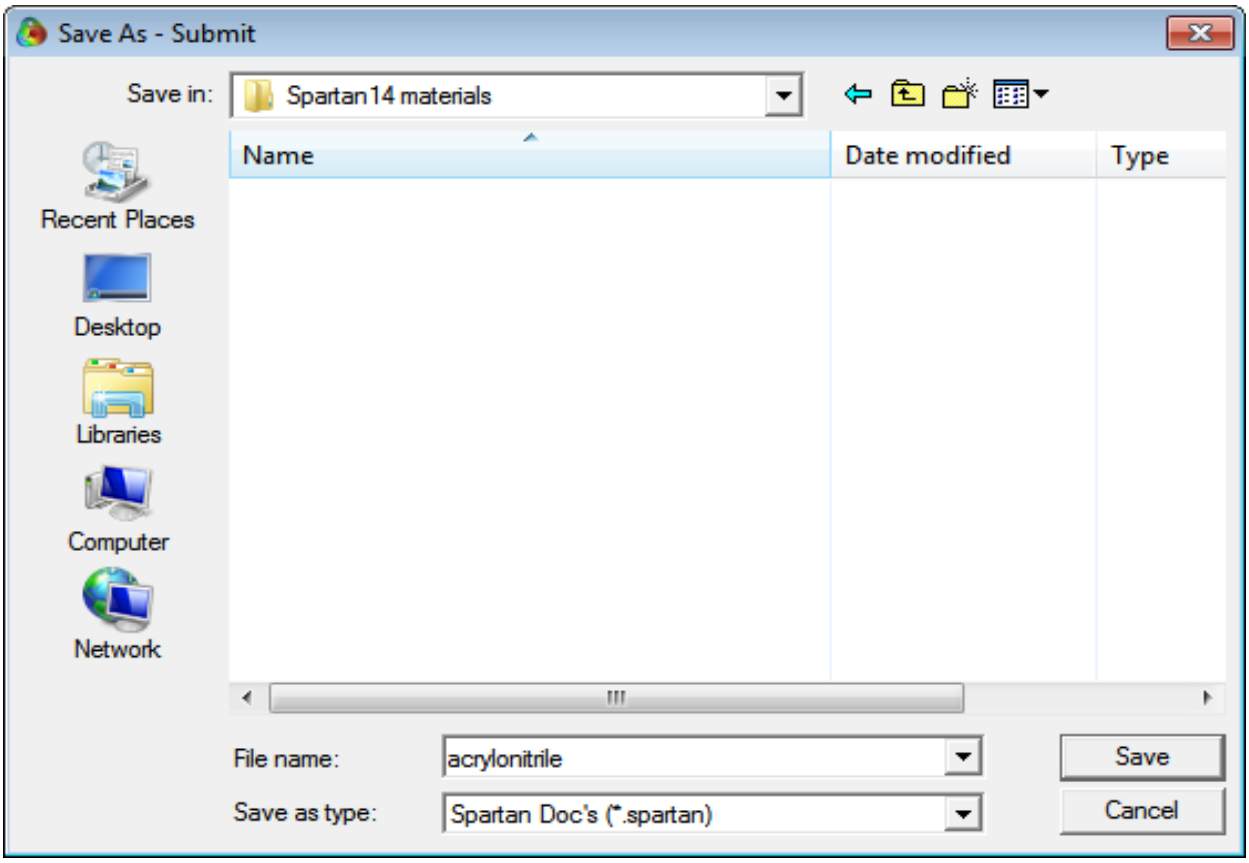


Figure 3.10 File browser

$$I = -E_{HOMO} \quad (1)$$

$$A = -E_{LUMO} \quad (2)$$

Absolute electronegativity, χ, and absolute hardness, η, of the molecules are given by

$$\chi = \frac{I+A}{2} \quad (3)$$

$$\eta = \frac{I-A}{2} \quad (4)$$

The absolute softness, σ, is the reciprocal of absolute hardness (Parr and Pearson, 1983).

$$\sigma = \frac{1}{\eta} \quad (5)$$

The absolute electrophilicity index is given by (Parr *et al*., 1999).

$$\omega = \frac{\mu^2}{2\eta} \quad (6)$$

where μ refers to the dipole moment.

3.2.5 Calculation of 0- 3D descriptors from PaDEL

PADEL software package was used for the calculation of the 0D, 1D, 2D and 3D molecular descriptors for each compound of the training/validation sets. The jar file in the Padel folder (Figure 3.11) was *double clicked* to launch the software and access the interfaceFigure 3.12. The folder containing the earlier optimised molecules was selected by clicking the *molecular directory/file* (Figure 3.13). 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*Figure 3.14. 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* as well as the 3Dboxes were checked to calculate 1D, 2D and 3D descriptors as shown in Figure 3.15.

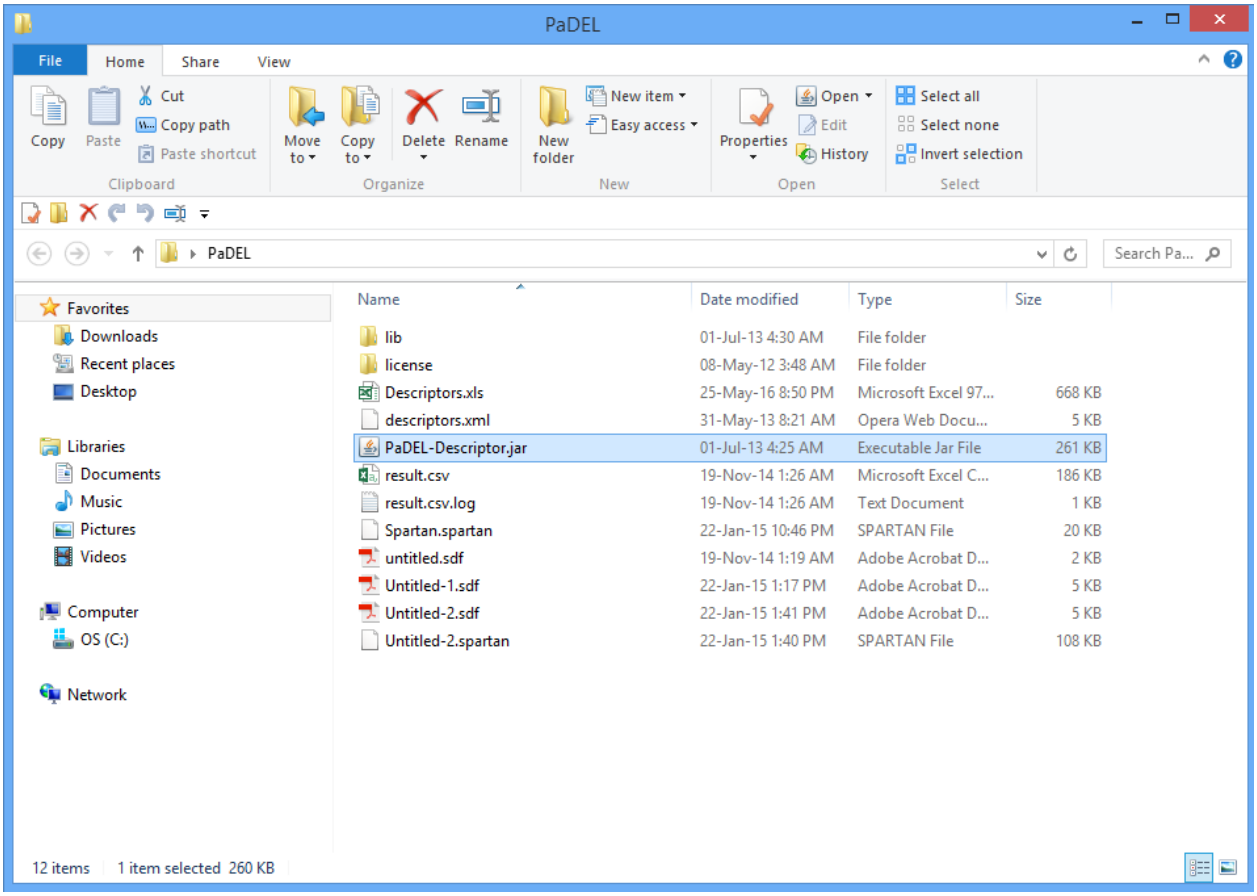


Figure 3.11 A folder of PaDEL-Descriptor jar

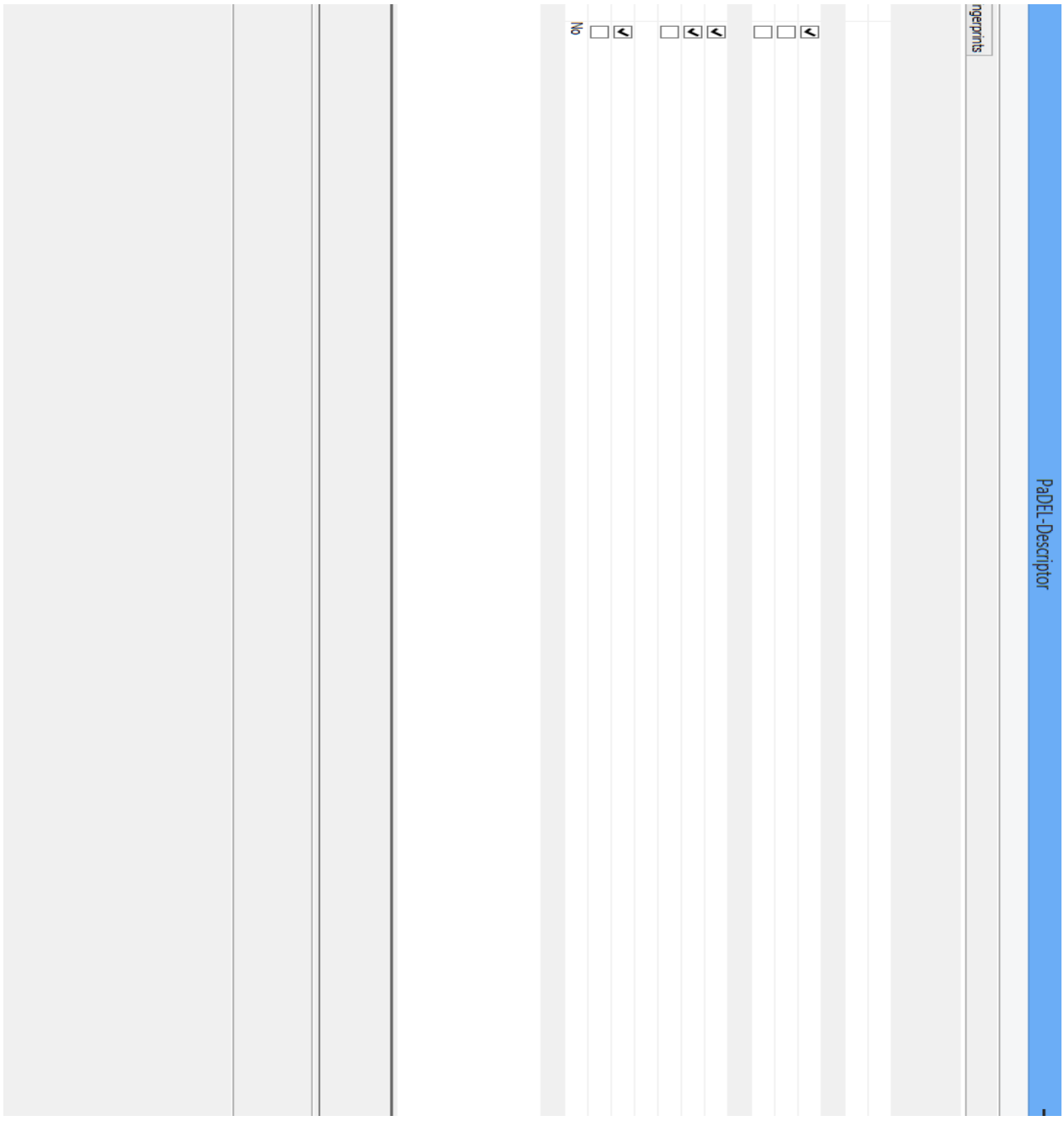


Figure 3.12 PaDEL home page

Figure 3.13 PaDEL file browser

Figure 3.14 PaDEL file output

Figure 3.15 PaDEL 1-3D molecular descriptors

The options "*Remove salt*", "*Detect aromaticity*" and "*Standardise tautomer*", were all *selected* in order to remove salts like NaCl, remove existing aromaticity from the molecule and standardise tautomer respectively before calculation of descriptors, finally the job was then submitted for calculation by clicking on *start***.** This was repeated for all the molecules.

3.2.6 Correlation matrix

The material studio software was launched by *double clicking* the software, which result in the appearance of a *Welcome to material studio* interface (Figure 3.16) asking whether to *create a new project* or to *open an existing project*, from which create a new project was selected and the *OK* was clicked to pop up a new project interface, which desire the project to be named before saving (Figure 3.17). The *OK* was clicked which introduces a blank sheet (Figure 3.18) from which icon  in the menu was clicked. This lead to the display of a dialogue box (Figure 3.19) from which *Study Table Document* was clicked to display the worksheet (Figure 3.20). The biological activities (pIC_{50}) of the molecules were copied from the saved file and pasted in the first column of the spread sheet while the descriptors (saved in csv file format as described in section 3.2.5) correspondingto the activities were also copied to occupy the subsequent columns see Figure 3.21. To create the correlation matrix, the whole data were selected.From *the Menu Bar*of Figure 3.21,*statistics*was*selected*, this brought up *Dialogue box* containing a number of *Analysis Tools* (Figure 3.22). The*Initial Analysis*was selected which again popped up another dialogue box containing number of analytical tools (Figure 3.23) from which*Correlation Matrix*at the bottom of the dialogue was *selected* to calculate the correlation matrix (Figure 3.24).

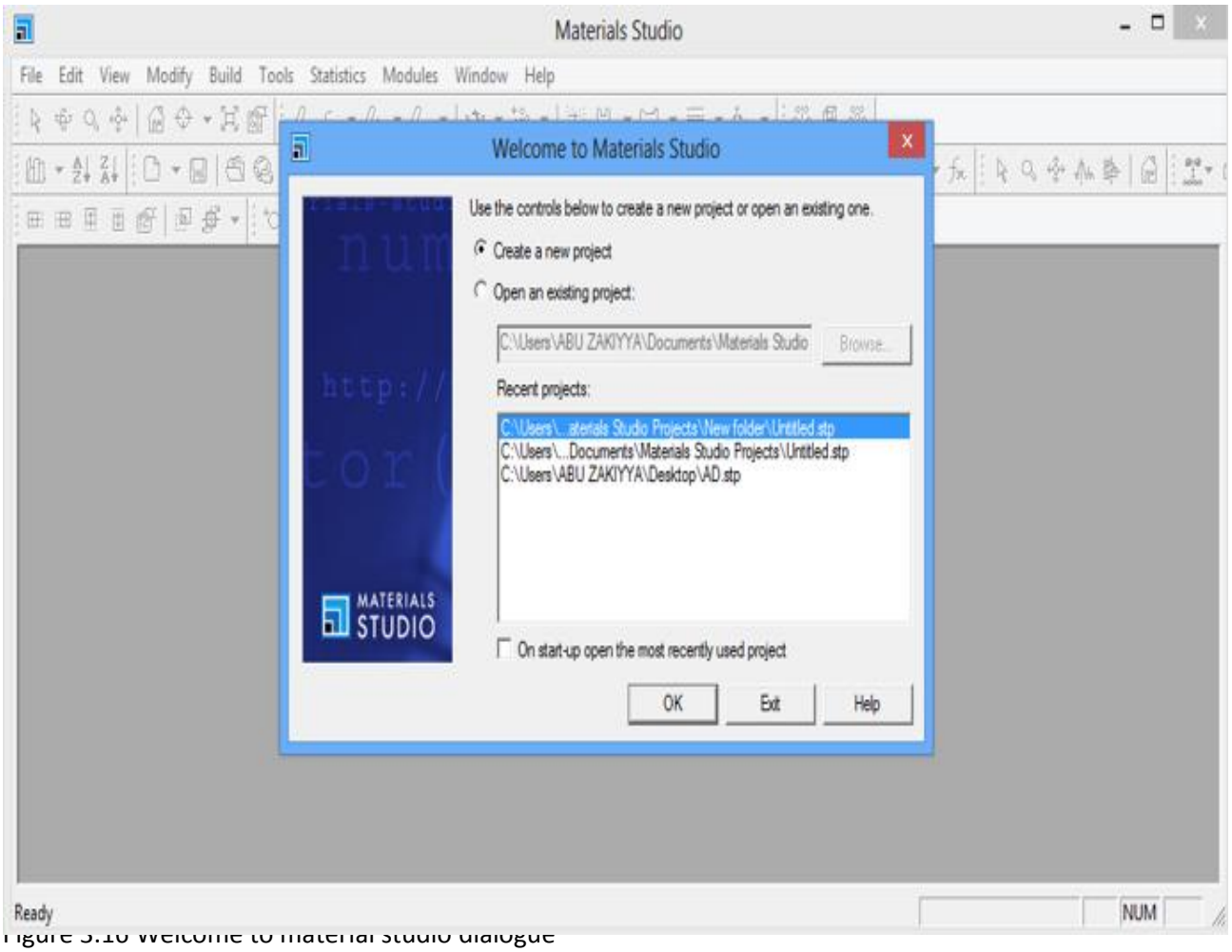


Figure 3.10 Welcome to material studio dialogue

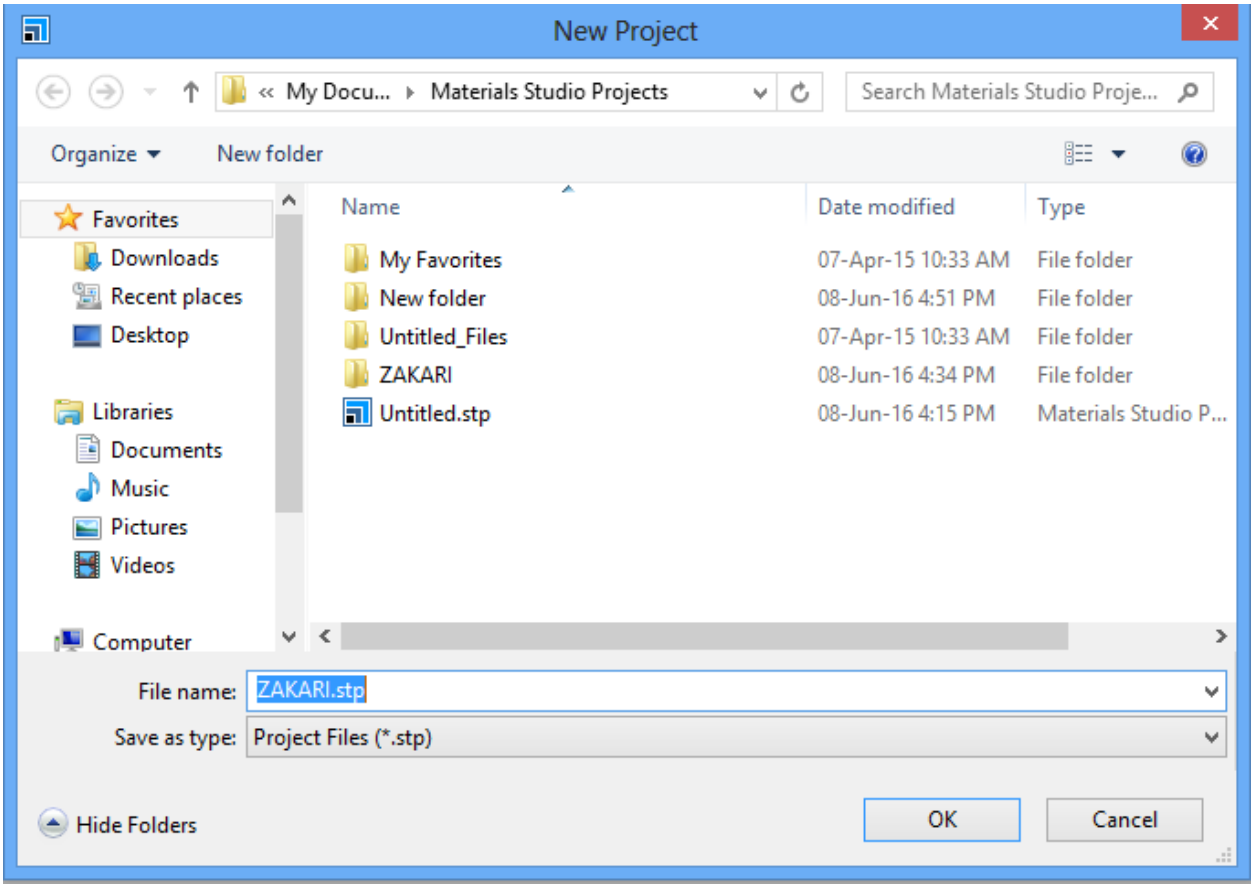


Figure 3.17 A folder of PaDEL-Descriptor jar

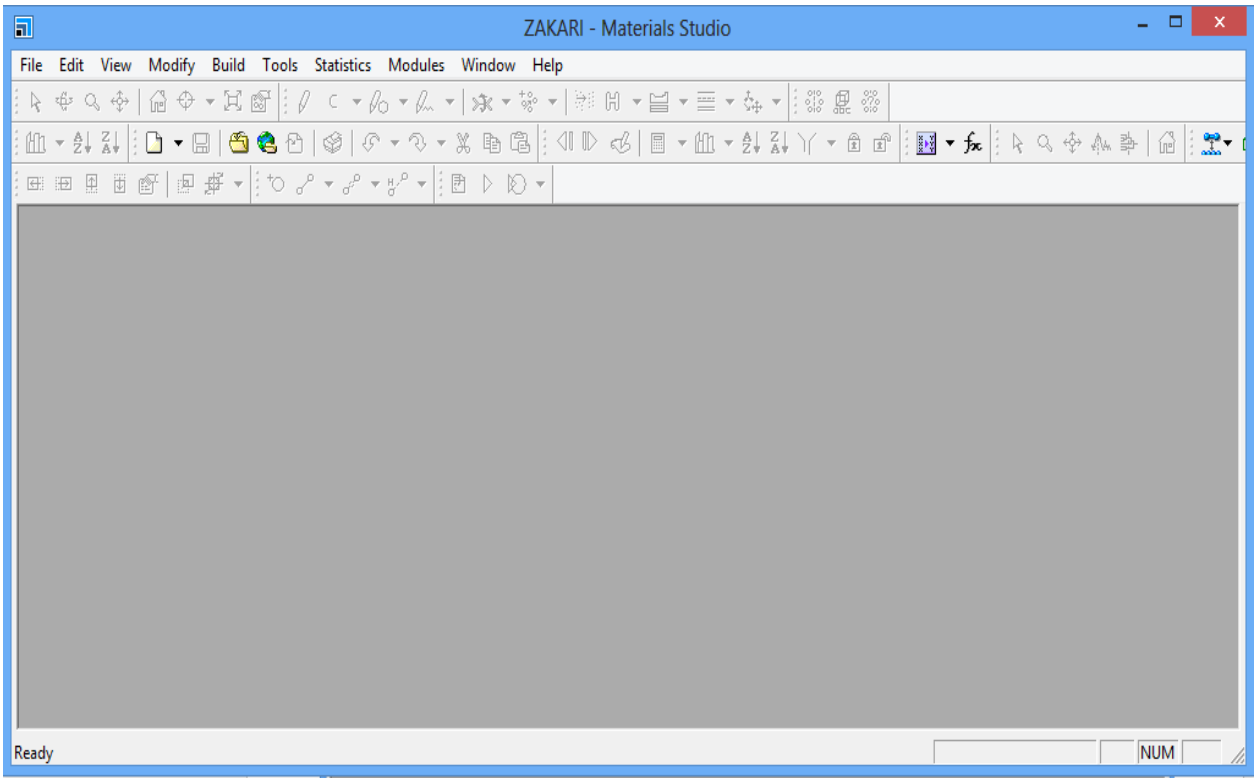


Figure 3.18 Material studio blank sheet

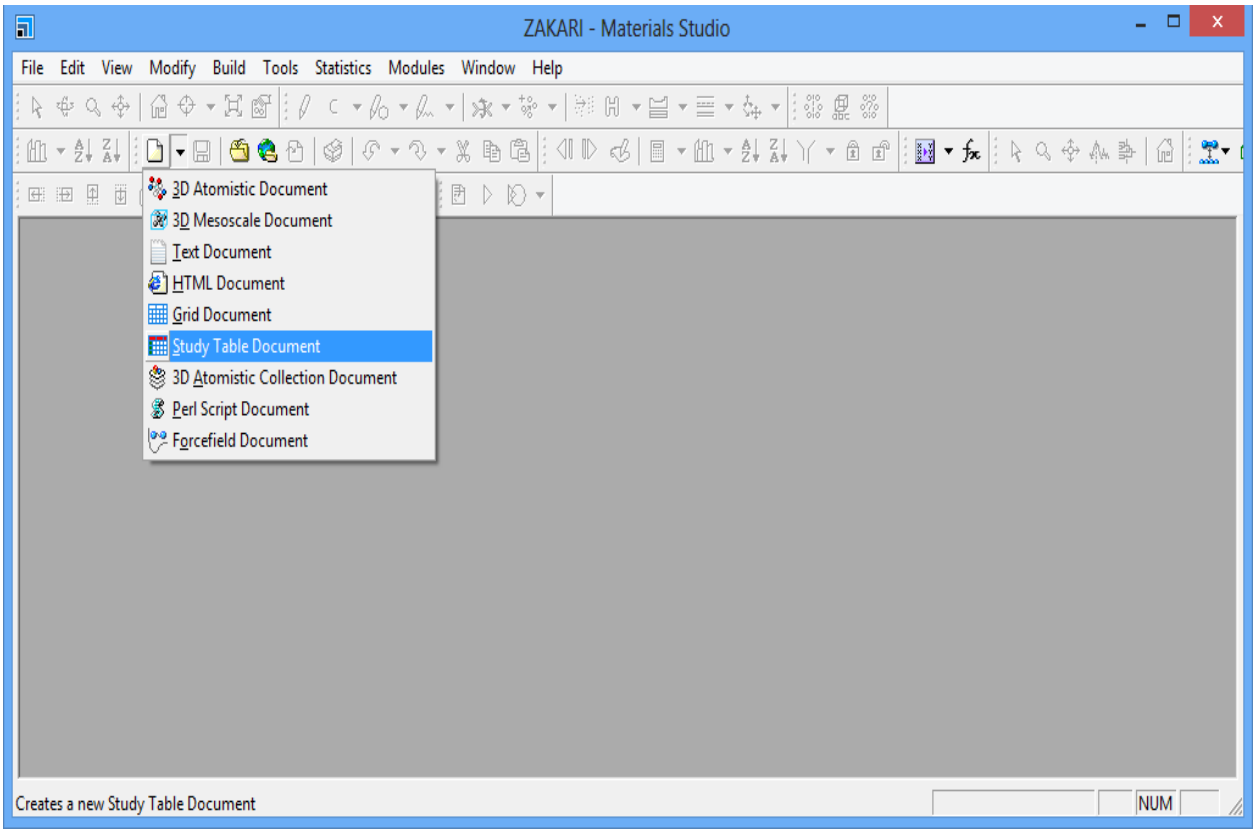


Figure 3.19 Study table dialogue

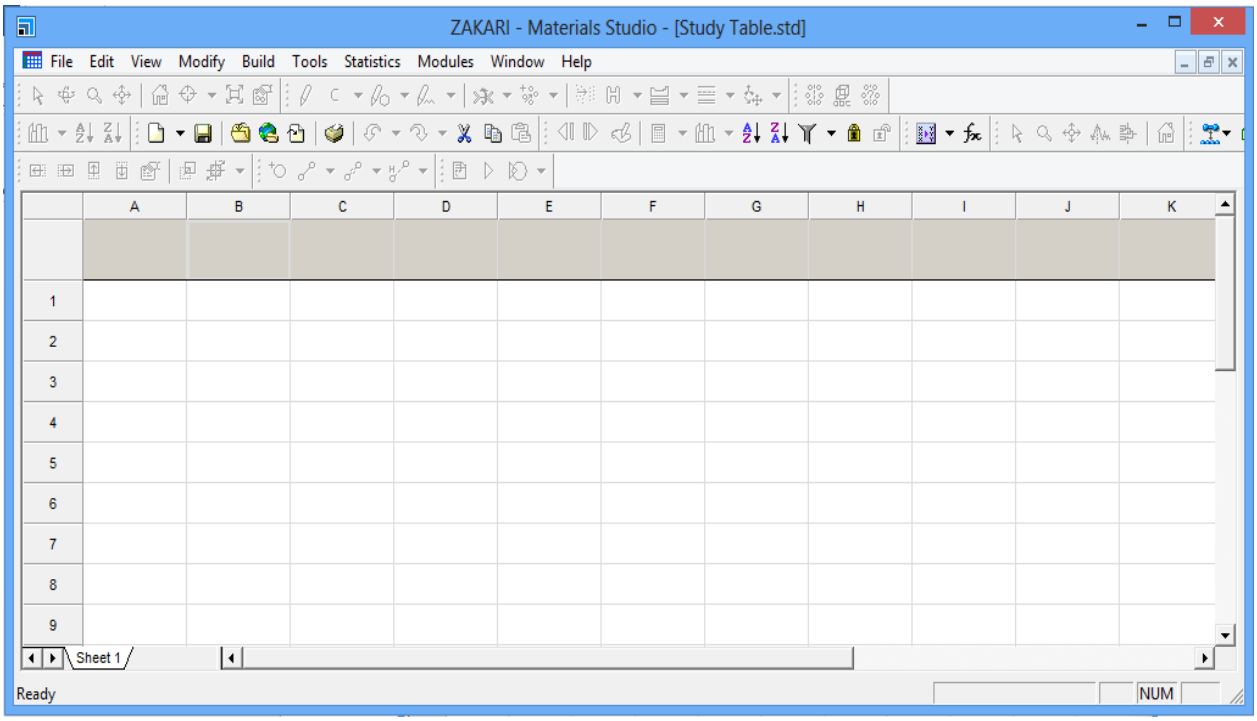
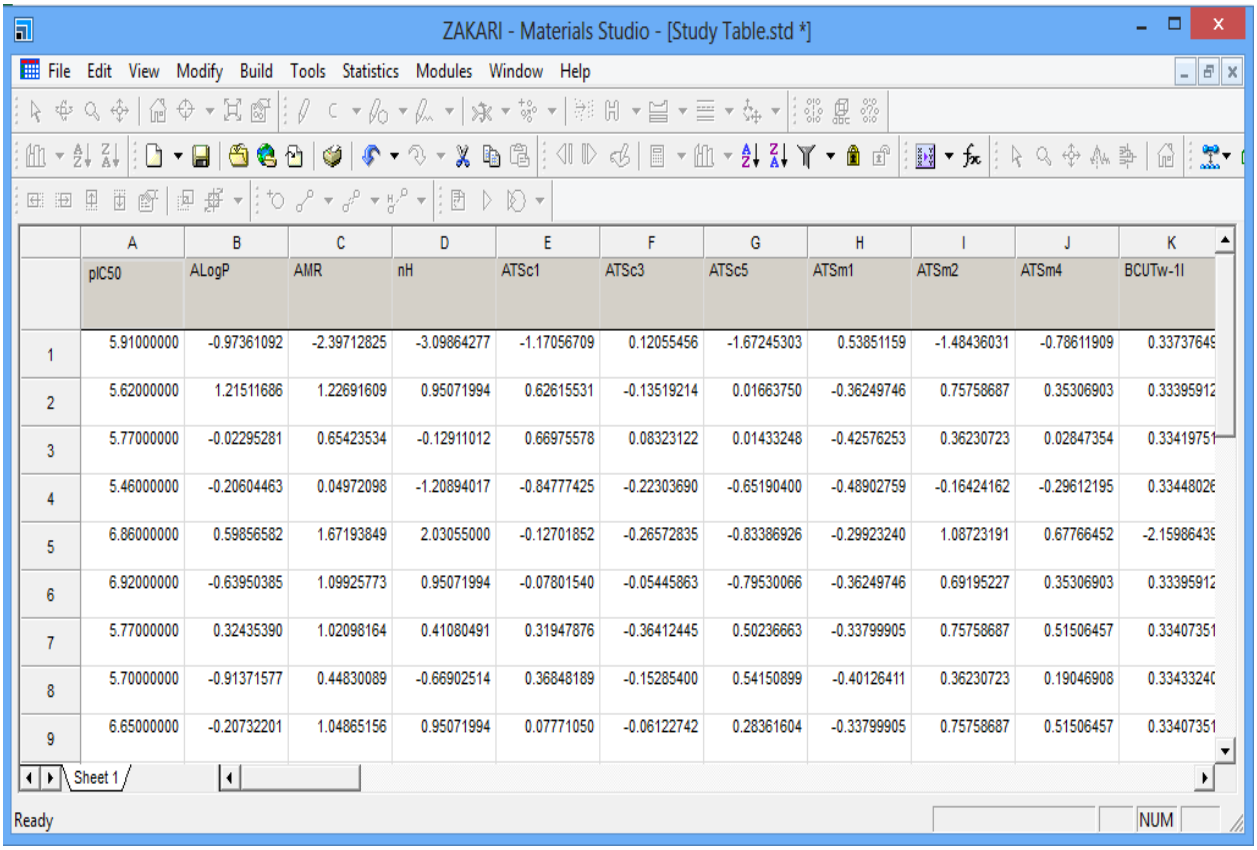


Figure 3.20 Material studio work sheet



	A	B	C	D	E	F	G	H	I	J	K
	pIC50	ALogP	AMR	nH	ATSc1	ATSc3	ATSc5	ATSm1	ATSm2	ATSm4	BCUTw-1l
1	5.91000000	-0.97361092	-2.39712825	-3.09864277	-1.17056709	0.12055456	-1.67245303	0.53851159	-1.48436031	-0.78611909	0.33737649
2	5.62000000	1.21511686	1.22691609	0.95071994	0.62615531	-0.13519214	0.01663750	-0.36249746	0.75758687	0.35306903	0.33395912
3	5.77000000	-0.02295281	0.65423534	-0.12911012	0.66975578	0.08323122	0.01433248	-0.42576253	0.36230723	0.02847354	0.33419751
4	5.46000000	-0.20604463	0.04972098	-1.20894017	-0.84777425	-0.22303690	-0.65190400	-0.48902759	-0.16424162	-0.29612195	0.3344802
5	6.86000000	0.59856582	1.67193849	2.03055000	-0.12701852	-0.26572835	-0.83386926	-0.29923240	1.08723191	0.67766452	-2.1598643
6	6.92000000	-0.63950385	1.09925773	0.95071994	-0.07801540	-0.05445863	-0.79530066	-0.36249746	0.69195227	0.35306903	0.33395912
7	5.77000000	0.32435390	1.02098164	0.41080491	0.31947876	-0.36412445	0.50236663	-0.33799905	0.75758687	0.51506457	0.33407351
8	5.70000000	-0.91371577	0.44830089	-0.66902514	0.36848189	-0.15285400	0.54150899	-0.40126411	0.36230723	0.19046908	0.3343324
9	6.65000000	-0.20732201	1.04865156	0.95071994	0.07771050	-0.06122742	0.28361604	-0.33799905	0.75758687	0.51506457	0.33407351

Figure 3.21 Material studio study table

Figure 3.22 Initial analysis dialogue box

Figure 3.23 Correlation matrix dialogue box

Figure 3.24 Correlation matrix

3.2.7Variable 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 values below the 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 = 0.75. To begin, the QSAR Statistical Software was launched by *double clicking* the software package on the screen of the laptop, this open up a file containing various type of operations that are do able by the software (Figure 3.25). From the folder, *Data Pre Treatment* was double clicked which introduces a data treatment dialogue containing the file (NEW_v-WSP version 1.2) that house the software that will carry out the elimination (Figure 3.26). On double clicking the NEW_v-WSP version 1.2 folder, it opens to display the content (Figure 3.27). The descriptors to be treated were saved in the folder titled *Data,* followed by double clicking the jar file (vWSP*.jar*) present in the vWSP program (Ambure*et al.,* 2015) folder to open -after some few seconds- a window (Figure 3.28) with few a queries. On the *Select Input File*, the *browse* button was *clicked*, an action that result in the popping up of a *choose file* dialogue (Figure 3.29) from which the file saved in *Data* was selected as input file and the *OK* was clicked. On the other hand, for the *Select Output Directory*, the *browse* button was *clicked*, an action that result in the popping up of a *Save* dialogue (Figure 3.30) to save the result of the variable reduction in the folder titled *Output*seen in Figure 3.27, and the output file name was defined and *Save* was clicked. On the "*Enter Variance cut-off*" section,the variance cut-off value based on which the constant variables will be removed was set at *0.0001* which was a default value.

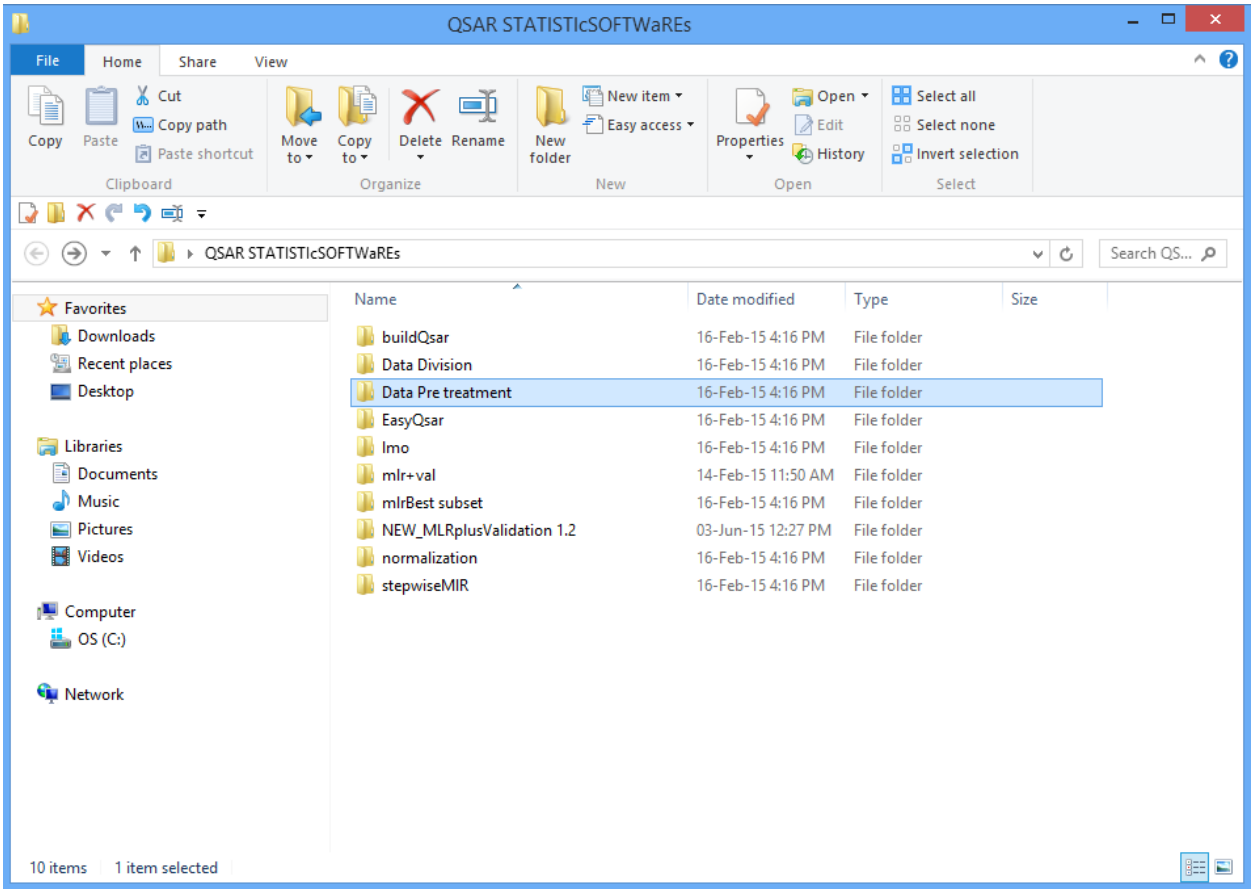


Figure 3.25 QSAR statistical software

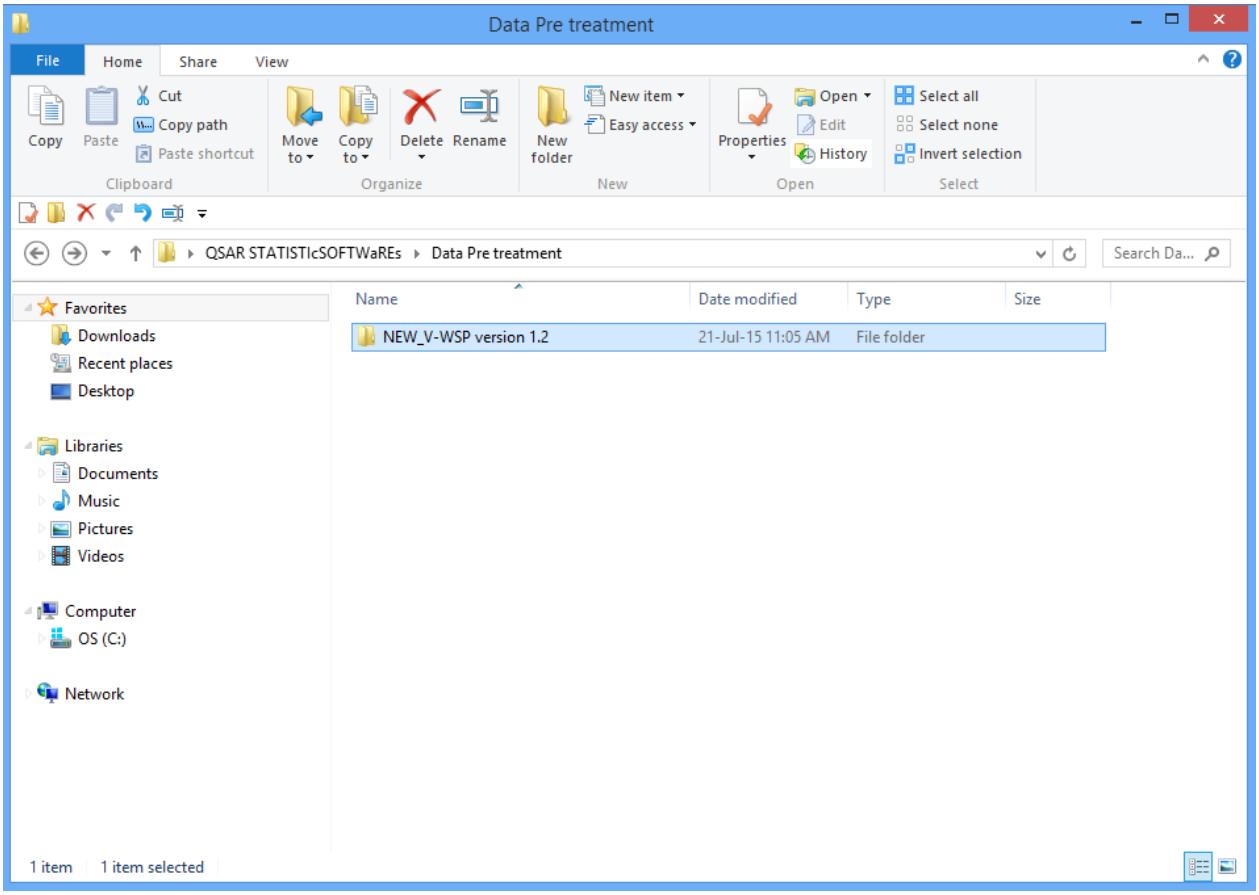


Figure 3.26Data pretreatment dialogue

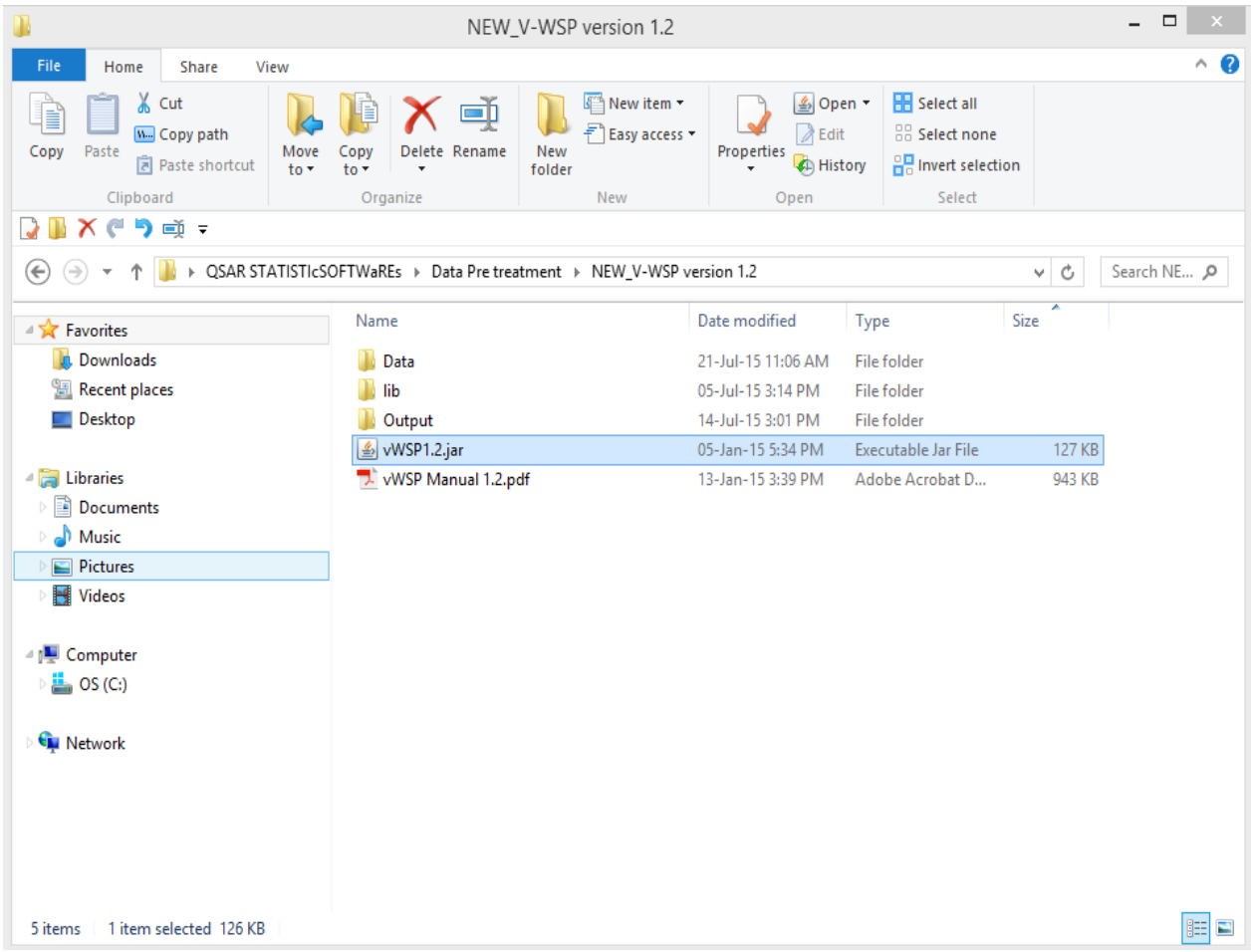


Figure 3.27 NEW_v-WSP dialogue

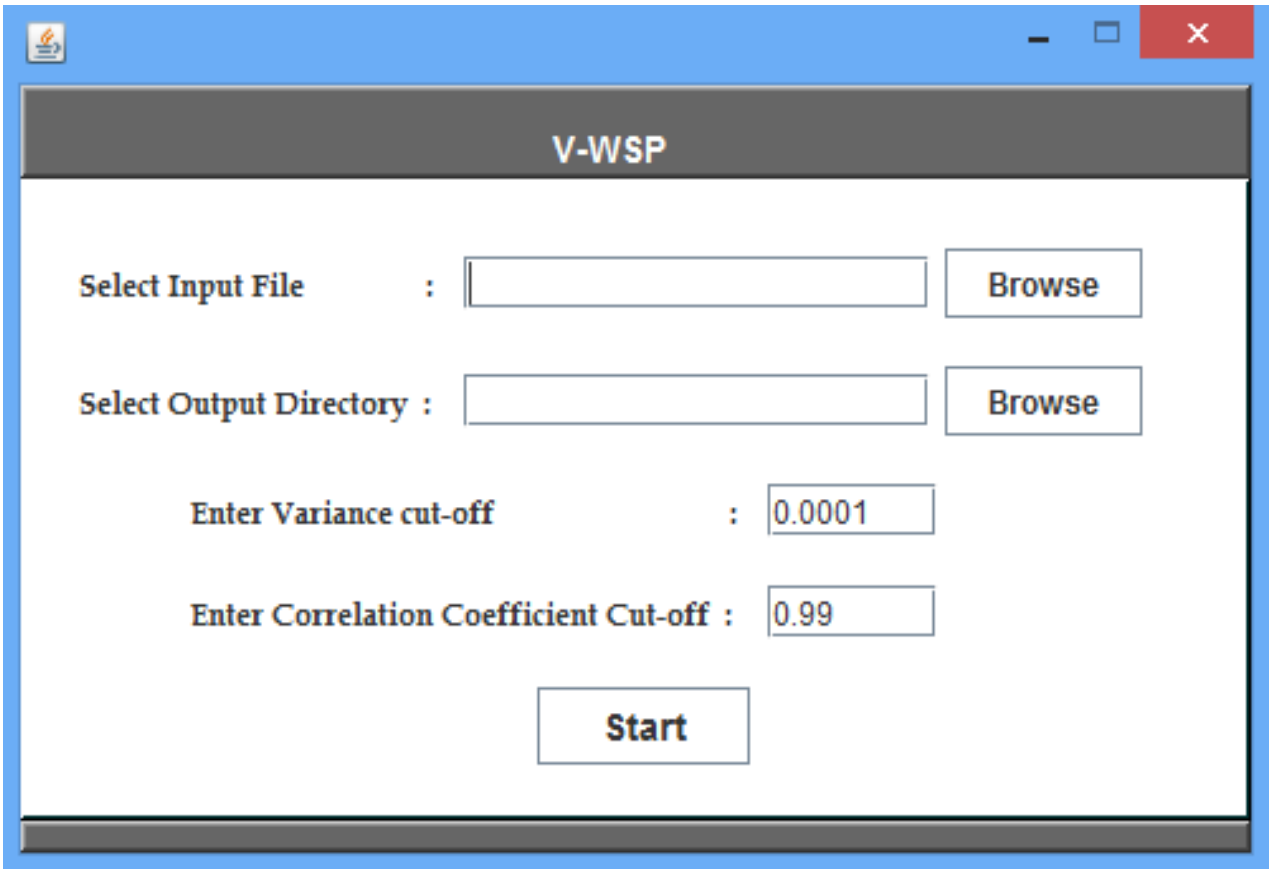


Figure 3.28 Queries window

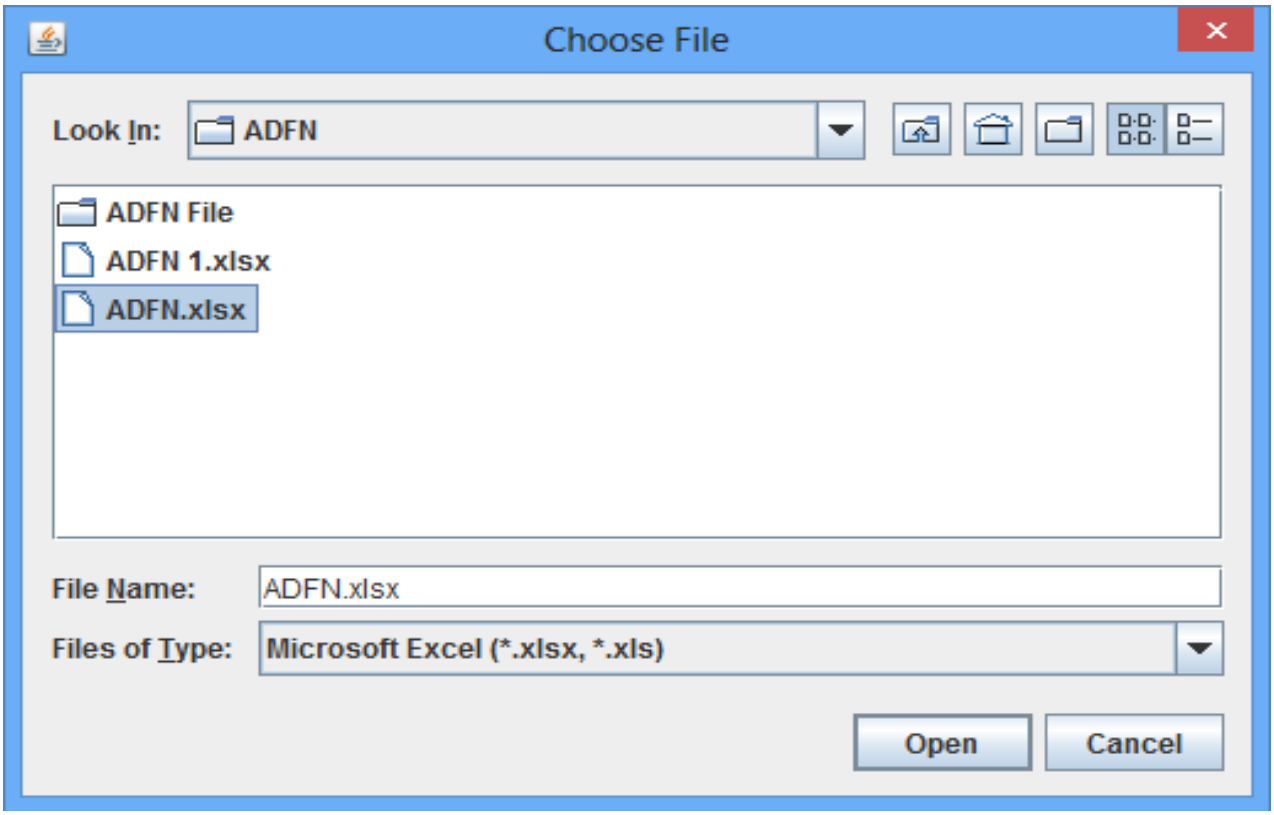


Figure 3.29 A choose file dialogue

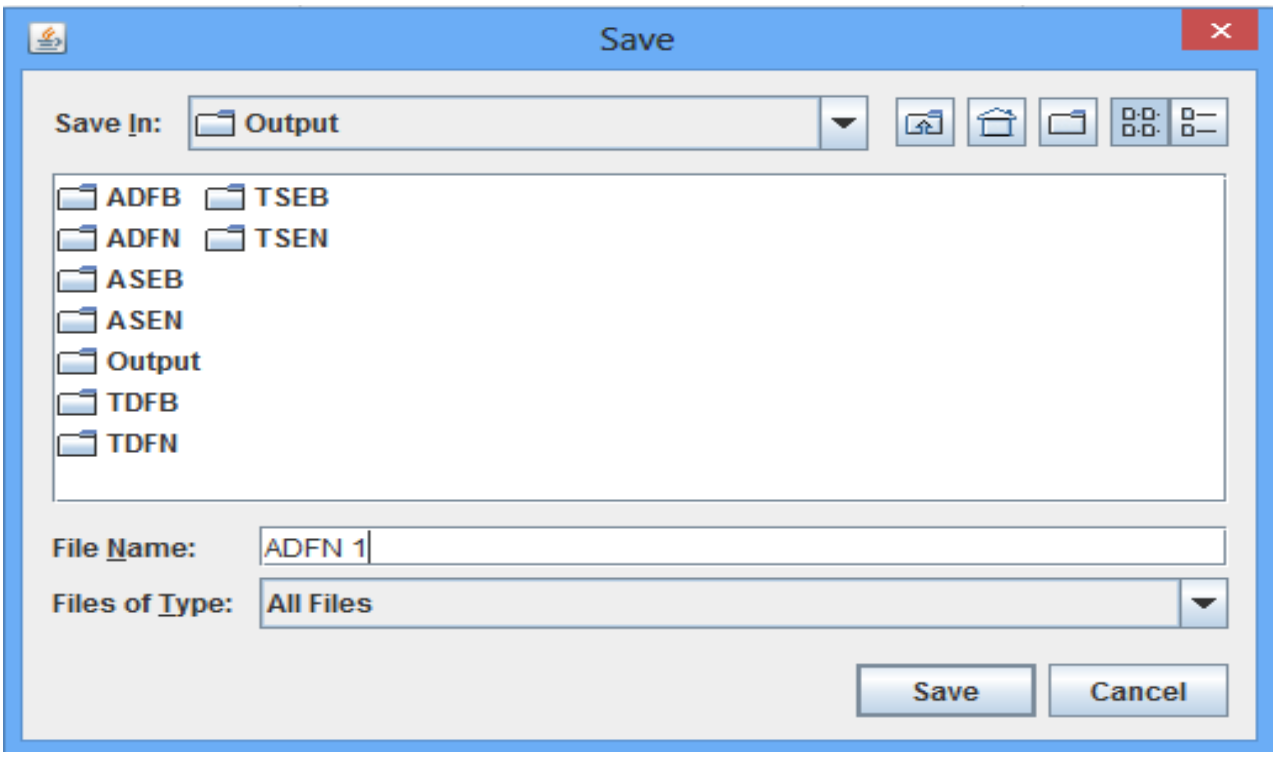


Figure 3.30 A save dialogue

On the "*Enter correlation coefficient cut-off*" sectionthe inter-correlation coefficient cut-off value based on which the inter-correlated variables will be removed, was set to *0.65* (this is to reduce the inter correlation between two independent descriptor values) resulting in Figure 3.31, the *Start* button was then clicked to begin the data treatment by variable reduction. This left a total of 307 descriptors which were then individually standardised, as described in section 3.2.8 and used for QSAR analysis.

3.2.8 Data standardisation

Data Standardisation ensures data reflect how many standard deviations from the average that data lies. Data Standardisation was carried out using the material studio software. After the software was launched, the data were imported into the work sheet (Figure 3.20) after which all the data were highlighted except the pIC_{50} column. Then from the menu bar, *statistics*was*selected*, this brought up Dialogue boxas shown inFigure 3.22, the dialogue contains number of *Analysis Tools*. From which the*Initial Analysis*was selected which again popped up another dialogue box containing number of analytical tools (Figure 3.32) from which*Standardise Data*was *selected* and the whole data were then standardised.

3.2.9 Division of data into training and test sets

Separating data into training and test set is an important part of evaluating QSAR models. It involves dividing the overall data set into two groups i.e. one group for constructing the QSAR model (usually referred to as the training set) while the other group is used to validate the model (usually referred to as the test sets). Commonly, this splitting is performed using random division (Todd *et al.,* 2012).

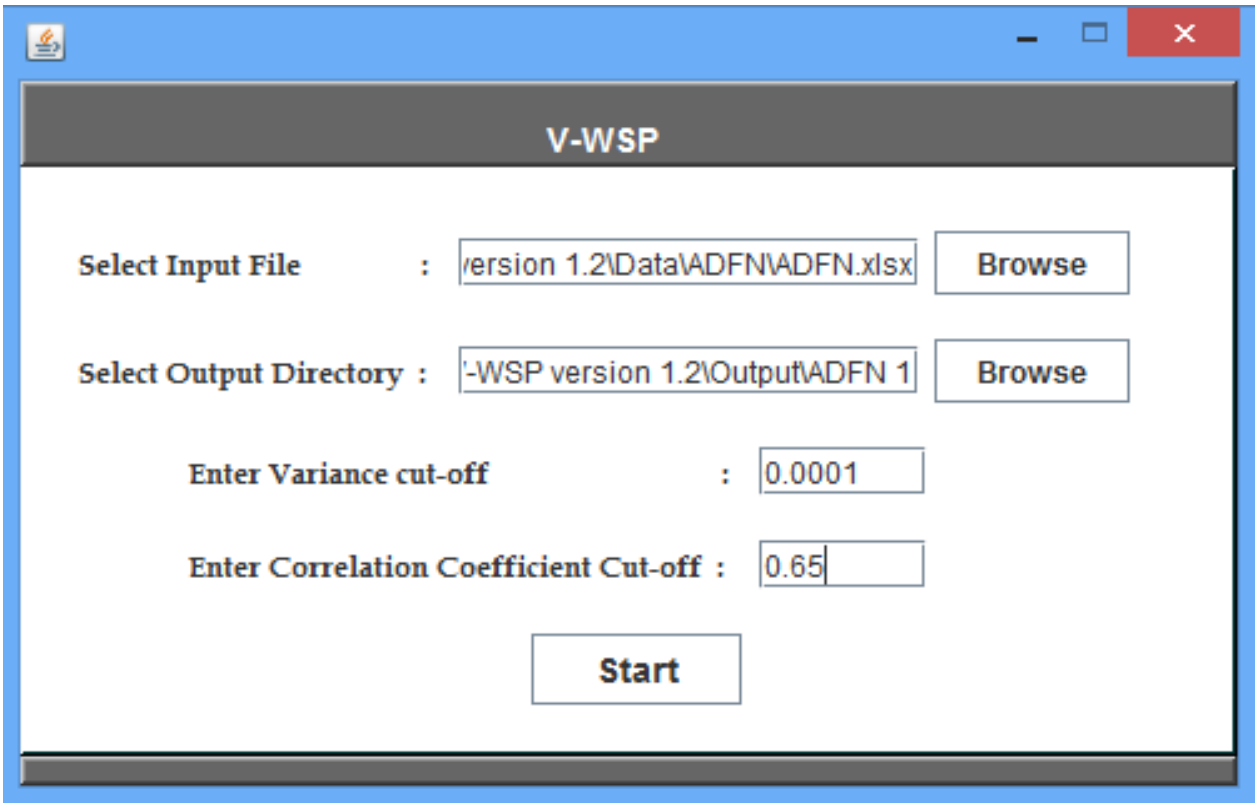


Figure 3.31 A V-WSP dialogue

Figure 3.32Standardise data dialogue

For the size of the training set in comparison with the whole data set, no general guideline exist. Various QSAR groups carry out this task arbitrarily and there is no uniformity in published QSAR reports regarding the ratio of sizes of training and test sets (Roy*et al.,* 2008).In this study each data set was randomly divided into about 75%training set and 25% external validation set. The data set for QSAR study of the activity of naphthylisoquinoline derivatives with *Plasmodium falciparum* comprises of 44

compounds, as compound 9 does not have activity. This constitute 34 training sets and 10 test sets. Also, data set for the QSAR study of the cytotoxicity of naphthylisoquinoline derivatives with L6 cell linecomprises of 44 compounds (34 training sets and 10 test set), as compound 1 is not toxic.The data set for QSAR study of the activity of bisbenzylsoquinoline derivatives with *Plasmodium falciparum* comprises of 20 compounds (16 training sets and 4 test sets). Also, data set for the QSAR study of the cytotoxicity of bisbenzylsoquinoline derivatives with KB cell linecomprises of 18 compounds (since compound numbers 9 and 17 are not toxic), in which 13 compounds were used as training sets and 5 compounds as the test sets. Their experimental pIC_{50}values were determined *in vitro*.

3.2.10 Development of a quantitative structure- activity relationship model

This study was done in four categories;

i) QSAR study of *in-vitro* antimalarial activity (pIC_{50}) of naphthylisoquinoline derivatives against K1 strain of *Plasmodium falciparum*.

ii) QSAR of an*in-vivo* cytotoxicity (pIC_{50}) study of naphthylisoquinoline derivatives against L6 cell line.

iii) QSAR study of the *in-vitro* antimalarial activity (pIC_{50}) of bisbenzylisoquinoline derivatives against K1 strain of *Plasmodium falciparum*.

iv) QSAR of an*in-vivo* cytotoxicity (pIC_{50}) study of bisbenzylisoquinoline derivatives against KB cell line.

3.2.11Descriptor selection and model development

The selection of relevant descriptors, which relates molecular structures of the compounds and their biological activity/cytotoxicity values is an important step in QSAR modelling. For this purpose variable selection genetic algorithm (GA) and multiple linear regression analysis (MLRA) methods were used.

Genetic Algorithm has found numerous applications, specifically, in recent studies as a powerful tool to address many problems in cheminformatics. Genetic algorithm 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).At the end, the MLR was 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 which is an extension of simple linear regression analysis to more than one dimension (Berk, 2003).Multiple linear regression 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 (Deoliveira and Gaudio, 2001) software.The GA began with a population of 100 random models, 1000 iterations to evolution, and the mutation probability of 35%. The selection of most predictive model was done by systematic search among models with between one and six descriptors. The selected equations were analysed 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 program was launched by *double clicking* the icon  present on the desktop of the laptop, which result in the BuildQsar worksheet shown in Figure 3.33. From the Menu Bar,*File menu* was selected (Figure 3.34) then the *New…..*Submenu was selected, this immediately popup a *Create New Dataset*dialogue (Figure 3.35).On *the Dataset size*section the number of activities,Descriptors and compounds were specified accordingly. The *Ok* was *clicked* which brought up an excel sheet 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, and the values of the activity were copied and pasted in the first column(yellow) of the excel sheet, then the descriptor values were copied and pasted in the subsequent (green) columns (Figure 3.36). On the Menu Bar,the *QSAR*was selected, which display a dialogue containing some operations (Figure 3.37). On placing the cursor on*Variable Selection*, two options were displayed; *Systematic Search…* and *Genetic Algorithm…..* (Figure 3.38). On selecting the*Genetic Algorithm*, a genetic search dialogue appeared (Figure 3.39) on which the *Number of descriptors per model, No of generation* and *Models per generation* were all specified. For the QSAR study of *in vitro* antimalarial activity (pIC_{50}) of naphthylisoquinoline derivatives against K1 strain of *Plasmodium falciparum*, the number of descriptors per model was first set at six, five, four, three, two and one respectively, since the recommended ratio of number of compounds to the number of molecular descriptors should be at least 5:1 (Walker, 2003).The same for the QSAR of an*in vitro* cytotoxicity (pIC_{50}) study of naphthylisoquinoline derivatives against L6 cell line, the number of descriptors per model was first at six, five, four, three, two and one respectively. For the QSAR study of the *in vitro* antimalarial activity (pIC_{50}) of bisbenzylisoquinoline derivatives against K1 strain of *Plasmodium falciparum*, the number of descriptors per model was set atfour, three, two and one respectively.

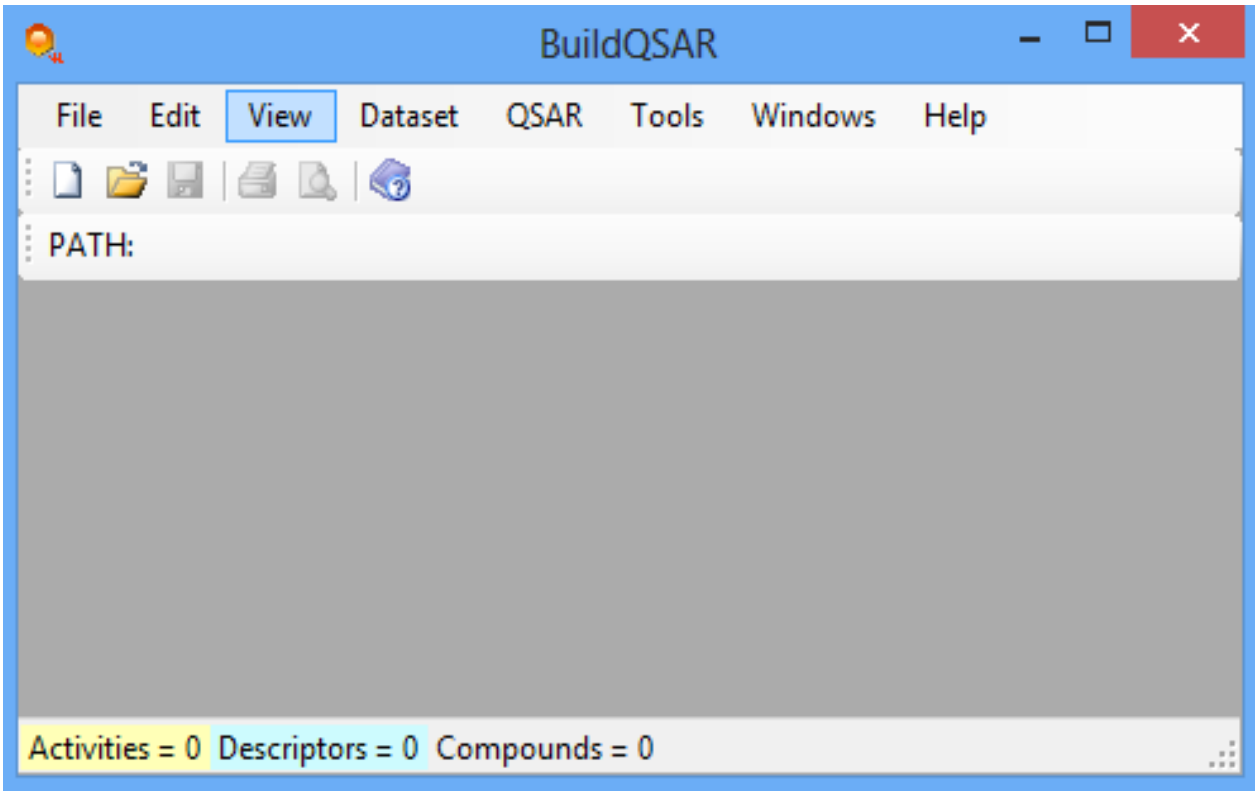


Figure 3.33 BuildQsar worksheet

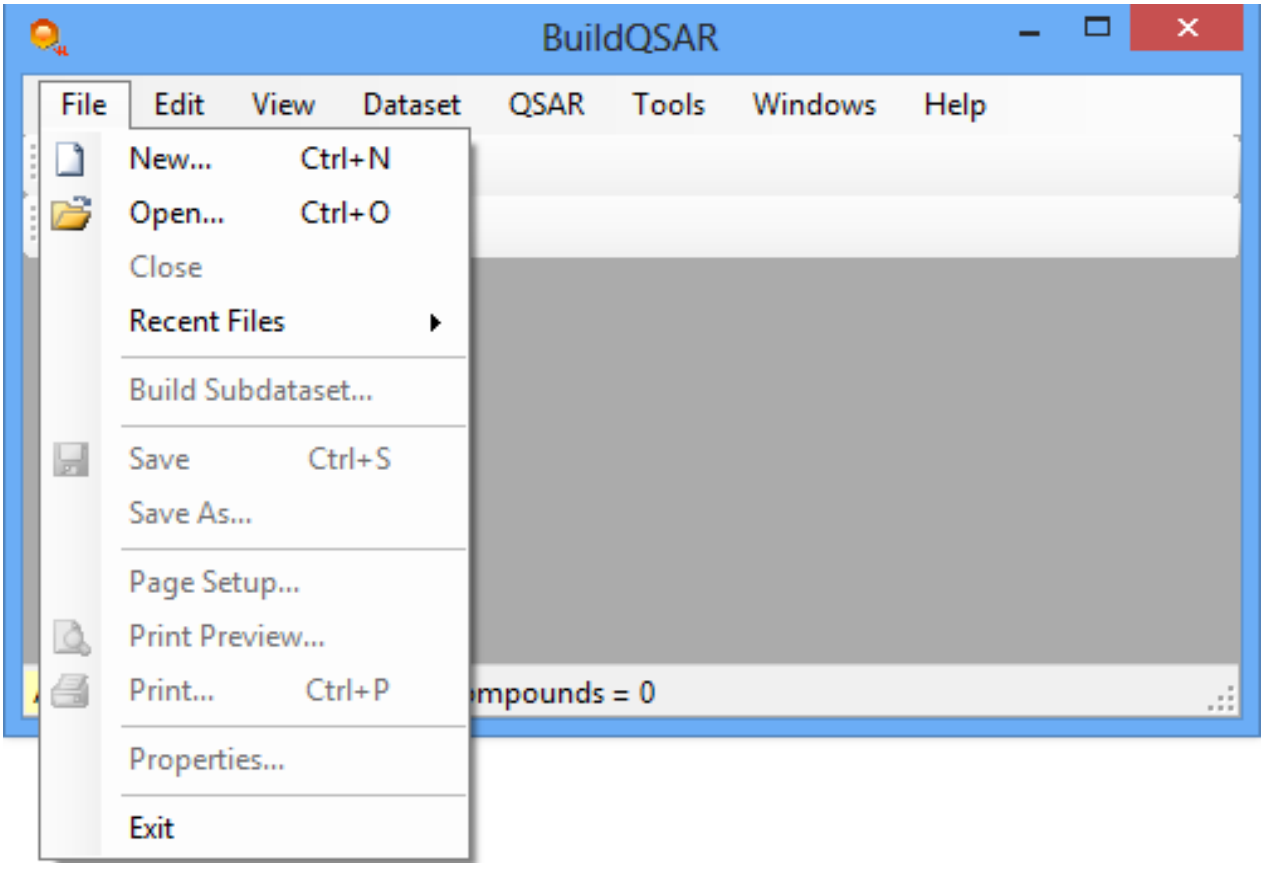


Figure 3.34 A file dialogue

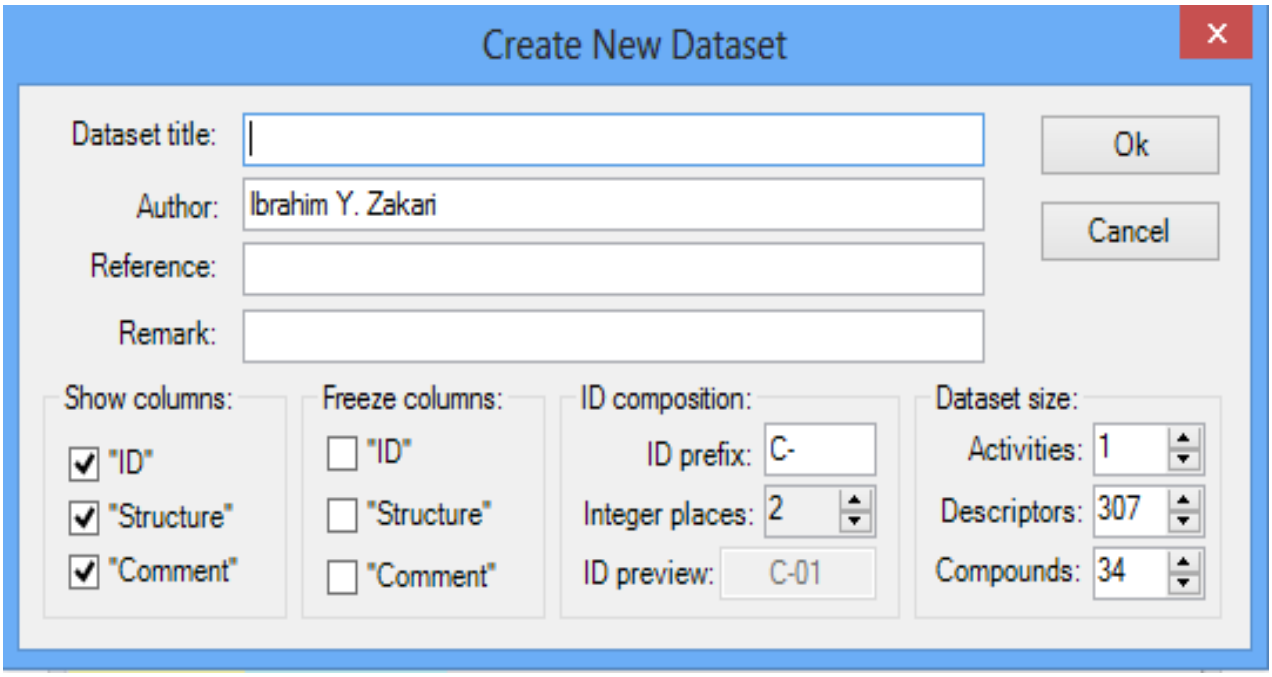


Figure 3.35 Create new data dialogue

ID	Structure	Comment	Y1	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
C-01			5.91	-1.41778447	-0.43408923	2.06941648	-1.45841531	0.70068027	-0.80194947	-0.70004008	2.89678292	0.25378311	1.24144695	-2.19426154
C-02			5.62	-0.58350882	-0.0549464	-0.41544653	0.56589853	-0.58946592	0.25229085	0.21531813	-0.70866194	-0.15478674	-0.29809139	0.70727667
C-03			5.77	0.59331018	-0.07475328	-0.32592108	0.78453863	-0.50835665	-0.15166146	0.54217047	-0.62409845	-0.06513467	-0.98949645	-0.1921048
C-04			5.46	1.1136245	0.11026713	0.09086213	0.06887629	-0.65186777	-0.28236264	0.41635691	-0.41499994	0.43700205	-1.02002047	-0.36106913
C-05			6.86	1.85972398	-0.01148726	-0.03118651	1.25341373	0.63658724	-0.16021166	0.44658139	-0.21939823	0.76696201	1.55022155	0.54012431
C-06			6.92	-0.59703122	-0.08798762	-0.32628461	1.33359939	0.45056808	-0.37400489	0.63491706	-0.26291167	0.62260732	-0.12041347	-0.45036854
C-07			5.77	0.09209132	-0.00768678	-0.23929074	0.56595608	-0.75757	0.07141977	0.36960479	-0.75648762	-0.07627962	-0.07900104	0.79587119
C-08			5.7	-0.50452748	-0.12163893	-0.40444022	0.71857534	-0.65251335	-0.31562793	0.68028806	-0.64032611	-0.04179818	-0.62850407	-0.12813948
C-09			6.65	-0.09122216	-0.05197058	-0.34028701	-1.70386055	0.22364358	2.32685059	-2.16115103	-0.17218908	-1.55132587	-0.2068012	1.58444919
C-10			6.58	1.53172426	-0.00379583	-0.14817666	0.79084531	-0.67866386	-0.28537205	0.69446049	-0.69437501	-0.10405627	0.04958387	0.36790516
C-11			4.72	-0.79173039	-0.36947871	1.63956956	-2.06548739	0.78828784	0.55012668	-2.41212807	1.69808773	-2.76102552	2.19676617	1.23002974
C-12			5.73	0.98181987	-0.77201649	-0.57877048	0.09496804	0.47049292	-0.56112412	0.36117778	0.38818241	0.08990566	0.46257077	-0.54287667

Figure 3.36 BuildQSAR raw data

Figure 3.37 BuildQSAR variable selection dialogue

Figure 3.38BuildQSARtools bar

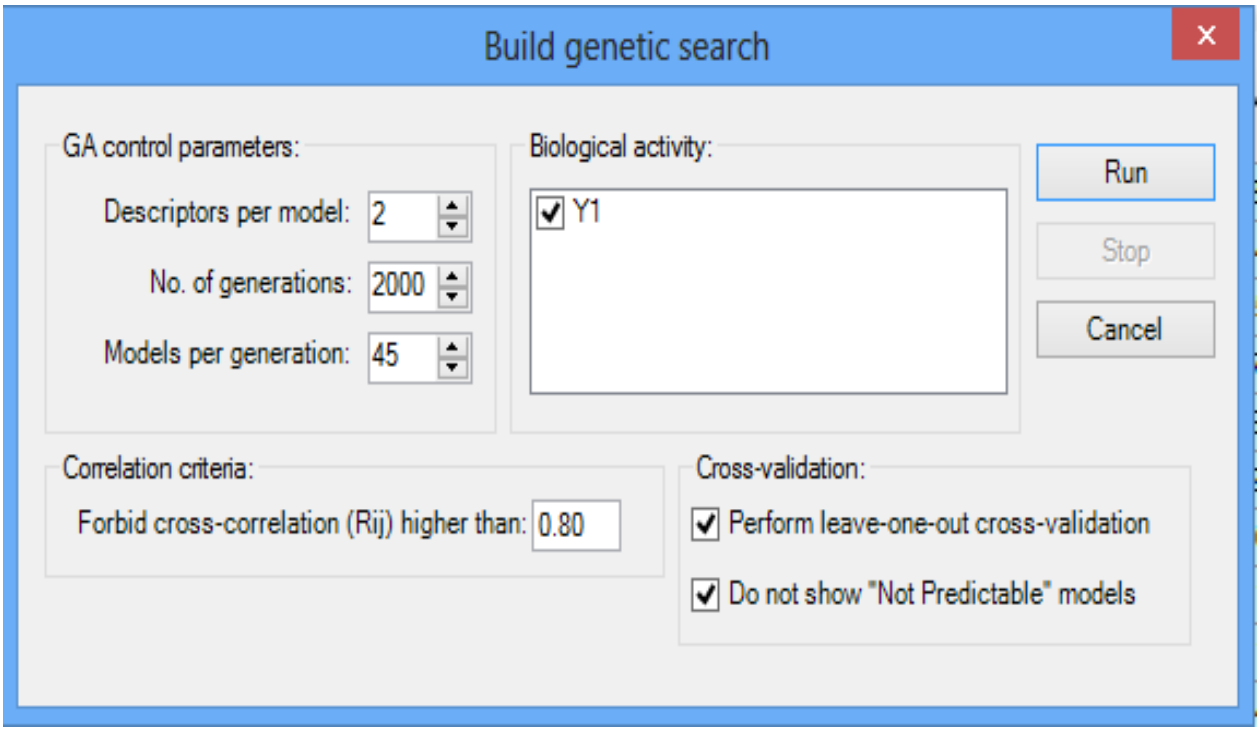


Figure 3.39 Genetic search dialogue

For the QSAR of an*invitro* cytotoxicity (pIC_{50}) study of bisbenzylisoquinoline derivatives against KB cell line, the number of descriptor per model was set at four, three, 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 *Run* in Figure 3.39 was clicked which took some seconds in performing the pairing to generate the models in each case. The results of the genetic search as well as their statistical parameters were display in the *Variable selection results* (Figure 3.40). From the variable selection results, four of the best models were selected based on the criteria suggested by Ravichandran *et al*., 2011.

3.2.12 Model validation

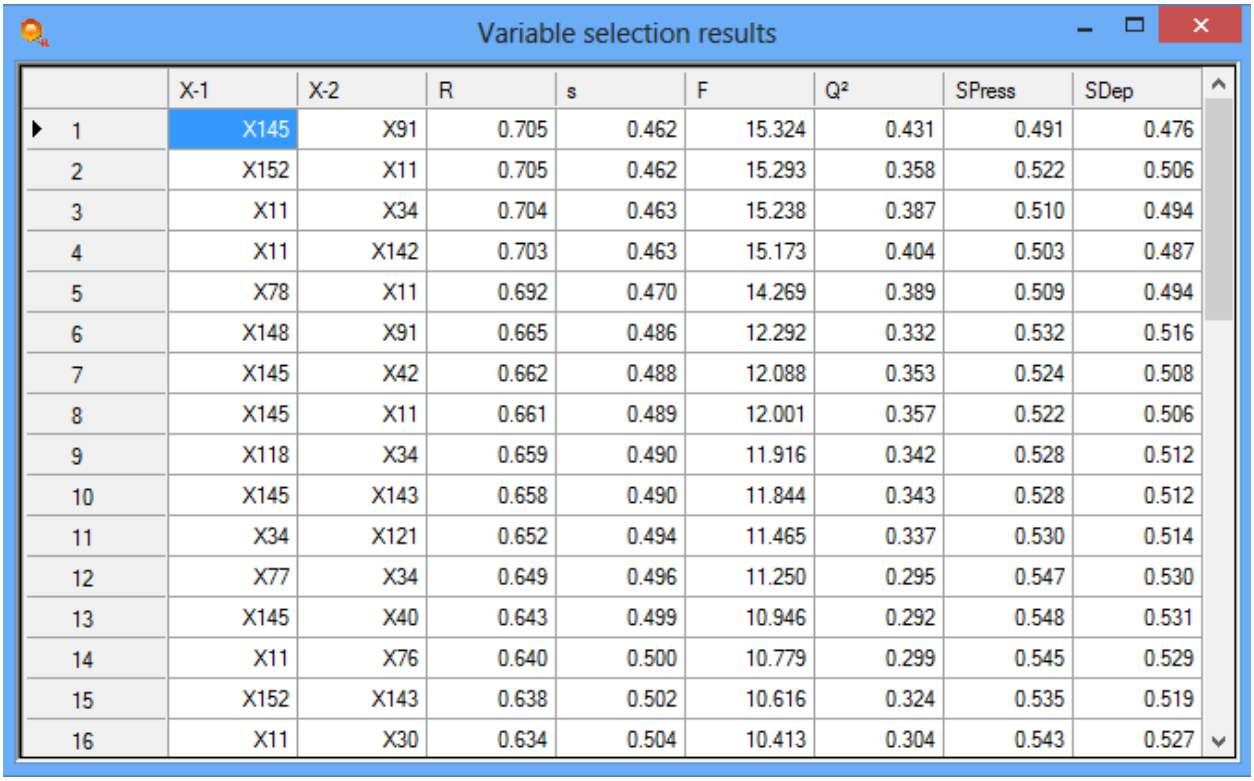
3.2.12.1 Internal validation

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 complete, the mean is taken of all the Q^2 values (Ravichandran *et al*., 2011). The Q^2_{LOO} was calculated according to equation (10) using the Material Studio software 7.0

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

where $N_{tr}$is the total number 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 as decribed insection 3.2.6 up untill the procedure that lead to the creation of Figure 3.21.



	X-1	X-2	R	s	F	Q²	SPress	SDep
1	X145	X91	0.705	0.462	15.324	0.431	0.491	0.476
2	X152	X11	0.705	0.462	15.293	0.358	0.522	0.506
3	X11	X34	0.704	0.463	15.238	0.387	0.510	0.494
4	X11	X142	0.703	0.463	15.173	0.404	0.503	0.487
5	X78	X11	0.692	0.470	14.269	0.389	0.509	0.494
6	X148	X91	0.665	0.486	12.292	0.332	0.532	0.516
7	X145	X42	0.662	0.488	12.088	0.353	0.524	0.508
8	X145	X11	0.661	0.489	12.001	0.357	0.522	0.506
9	X118	X34	0.659	0.490	11.916	0.342	0.528	0.512
10	X145	X143	0.658	0.490	11.844	0.343	0.528	0.512
11	X34	X121	0.652	0.494	11.465	0.337	0.530	0.514
12	X77	X34	0.649	0.496	11.250	0.295	0.547	0.530
13	X145	X40	0.643	0.499	10.946	0.292	0.548	0.531
14	X11	X76	0.640	0.500	10.779	0.299	0.545	0.529
15	X152	X143	0.638	0.502	10.616	0.324	0.535	0.519
16	X11	X30	0.634	0.504	10.413	0.304	0.543	0.527

Figure 3.40 Variable selection results dialogue

But unlike in section 3.2.6, where the whole values were copied and pasted in study table, in the validation procedure, only the activity and the corresponding descriptors of the training set were copied and pasted to give Figure 3.21. To carry out the internal validation,the column containing the values of the activities was highlighted(Figure 3.41), on the Menu bar, the *Statistics*was selected, resulting in a drop-down menu (Figure 3.42) from which *Model Building*was selected, resulting in another drop-down menu (Figure 3.43) from which *Multiple Linear Regression*was selected. This produced the dialogue shown in Figure 3.44. From Figure 3.44, the boxes corresponding to descriptors of a model on the *Variable Selection*was first checked. Next, the *Stepping* option on Figure 3.44, was clicked, a dialogue box appeared (Figure 3.45) on which in the *Type of Stepping*section, 2-way (bidirectional) was selected, and both*F to enter*(determines how significant the contribution of a variable in the regression equation has to be in order for it to be added to the equation)and *F to leave*(determines how insignificant the contribution of a variable in the regression equation has to be in order for it to be removed from the equation) were set at 4 (the usual rule-of-thumb value for the significance at 5% level), on the *Exclude if Correlation*greater section 0.8 (beyond which multicollinearity sets in) was imputed. The *Maximum number of steps* was set at 10. Next the *Validation* option on the dialogue was selected and the dialogue now appeared as shown in Figure 3.46.The *Omit single rows in turn*section was selected, *use random trials* was selected and the number of trials set at 5,*use randomisation of response*variable was selected and the number of trials set at 10. The *Ok*button was *clicked* and the result was displayed on the screen (Figure 3.47).

3.2.12.2 External validation

Figure 3.41 Training sets working sheet

Figure 3.42 Statistics menu brought down

Figure 3.43 Model building menu brought down

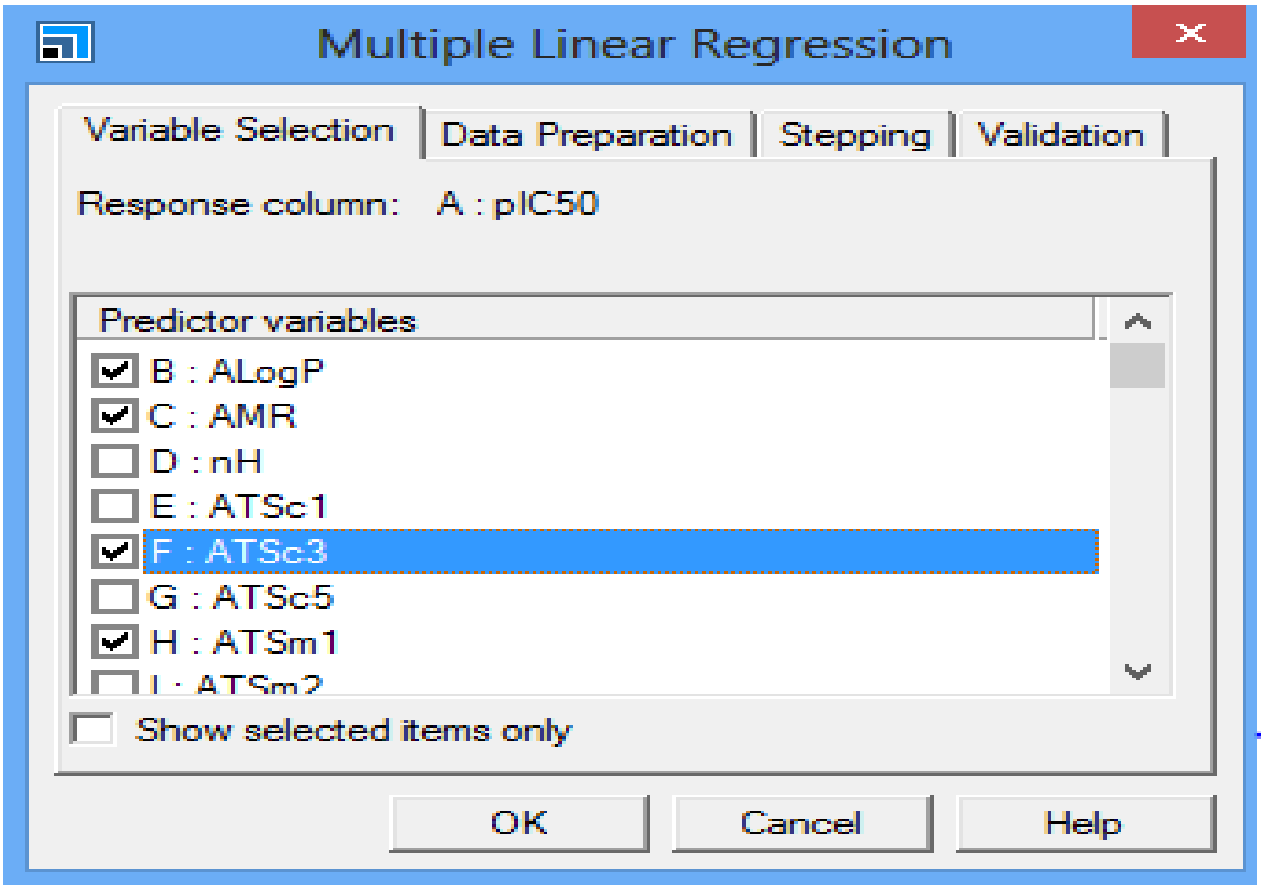


Figure 3.44 Multiple linear regressiondialogue

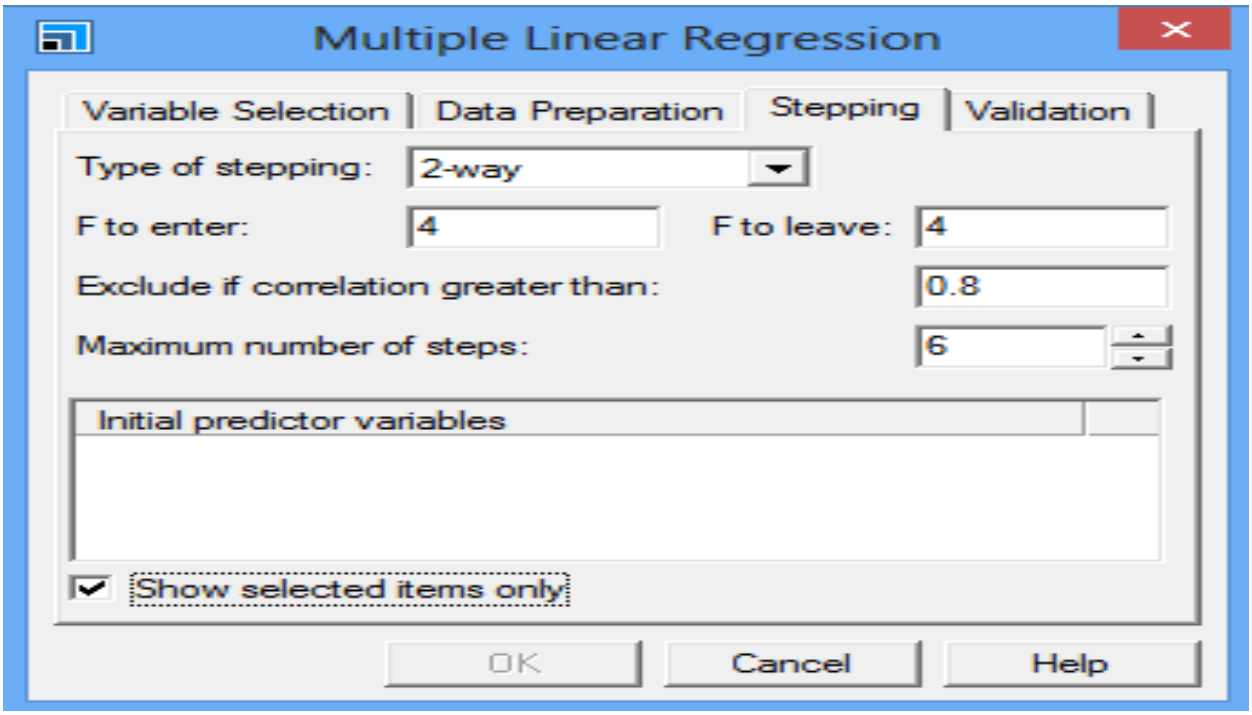


Figure 3.45Stepping of multiplelinear regression dialogue

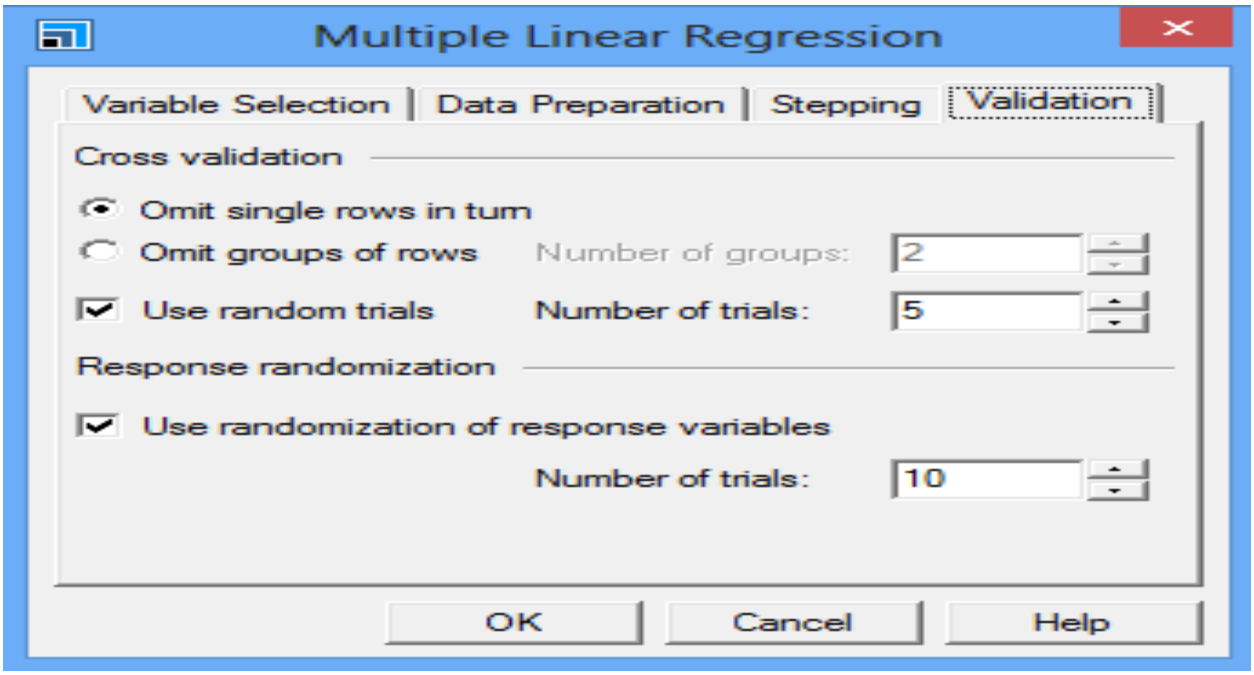
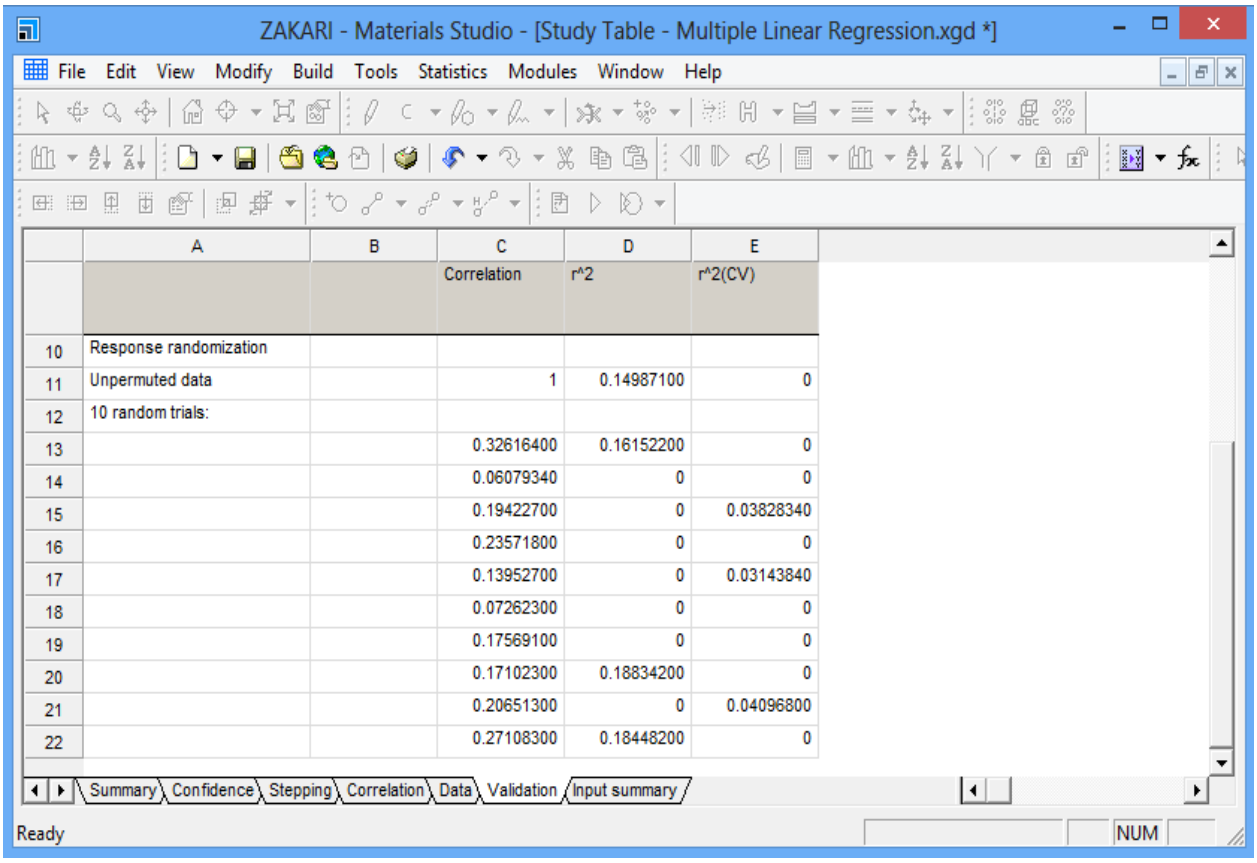


Figure 3.46 Validation of multiple linear regression dialogue



	A	B	C	D	E
			Correlation	r^2	r^2(CV)
10	Response randomization				
11	Unpermuted data		1	0.14987100	0
12	10 random trials:				
13			0.32616400	0.16152200	0
14			0.06079340	0	0
15			0.19422700	0	0.03828340
16			0.23571800	0	0
17			0.13952700	0	0.03143840
18			0.07262300	0	0
19			0.17569100	0	0
20			0.17102300	0.18834200	0
21			0.20651300	0	0.04096800
22			0.27108300	0.18448200	0

Figure 3.47 Model internal validation result dialogue

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 was not only based on R^2, Q^2 and R^2_{pred} (Golbraikh andTropsha, 2002), as they might mislead the search for the ideal predictive model. Therefore, Ravichandran *et al*., 2011 suggested additional statistical parameters that must be fulfilled in model selection. 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. $\left[\frac{|R^2 - R_0^2|}{R^2} < 0.1\right]$
5. $0.85 < k < 1.15$
6. $R_m^2 \geq 0.5$

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), slopes k of regression lines through the origin. These additional statistical parameters were calculated, with the aid of the program MLRplusValidation (Ambure *et al.,* 2015). To launch the MLRplus Validation software, the *NEW_MLRplusValidation 1.2* folder in Figure 3.25 was double clicked which lead to the display of Figure 3.48. The training set that was to be used to

development the models was arranged such that the column of the descriptors were sandwich between the compounds numbers and their activities as shown in Table 3.3, the test sets to be used to validate the model was also arranged in the same format as in the training sets describe in Table 3.3. Both the training and the test sets were saved in the *Data* component of Figure 3.25 using different Excel names.Then the executable jar file contained in the folder Figure 3.48, was *double clicked*. This produced the dialogue in Figure 3.49. The rectangular box in front of *Select Training Set* in Figure 3.49 was clicked and the *Browse* in front of the box was selected. This popup a dialogue (Figure 3.50) from which the saved training set was selected and the *Open* was clicked to pick the training set. The rectangular box in front of *Select Test Set* in Figure 3.49 was clicked and the *Browse* in front of the box was selected. This popup a dialogue (Figure 3.51) from which the saved test set was selected and the *Open* was clicked to pick the test set. The rectangular box in front of *Select Output File* in Figure 3.49 was clicked and the *Browse* in front of the box was selected. This popup a dialogue (Figure 3.52) in which the result of the validation was saved to the *Output* component of Figure 3.48. Save was clicked to save the file. The Euclidean based applicability domain was checked and *Submit* was clicked to run the validation. The result of the validation was found by double clicking the Output component of Figure 3.48. The External Validation Parameters and also Ravichandran *et al*., acceptable model criteria was displayed (Figure 3.53).

3.2.13 Y-Randomisation

The final Model was validated by the Y-randomisation (randomisation of response). This approach is widely used to validate the robustness of QSAR models (Baumann**,** 2003**)**.

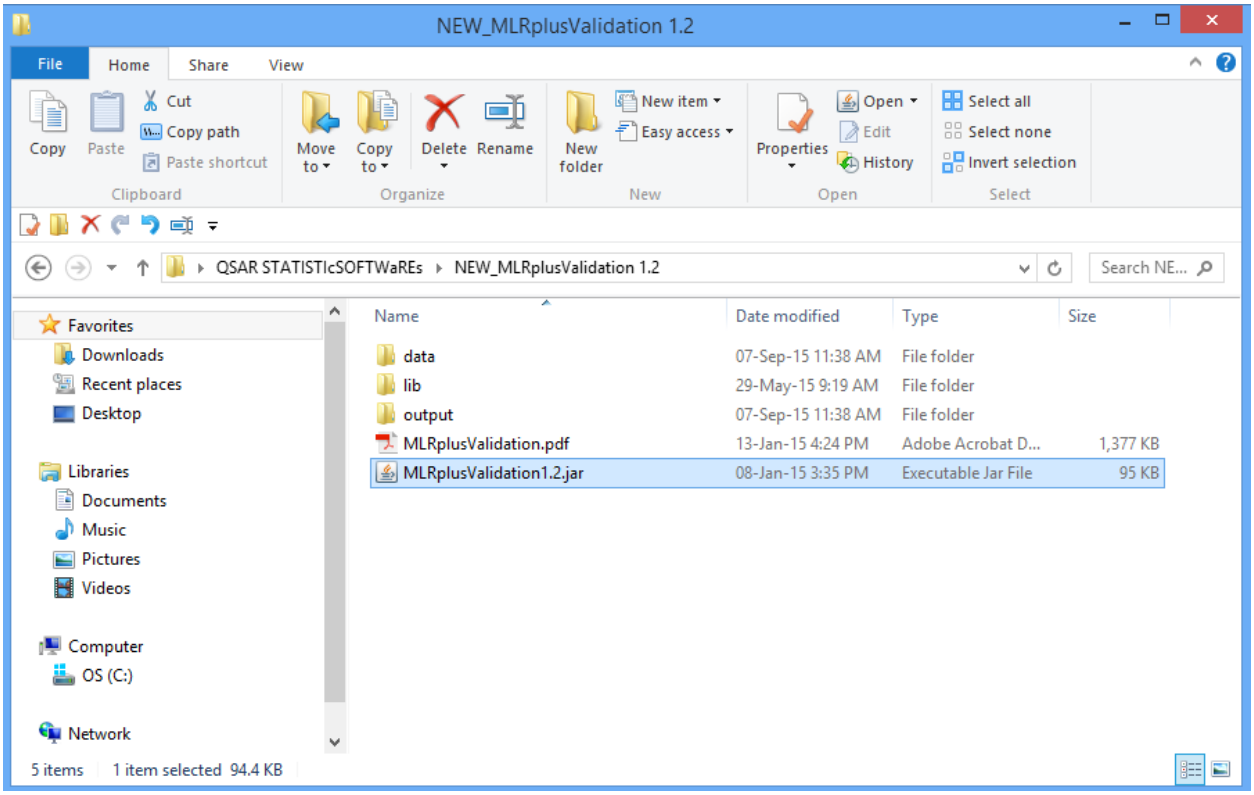


Figure 3.48 MLRplus validation folder

Table 3.3 Format of training set compounds and their activities in the data folder of MLR plus validation

Comp.No.	BCUTc-1l	nHBd	nssCH2	Wnu1.unity	μ	LogP	pIC_{50}
2	-1.2165	0.1703	0.3624	-2.1203	-0.1551	-0.9751	5.91
3	-0.2447	-0.9488	-0.8283	1.2522	0.5779	1.2469	5.62
4	-0.2445	-0.9488	0.3624	1.6644	0.4917	0.7777	5.77
5	-1.2318	1.2893	0.3624	2.2128	-1.1038	0.3085	5.46
6	1.6465	0.1703	1.5531	1.1259	0.7504	2.1852	6.86
7	-0.2450	0.1703	2.7438	-0.1429	-0.3168	1.7072	6.92
8	-0.2388	1.2893	-0.8283	1.1969	0.8366	0.6715	5.77
9	-0.2386	1.2893	0.3624	1.0064	0.9552	0.1934	5.70
10	-0.2390	2.4084	0.3624	1.1686	0.9121	0.1049	6.65
11	-0.2388	2.4084	1.5531	1.0472	0.9767	-0.3732	6.58
12	-0.2361	-0.9488	-0.8283	-0.1497	0.0820	-0.8512	4.72
13	-0.2370	-0.9488	-0.8283	-0.1870	-0.1875	-0.0987	5.73
14	1.6577	0.1703	-0.8283	0.1601	-0.0258	-0.3289	6.09
15	-0.2371	-0.9488	-0.8283	-0.7366	-0.0689	0.3351	5.60
16	1.6575	0.1703	-0.8283	-0.4857	0.1144	0.0960	6.36
17	-0.2400	0.1703	-0.8283	-0.4016	-0.6079	-0.4440	4.97
18	-1.3437	1.2893	-0.8283	-0.0688	-0.3815	-0.6741	5.00
19	-0.2371	-0.9488	-0.8283	-0.2338	0.3515	0.3970	5.19
20	1.6575	0.1703	-0.8283	0.1074	0.0389	0.1580	6.27
22	1.6562	0.1703	-0.8283	-0.4398	0.5348	0.4855	6.34
23	-0.2375	-0.9488	-0.8283	-0.5725	-0.2198	0.7865	5.10
24	1.6575	0.1703	-0.8283	0.2092	-0.8019	0.1580	6.07
25	-0.2375	-0.9488	-0.8283	1.5363	-0.1875	0.7865	5.16
26	1.6566	0.1703	-0.8283	1.4362	-0.0581	0.5475	6.12
27	-0.2375	-0.9488	-0.8283	-0.5725	-0.2198	0.7865	5.93
29	-0.2984	-0.9488	-0.8283	-1.0332	0.3515	-1.4974	4.76
33	-0.2986	0.1703	0.3624	-1.2582	-2.5375	-1.6833	4.81
35	-0.2652	-0.9488	1.5531	0.3738	0.7288	0.9193	5.96
37	-0.2645	0.1703	0.3624	-1.3469	0.7611	-0.4794	5.87
39	-0.2371	1.2893	0.3624	0.8792	-0.0366	0.3351	6.02
42	-0.2701	-0.9488	1.5531	1.2080	2.7877	-3.3299	4.96
43	-0.2722	0.1703	0.3624	-0.8001	-2.2141	-1.5683	4.58

44	-0.2722	0.1703	0.3624	-0.8688	-0.2737	-0.3732	5.97
46	-1.8247	2.4084	1.5531	0.5713	-1.5781	-0.9397	4.95

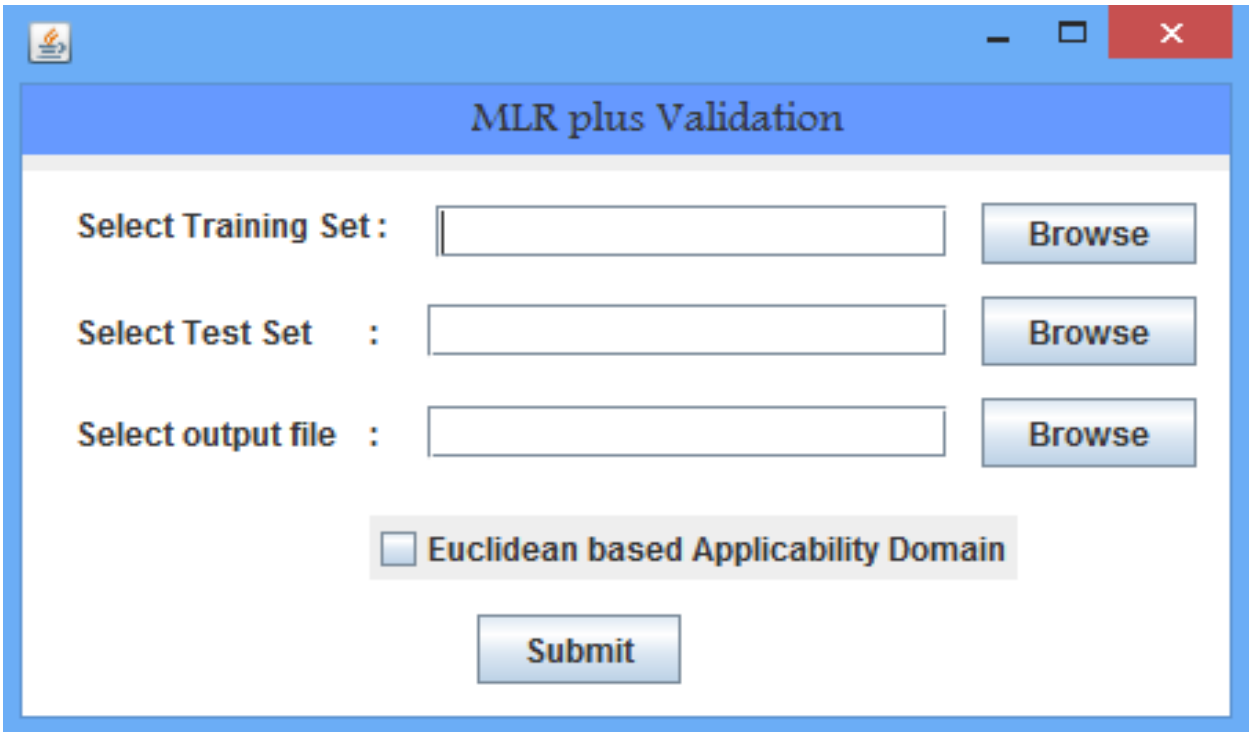


Figure 3.49 MLR plus validation dialogue

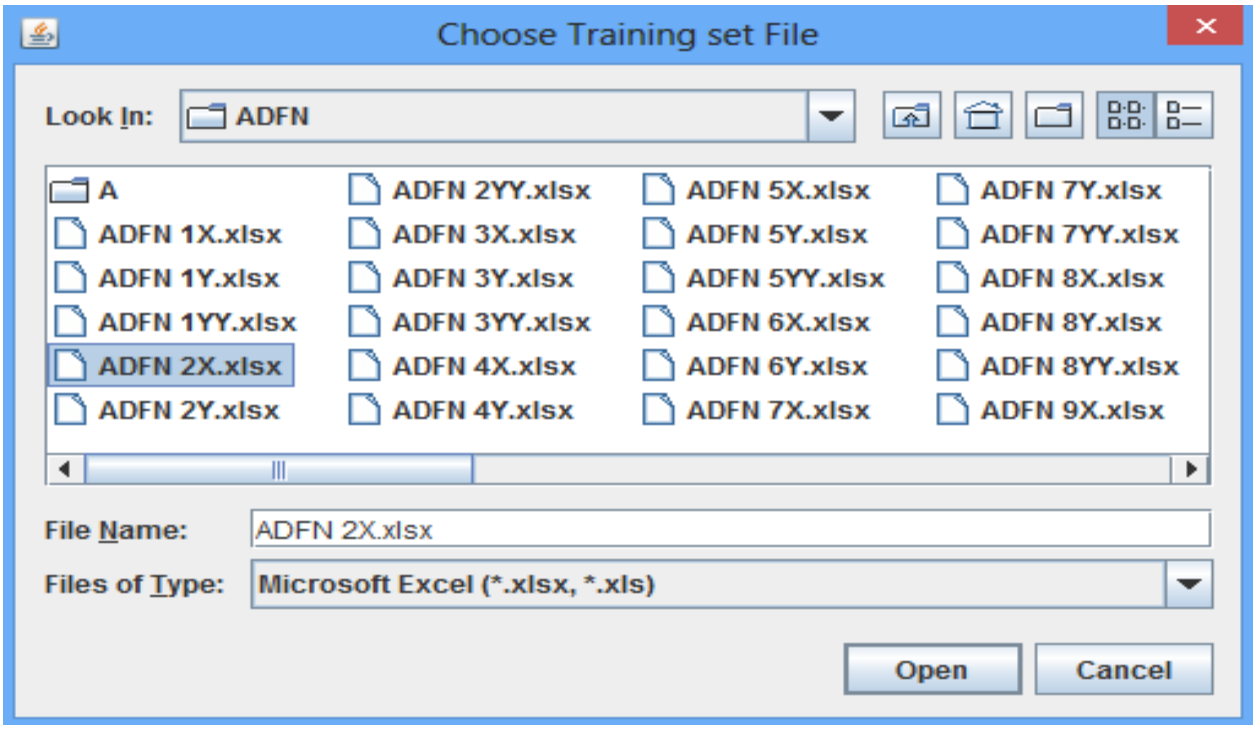


Figure 3.50 Training set file dialogue

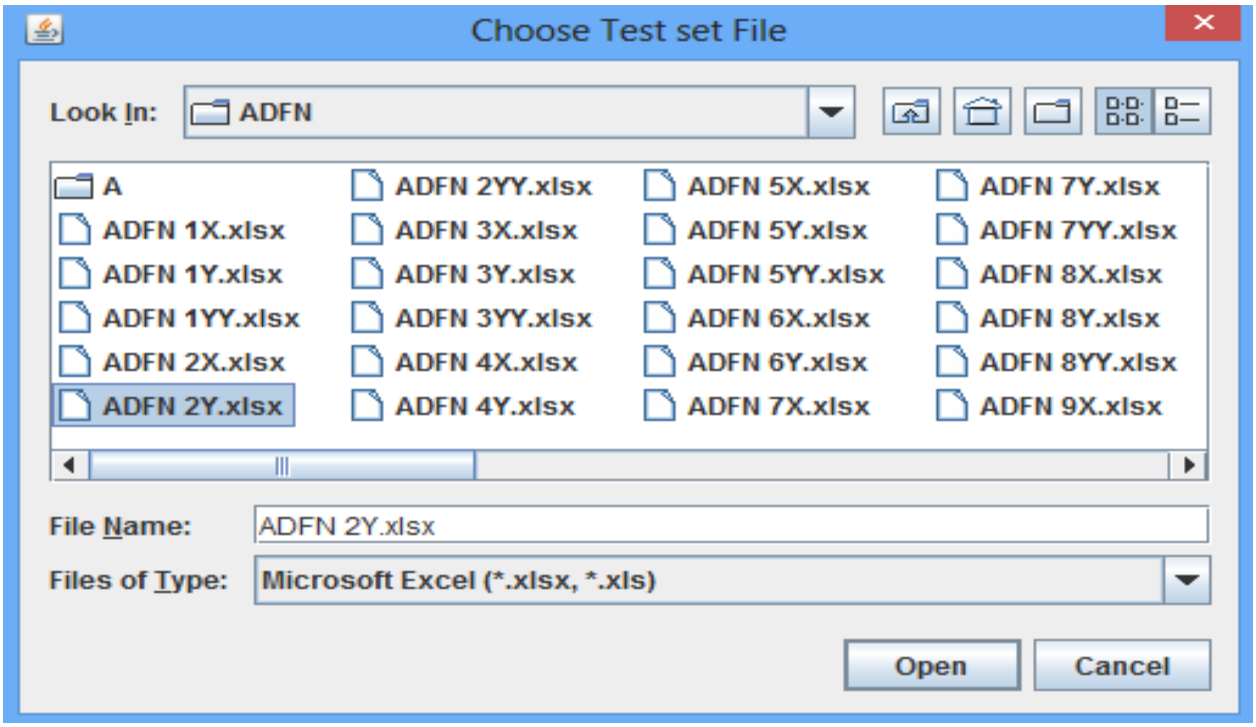


Figure 3.51 Test set file dialogue

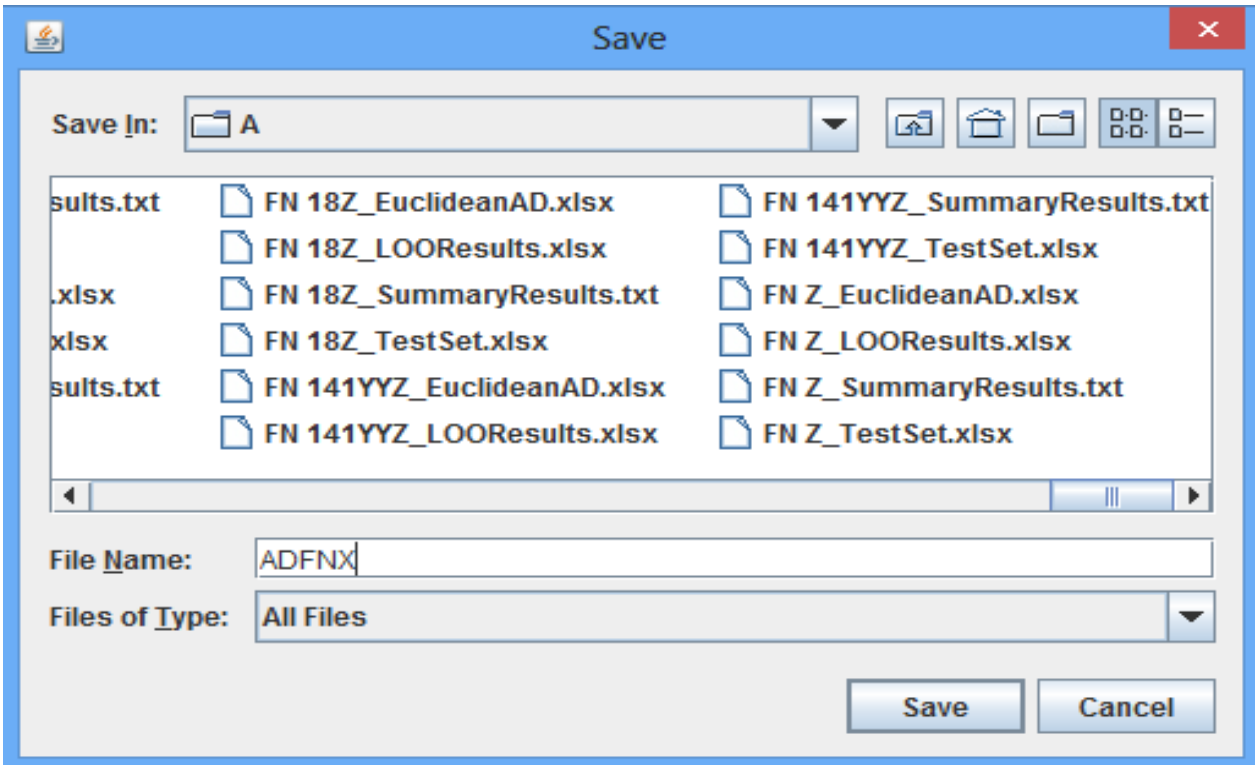


Figure 3.52 Save file dialogue

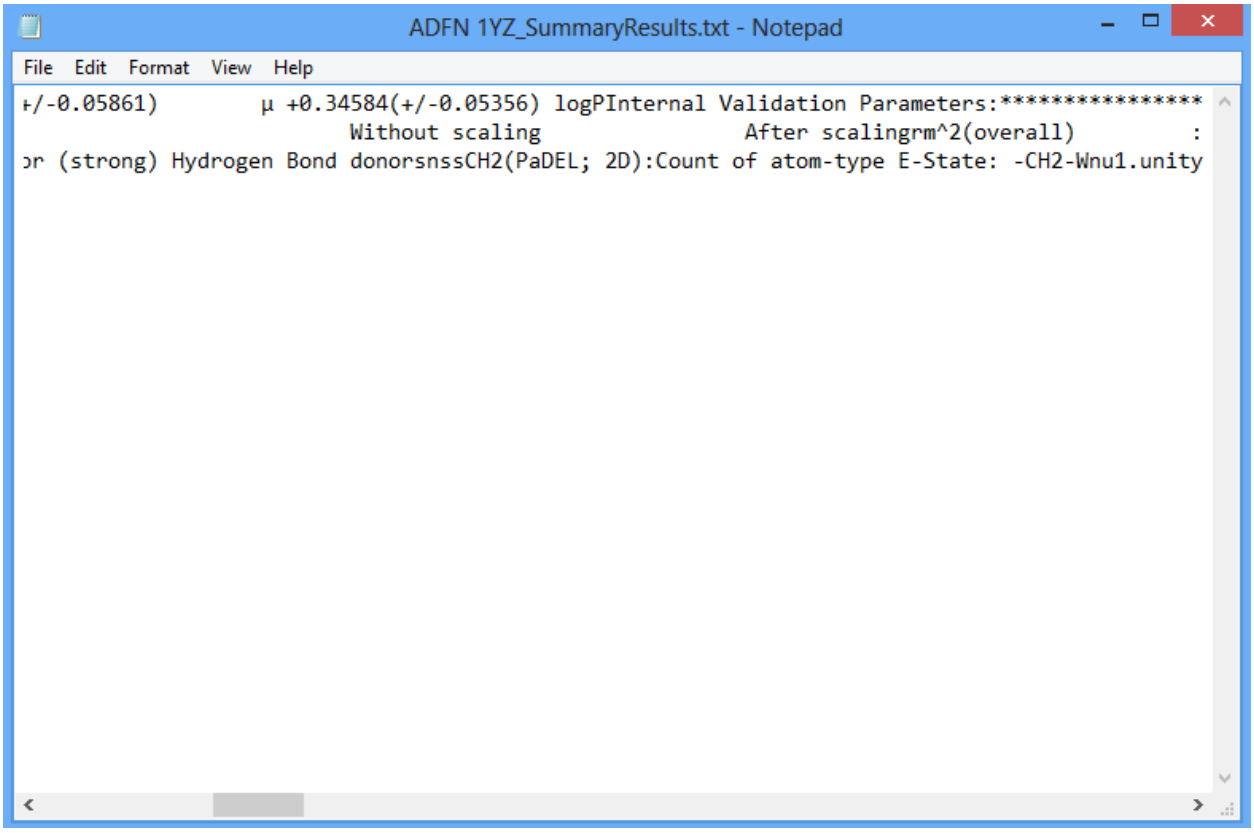


Figure 3.53 Result summary pad

The activity/cytotoxicity are randomly shuffledand a new QSAR model is developed using the original descriptors matrix (Afantitis *et al*., 2006a; Pramod *et al*., 2008; Jalali-Heravi *et al*., 2008 and Afantitis *et al*., 2006b). It is expected that the new QSAR models (after several repetitions) obtained for the training set with randomised activities should have significantly lower values of pIC_{50} for the training set than themodels built using training set with real activities.If the opposite happens then an acceptable QSAR model cannot be obtained for the specific modelling method and data. Hence the original models should be discarded.Y-randomisation was done with the aid of Material Studio Software. In the validation step of multiple linear regression as described in section 3.2.12.1 above, the Y-randomisation box of Figure 3.46 was checked and the number of trials was set at 10 (as shown in the figure) i.e. the activities/cytotoxicities were 10 times randomly shuffled. On clicking the Ok key of Figure 3.46, the calculation was done by the software and the result was displayed in Figure 3.47.

3.2.14 Applicability domain

Applicability domain isdefined as a "physico-chemical, structural, or biological space, knowledge or information on which the training set of the model has been developed. Applicability domainplays a crucial role for estimating the uncertainty in the prediction of a particularcompound based on how similar it is to the compounds employed to constructthe QSAR model.

The organization for economic co-operation and development(OECD) principles for validation of the QSAR models for regulatory purposes, clearly stated that a model should be used within the boundaries of its applicability domain (AD) (OECD, 2004). Hence, only the predictions forchemicals that falls 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 de 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 standardised residuals (William plot) (Gramatica, 2007). This method offers a graphical assessment of the leverage values (h_{ii}) as a function of the standardised 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. 12)

$$h_{ii} = x_i^T (X^T X)^{-1} x_i \quad \text{..... (12)}$$

where $x_i$is the descriptor row matrix of the compound under consideration and X 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. (13)

$$h^* = \frac{3(P+1)}{n} \quad \text{..... (13)}$$

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

3.2.15 Generation of the Williams plot

The Williams plot, being the plot of the standardised residuals against the leverage values (hi), was generated by first of all determining the hat matrix, H, of the descriptors. Then the leverages-the diagonal of a hat matrix- was extracted for every compound and then plotted against the standardised residuals. The hat matrix of a given set of molecular descriptors was determined through implementation of a simple set of matrix algebra rules (Anscombe and Tukey, 1963; Hoaglin and Welsch, 1978). Xtr was used as an n xm matrix containing the structural information m_j(molecular descriptors) for each training set compound $n_i$separately. The calculation of the hat matrix H of the original Xtr matrix was achieved through several consecutive steps. First of all, the transposed version (X^T_{tr} = m x n) of the original descriptor matrix X_{tr} was generated.

In generating this transposed matrix, the excel software was launched by double clicking the icon on the desktop of the computer, which lead to the display of the excel worksheet (Figure 3.54).The original descriptor matrix was copied and pasted on the excel worksheet, with unity typed down the descriptors in the first column (Figure 3.55). This m x n matric was transposed by copying the matrix and repasted beside the original matrix by clicking a box in the area where the transpose was to be pasted and then the cursor key of the laptop was right clicked to display

the pull down dialogue (Figure 3.56). From the special option of the appeared dialogue box, icon  denoting a transpose was selected and the original matrix was transposed (Figure 3.57).

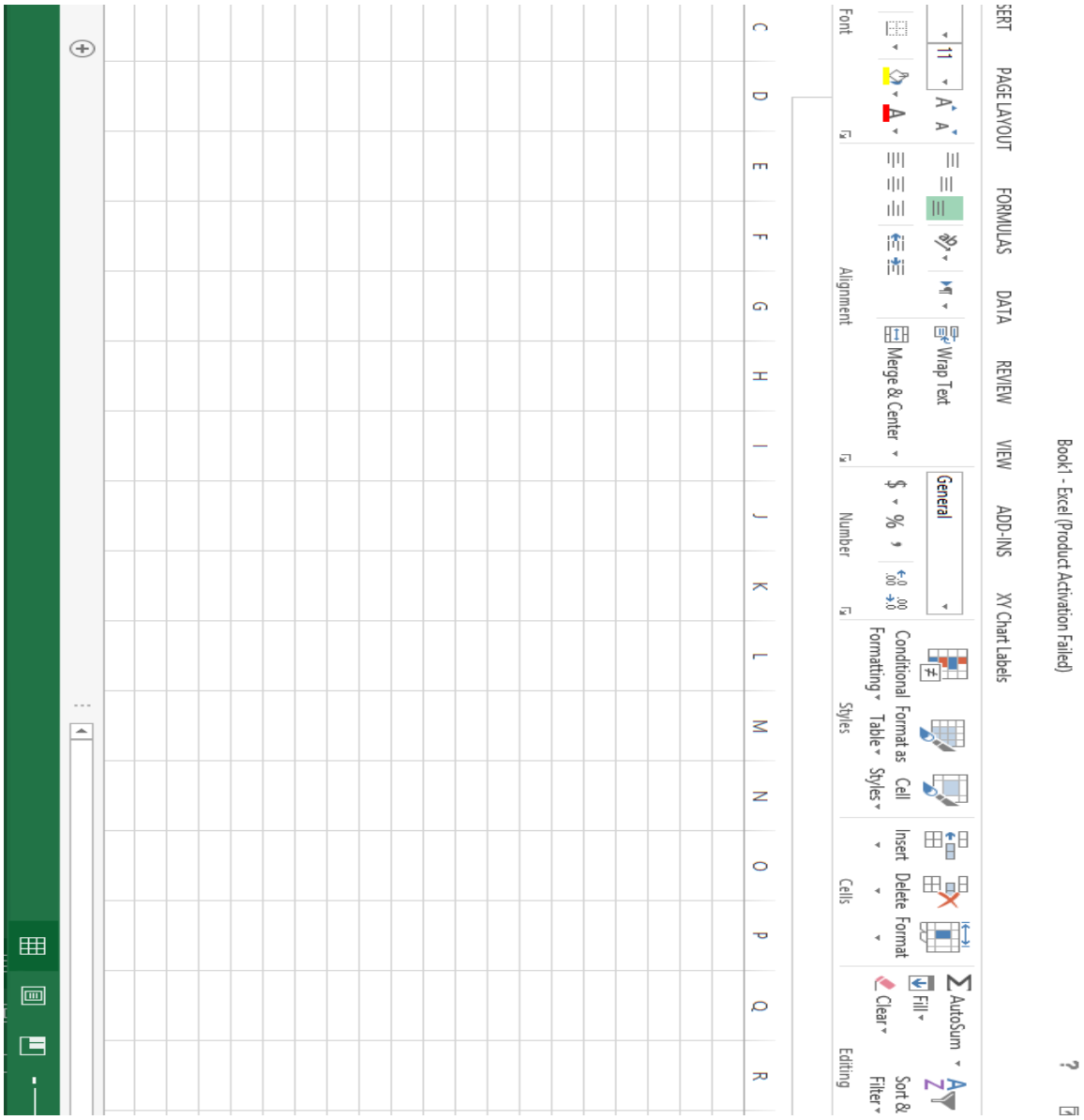


Figure 3.54 Excel worksheet-A

1.1706	0.3374	-1.2165	0.3624	-0.1551	-0.6823														
1.6262	0.3340	-0.2447	-0.8283	0.5779	0.3800														
1.6698	0.3342	-0.2445	0.3624	0.4917	0.0751														
1.8478	0.3345	-1.2318	0.3624	-1.1038	1.0243														
1.1270	-2.1599	1.6465	1.5531	0.7504	1.1892														
1.0780	0.3340	-0.2450	2.7438	-0.3168	0.8844														
1.3195	0.3341	-0.2388	-0.8283	0.8366	1.6615														
1.3685	0.3343	-0.2386	0.3624	0.9552	1.3567														
1.0777	0.3341	-0.2390	0.3624	0.9121	1.8980														
1.1027	0.3343	-0.2388	1.5531	0.9767	1.5931														
1.4444	0.3361	-0.2361	-0.8283	0.0820	-0.9343														
1.5592	0.3346	-0.2370	-0.8283	-0.1875	-0.6869														
1.3205	-2.4193	1.6577	-0.8283	-0.0258	-0.2124														
1.5474	0.3345	-0.2371	-0.8283	-0.0689	-0.5151														
1.3086	-2.4193	1.6575	-0.8283	0.1144	-0.0405														
1.1201	0.3346	-0.2400	-0.8283	-0.6079	0.1201														
1.8813	0.3348	-1.3437	-0.8283	-0.3815	0.5946														
1.3569	0.3346	-0.2371	-0.8283	0.3515	-0.3804														
1.1181	0.3956	1.6575	-0.8283	0.0389	0.0942														

Figure 3.55 A descriptor matrix

Figure 3.56 A paste option pull down menu

Figure 3.57 Transpose matrix

The transposed version (X^{T}_{tr} = m x n) of the original descriptor matrix (X_{tr})generated still on the worksheet, was then multiplied by the original matrix X_{tr} 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, *=MM*was typed and return the function 

which when double clicked, results in *=MMULT (array1, array2)* (Figure 3.58). Array 1 is the range of cells covered by the original matrix on the worksheet e.g. A4:G37, while array2 is the range of cells covered by the transpose matrix on the worksheet e.g. J4:AQ10. Array 1 was selected by highlighting the section covered by the original matrix, a *comma* was inserted. Followed by highlighting the section that covered the transpose matrix for array 2, then the bracket was closed. The *CTRL+ SHIFT + ENTER* was pressed to give the result of the multiplication (Figure 3.59). 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. In the first cell of the highlighted section,**=** *MI* was typed which return a pull down dialogue box (Figure 3.60) in which MINVERSE was double clicked to give *MINVERSE (* as shown in Figure 3.61. Then the result of the multiplication matrix was selected and the bracket was closed. The*CTRL+ SHIFT + ENTER* was pressed to give the result of the operation as inverse of the original matrix $(X^{T}_{tr}\ X_{tr})^{-1}$ as shown in Figure 3.62. Finally multiplication of the original Xtr, the inverse $(X^{T}_{tr}\ X_{tr})^{-1}$ and thetransposed matrix X^{T}_{tr}were performed in a similar manner as describe above to obtain a n × n symmetric matrix (hat matrix)whose diagonal elements (leverages) (Figure 3.63) describe the distance of each training

set compound to the structural centroid of the model; the leverages were extracted from the hat matrix.

Figure 3.58 Multiplication worksheet

Figure 3.59 Multiplication of transpose and the original matrix

Figure 3.60 Inverse pull down dialogue

Figure 3.61 Inverse worksheet

Figure 3.62 Inverse of a descriptor matrix

Figure 3.63 Extracted diagonal element of 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, $(X^{T}_{tr}\ X_{tr})^{-1}$ was also generated, the original $X_{te/ext}$, the inverse $(X^{T}_{tr}\ X_{tr})^{-1}$ and the transposed matrix $X^{T}_{te/ext}$ were multiplied. The combine standardised residuals (the difference between the experimental and calculated activity) for the training and test set were plotted against the combine leverages to obtain the Williams plot. The Williams plot was done using the excel sheet, where the standardised residuals as well as the leverages in a fresh excel sheet. The combine data was highlighted followed by clicking INSERT from the menu bar of the excel sheet and Figure 3.64 was displayed. The icon  under the Chats was selected to display a scatter pull-down menu (Figure 3.65) from which the icon  was selected to plot the residual against the leverages (Figure 3.66). The axis were formatted by clicking first of all the Y- axis followed by right clicking the cursor to display a pull down menu (Figure 3.67) from which *Format Axis…* was selected from which a Format Axis dialogue appeared on the right hand side of the screen (Figure 3.68). In the minimum and maximum boxes of the Format Axis dialogue, - 4.0 and 4.0 were typed, respectively to format the Y- axis, a little down the dialogue, the horizontal axis was set to cross the vertical axis at -4.0. The X- axis was also formatted likewise by setting the minimum and maximum values as 0.0 and 1.0, respectively. The hatch lines were then used to mark the ±3, a value for standardised residual that is commonly used as a cut-off value for accepting predictions, as well as inserting the warning leverage (h*). The XY Chart Label on the menu of excel (Figure 3.68) was used to distinguish

the test set from the training set and also properly naming the axis (Figure 3.69). The resultant plot is called the Williams plot.

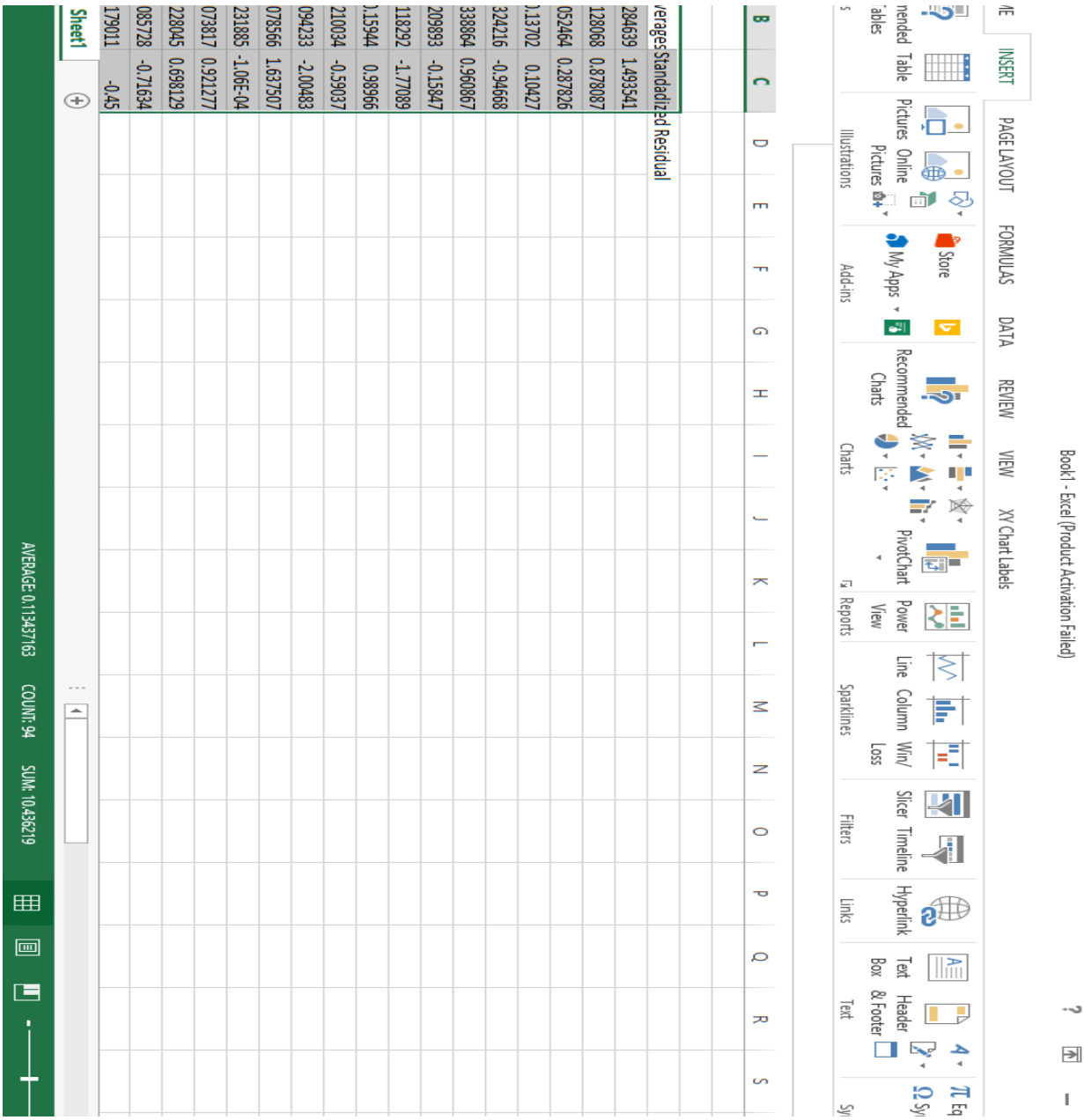


Figure 3.64 Highlighted residual and leverage values

Figure 3.65 A Scatter pull down

Figure 3.66 A plot of residual against the leverage values

Figure 3.67 A pull down menu from axis

Figure 3.68 Format axis dialogue

Figure 3.69 Generated Williams plot

CHAPTER FOUR

4.0 RESULTS

This research was focused on developing valid models that are capable of predicting the*in-vitro* antimalarial activity (IC_{50}) of some naphthylisoquinoline alkaloids (NIQs)and bisbenzylisoquinoline alkaloids (BBIQs)against K1 strain *P. falciparum*, as well as their cytoxicity. The generated models could be used further to design new potential antimalarial drug candidates with higher potency and less toxic. It was on this basis that QSAR approach was utilized. One of the most important steps in QSAR modelling is to define the number of independent variables in the model equation obtained. This will help to eliminate possible cases of over- parameterization of the mathematical model as well as the chance correlation between the molecular descriptors (Topliss and Edwards, 1979). Since the modelling procedure started with a pool of several theoretical molecular descriptors, there is a possibility of encountering chance correlation in cases where the number of examined variables are higher than the number of observations.

Combination of the most relevant descriptors that will produce models with the highest predictive power for the activities of interest were selected from the pool of theoretically calculated descriptors with the aid of genetic algorithm (GA). Eventually the MLRA was employed to build different QSAR models. The usefulness of QSAR models is not just in their ability to reproduce known data, verified by their fitting power (R^2), but is mainly in 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, with some of these models havinghigh squared correlation coefficient R^2, low standard error s and high Fisher coefficient F.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 byRavichandran*et al*., 2011.

4.1 Results of Optimisation and Calculation of Quantum-chemical Descriptors

Geometry optimisation of the structures took an average time of 9 hours to complete depending on the structure. The optimisation process ran smoothly except in few cases where the optimisation process ran out of cycles and the number of cycles had to be increased with the aid of '*Options*' section on the Calculationsdialogue box (Figure 3.9). 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, dipole moment, polarisability, ovality, enthalpy, entropy and Gibbs free energy. The descriptors values of these individual molecules were obtained once the optimisation has ended by selecting *Properties*from the *Display*menu of Spartan 14 software to bring up the *Molecule Properties*dialogue.Quantum descriptors such asionisation energy (IP), electron affinity(EA), global hardness (η),chemical potential (μ),global electrophilicity index (ω) and total softness were calculated using their various expressions.

4.2 Results of the Calculation of 0D, 1D, 2D and 3D Descriptors using Padel

From the Padel software, the pool of filtered calculated 0-3D molecular descriptors which amount to a total of 1500, 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.

4.3 Results of Variable Elimination and Selection

The variable elimination, reduces the number of descriptors to 317 all of which 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.

4.4 Results of Antimalarial Activity (pIC_{50})Modelling of Naphthylisoquinoline Derivatives against K1 Strain of *Plasmodium falciparum.*

The four best models generated by the GA-MLRA analysis performed using the BuildQSAR program are presented below. Several models were first generated from which the four best were selected based on theirhigh squared correlation coefficient R^2, low standard error s and high Fisher

coefficient *F*. Statistical parameters of the internal validation parameters for each of the models are presented below the respective models;

MODEL 4.4.1

pIC_{50} = + 5.6287 (± 0.1514) + 0.3071 (± 0.1625) minHBd + 0.2692 (± 0.1569) µ + 0.3176 (± 0.1428) logP

(R^2= 0.6080, R^2_{adj} = 0.5688,F= 15.5097, s= 0.4318, Q^2_{LOO} =0.5187)

MODEL 4.4.2

pIC_{50} = + 5.6172 (± 0.1401) + 0.1779 (± 0.1447) SssCH2 + 0.2899 (± 0.1507) minHBd + 0.2697 (± 0.1448) µ + 0.2882 (± 0.1339) logP

(R^2 = 0.6781, R^2_{adj} = 0.6337,F = 15.2742, s = 0.3980, Q^2_{LOO} = 0.5728)

MODEL 4.4.3

pIC_{50} = + 5.6315 (± 0.1500) + 0.1129 (± 0.1793) VCH-7 + 0.3277 (± 0.1643) minHBd + 0.2458 (± 0.1598) µ + 0.2806 (± 0.1531) logP

(R^2 = 0.6292, R^2_{adj} = 0.5781,F = 12.3019, s = 0.4270, Q^2_{LOO} = 0.5020)

MODEL 4.4.4

pIC_{50} = + 5.6329 (± 0.1297) + 0.1364 (± 0.1552) BCUTc-1l + 0.2077 (± 0.1369) SssCH2 + 0.2273 (± 0.1320) minHsOH - 0.2077 (± 0.1368) Wlambda2.unity + 0.1674 (± 0.1418) µ + 0.2559 (± 0.1321) logP

(R^2 = 0.7483, R^2_{adj} = 0.6924,F = 13.3812, s = 0.3647, Q^2_{LOO} = 0.6042)

Although selection of a robust and well predictive QSAR model cannot be based on the basis of R^2, Q^2 and R^2_{pred} only since it capable of misleading, further statistical analysis were carried out on the selected models as suggested by Ravichandran*et al*., 2011. The result is presented in Table 4.1. Table 4.2 compares the predictive powers of the four models.

Model 4.4.4 was chosen as the best model since it has the highestR^2_m (Overall) value out of the four models.Figure 4.1 shows the corresponding plot of the experimental activities and the calculated activity

for both the training and external validation set.The Y-randomisation analyses results is shown in Table 4.3. The applicability domain which was obtained from the scatter plot defined by standardised residuals as a function of the leverage (Williams plot) is shown in Figure 4.2. Table 4.4 shows the correlation matrix of the descriptors in model 4.4.4.

Table 4.1Ravichandran*et al*., parameters for the activity models of Naphthylisoquinoline Alkaloids.

Model	$Q^2>0.5$	$R^2_{pred}> 0.6$	$\|R_o^2 - R'^2_o\|< 0.3$	$[(R^2 – R_o^2)/R^2]< 0.1$	$0.85<k>1.15$	$R^2_m\geq 0.5$
1	0.5187	0.7469	0.0841	0	1.0302	0.5156
2	0.5728	0.8815	0.2262	0.0798	1.0469	0.5515
3	0.5020	0.8057	0.0226	0.0030	1.0296	0.5013
4*	0.6042	0.7271	0.0262	0.0147	1.0355	0.5550

*= Best of the selected model

Table 4.2 Comparison of observed activity to predicted activity of Naphthylisoquinoline Alkaloids Models.

Comp. No.	Observed activity	Predicted activity			
		Model 1	Model 2	Model 3	*Model 4
1^	6.15	5.86	5.71	5.84	5.74
2	5.91	5.80	5.88	5.64	5.82
3	5.62	5.87	5.70	5.87	5.44
4	5.77	5.69	5.74	5.57	5.91
5	5.46	5.81	5.78	5.79	5.86
6	6.86	6.80	7.03	6.86	7.00
7	6.92	6.42	6.85	6.37	6.74
8	5.77	6.37	6.15	6.46	5.91
9	6.65	5.86	5.84	6.06	5.99
10	6.58	5.89	6.10	5.91	6.39
11	4.72	5.07	5.00	4.99	5.13
12	5.73	5.23	5.12	5.07	5.13
13^	6.09	5.87	5.71	5.76	5.92
14	5.60	5.40	5.28	5.33	5.31
15	6.36	6.03	5.85	6.01	6.08
16	4.97	5.85	5.68	5.81	5.52
17	5.00	5.66	5.50	5.62	5.36
18^	5.19	5.54	5.41	5.47	5.34
19	6.27	6.02	5.85	6.02	6.04
20	5.52	5.82	5.63	5.65	5.96
21	6.34	6.28	6.04	6.18	6.20
22	5.10	5.51	5.37	5.55	5.42
23^	6.07	5.79	5.62	5.81	5.90
24	5.16	5.51	5.37	5.55	5.45
25^	6.12	6.08	5.89	6.16	6.12
26	5.93	5.51	5.37	5.55	5.42
27^	6.32	6.09	5.90	6.18	6.18
28^	4.76	4.93	4.85	4.79	4.59
29^	5.92	5.64	5.55	5.54	5.35
30	4.83	5.36	5.22	5.28	4.79
31	4.64	5.10	5.00	4.90	4.87
32	4.81	4.40	4.52	4.43	4.47
33	5.68	5.45	5.58	5.49	5.43
34	5.96	5.80	6.12	5.89	6.23
35	4.80	5.16	5.20	5.22	4.95
36	5.87	5.65	5.76	5.71	5.57
37	5.63	5.91	6.21	6.03	6.07
38^	6.02	5.66	5.67	5.86	5.36
39^	5.32	4.84	5.14	4.90	5.46
40	5.76	5.70	5.71	5.75	5.71
41	4.96	5.01	5.15	5.11	5.16
42	4.58	4.52	4.67	4.53	4.98
43	5.97	5.40	5.54	5.50	5.59
44	4.95	4.85	4.88	5.00	4.83

NB. The numbers with ^ represent the test set.

*= Best of the selected model

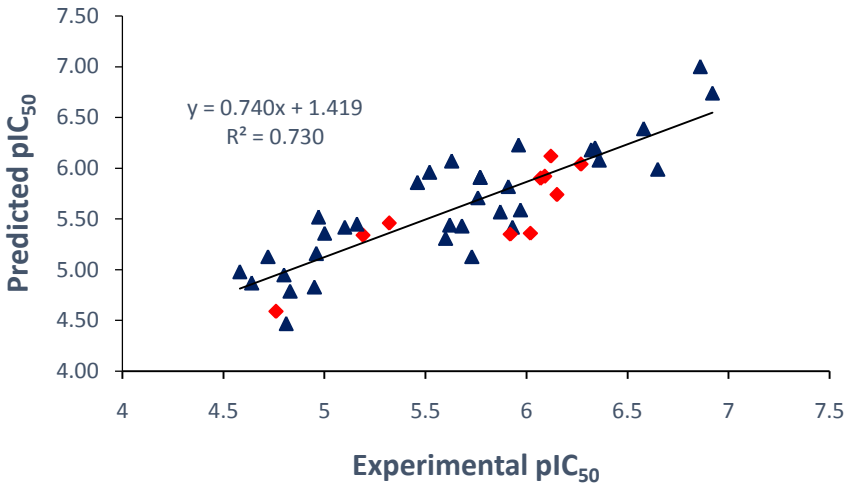


Figure 4.1 Plots of the observed and predicted pIC_{50}values for the training and test sets for activities of naphthylisoquinoline alkaloids.

NB: ▲ represents training sets while ◆ represents test sets.

Table 4.3 Y-randomisation results of the QSAR analysis of the activity of naphthylisoquinoline derivatives.

SN	Correlation	R^2	Q^2
1	0.3921	0.2382	0
2	0.0648	0.2540	0
3	0.2279	0.2334	0
4	0.2818	0.2188	0
5	0.1983	0.2728	0
6	0.0542	0.0829	0
7	0.1713	0.1209	0
8	0.3309	0.1487	0
9	0.0357	0.0878	0
10	0.2645	0.2115	0

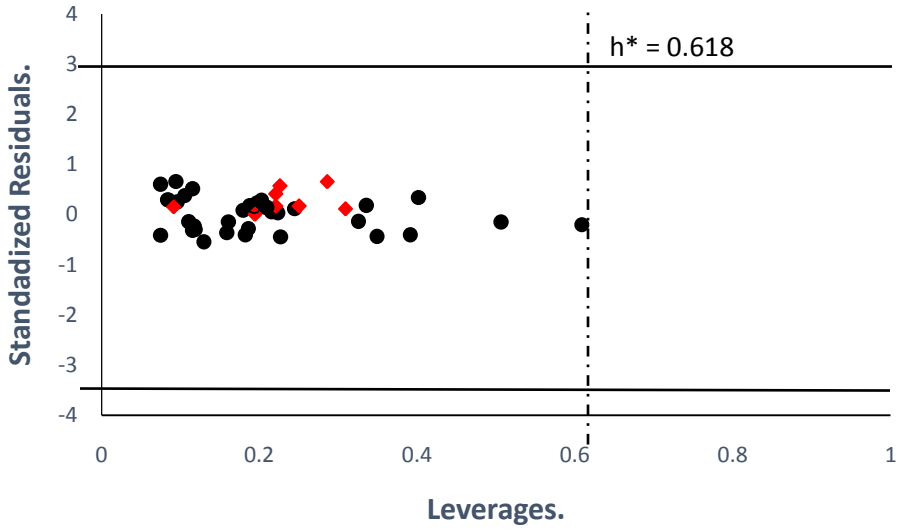


Figure 4.2 Williams Plot for an externally validated for activities of naphthylisoquinoline alkaloids. Cut-off value: h*= 0.618

NB: ● represents training sets while represents test sets.

Table 4.4 Correlation matrix of the descriptors selected for the activity of Naphthylisoquinoline derivatives model generation.

	pIC_{50}	BCUTc-1l	VCH-7	SssCH2	minHBd	minHsOH	Wlambda2.unity	μ	logP
pIC_{50}	1								
BCUTc-1l	0.3934	1							
VCH-7	0.3432	0.0121	1						
SssCH2	0.2635	-0.2308	0.3423	1					
minHBd	0.4431	0.1244	-0.1696	-0.0192	1				
minHsOH	0.4225	0.0579	-0.0205	-0.0464	0.8778	1			
Wlambda2.unity	-0.3462	-0.0818	0.1501	-0.0219	-0.3711	-0.0815	1		
μ	0.2816	0.2611	0.1800	-0.1164	-0.2599	-0.2032	-0.0446	1	
logP	0.5509	0.3043	0.4285	0.1851	0.0676	0.1176	0.0197	0.0180	1

4.4.1 Evaluation of model 4.4.4

The mathematical equation of the model 4.4.4 is as follows:

MODEL 4.4.4

pIC_{50} = + 5.6329 (± 0.1297) + 0.1364 (± 0.1552) BCUTc-1l + 0.2077 (± 0.1369) SssCH2 + 0.2273 (± 0.1320) minHsOH - 0.2077 (± 0.1368) Wlambda2.unity + 0.1674 (± 0.1418) μ + 0.2559 (± 0.1321) logP

(R^2 = 0.7483, R^2_{adj} = 0.6924,F = 13.3812, s = 0.3647, Q^2_{LOO} = 0.6042)

The high correlation coefficient R (0.865) indicates the susceptibility of descriptors (BCUTc-1l,SssCH2, minHsOH, Wlambda2.unity, μ and logP)to form model 4.4.4. Squared correlation coefficient (R^2) of 0.748 explains 75% variance in biological activity of the tested compounds. It also indicates the statistical significance >99.9% with F-values (13.3812). Cross-validated square correlation coefficient (Q^2) by LOO technique was 0.604 which showed a good internal predictive ability of the model. The model was also validated by applying the Y-randomisation test. The values of the activity were shuffled 10 times and each randomised set was used to construct a model which arein good agreement with the suggested limits (Eriksson *et al*., 2003).The low values of R^2 and $Q^2_{LOO}$of the random models shown in Table 4.3and the value of $R^2_{Pred} \geq 0.6$indicates that there is no chance correlation or structural dependency in the proposed model. Consequently, model 4.4.4can be considered as a very good model with both high statistical significant and excellent predictive ability. The applicability domain (AD) of model 4.4.4 (Williams plot) shown in Figure 4.2 was generated using the leverage approach. Training set objects used in the model development are presented as black circles, whereas the external validation set presented as green diamond. The analysis of the AD for the training set objects shows that no compound appeared after the thresh-hold value. Hence the compounds have no outlier ($h > h^* = 0.618$), i.e. 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 ±3.0σ.

4.5 Results of Antimalarial Activity (pIC_{50}) Modelling ofBisbenzylisoquinoline Derivatives against K1 Strain of *Plasmodium falciparum*.

Some of the models generated by the GA-MLRA analysis using the BuildQSAR program for bisbenzylisoquinoline derivatives against K1 strain are presented below.From the several models generated, four models were selected as the best based on theirhigh squared correlation coefficient R^2, low standard error *s* and high Fisher coefficient *F*, and are presented below along with the internal parameters:

MODEL 4.5.1

pIC_{50} = + 5.7633 (± 0.1154) + 0.1769 (± 0.1240) BCUTw-1h - 0.3147 (± 0.1172) CrippenLogP + 0.3746 (± 0.1240) maxssO

(R^2= 0.8845, R^2_{adj} = 0.8556,F = 30.6195, s = 0.2014, Q^2_{LOO} = 0.7942)

MODEL 4.5.2

pIC_{50} = + 5.7436 (± 0.1298) + 0.1734 (± 0.1767) BCUTp-1h - 0.3041 (± 0.1402) CrippenLogP + 0.3123 (± 0.1787) maxssO

(R^2= 0.8490, R^2_{adj} = 0.8112,F = 22.4835, s = 0.2303, Q^2_{LOO} = 0.7100)

MODEL 4.5.3

pIC_{50} = + 5.7448 (± 0.1356) - 0.3819 (± 0.1385) CrippenLogP + 0.4333 (± 0.1412) maxssO + 0.1266 (± 0.1521) MLFER_L

(R^2= 0.8363, R^2_{adj} = 0.7954,F = 20.4376, s = 0.2397, Q^2_{LOO} = 0.6853)

MODEL 4.5.4

pIC_{50} = + 5.7264 (± 0.1373) - 0.3304 (± 0.1426) CrippenLogP + 0.4134 (± 0.1454) maxssO - 0.0958 (± 0.1295) geomRadius

(R^2= 0.8286, R^2_{adj} = 0.7857,F = 19.3329, s = 0.2453, Q^2_{LOO} = 0.5838)

The four models have high values of R^2, Q^2 and R^2_{pred}, though high values of these parameters does not necessarily point to the robustness of the models. As models with low Q^2may turn out to have high R^2_{pred}. Therefore, selection of the best model becomes difficult. This made it necessary to determine more parameters such as R^2_m as suggested by Ravichandran*et al*., 2011, which are presented in Table 4.5. Table 4.6 compares the predictive power of the four models.

Model 4.5.1 was chosen as the best model based on its highest R^2_mvalue out of the four models. The plot of the experimental versus the calculated activities for both the training and external validation sets is shown in Figure 4.3. The Y-randomisation test is shown in Table 4.7 and the Williams plot, which shows the applicability domain of the model, is shown in Figure 4.4.Correlation matrix of the descriptors in model 4.5.1 is shown in Table4.8.

4.5.1 Evaluation of model 4.5.1

Model 4.5.1 is represented by the following equation:

pIC_{50} = + 5.7633 (± 0.1154) + 0.1769 (± 0.1240) BCUTw-1h - 0.3147 (± 0.1172) CrippenLogP + 0.3746 (± 0.1240) maxssO

(R^2= 0.8845, R^2_{adj} = 0.8556,F = 30.6195, s = 0.2014, Q^2_{LOO} = 0.7942)

In this model, the descriptors responsible for the activity of the compounds are BCUTw-1h, CrippenLogP and maxssO. Only descriptor CrippenLogP has negative correlation coefficient, all others have positive correlation coefficients. Decreasing the value of the CrippenLogP descriptor and increasing the values of the other two descriptors will increase the activities of the compounds.Squared correlation coefficient (R^2) of 0.885 explains 88.5% variance in biological activity of the tested compounds. The real

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. Thus, the leave-one-out (LOO) cross-validation method was carried out to ensure the robustness of the model. The value of the leave one out cross-validation, Q^2_{LOO} value, 0.7942 suggests a good internal validation and indicates the robustness and stability of the formulated model. The Y-randomisation test carried out indicates that the model do not occur by chance. The analysis of the Williams plot for the training and validation set shows that all the compounds were within the applicability region. Hence the compounds contain no outlier.

Table 4.5Ravichandran*et al*.,parameters for the activities models of Bisbenzylisoquinoline Alkaloids.

Model	$Q^2>0.5$	$R^2_{pred}> 0.6$	$\|R_o^2 - R'^2_o\|< 0.3$	$[(R^2 – R_o^2)/R^2]< 0.1$	$0.85<k>1.15$	$R^2_m \geq 0.5$
1*	0.7942	0.9774	0.0892	0.0746	1.0446	0.7679
2	0.7100	0.8926	0.2456	0.0915	1.0575	0.7043
3	0.6853	0.8775	0.1023	0.0315	1.0553	0.6862
4	0.5838	0.9065	0.1460	0.0634	0.9065	0.5889

*= Best of the selected model

Table 4.6 Comparison of observed activity to predicted activity of Bisbenzylisoquinoline Alkaloids Models.

Comp.No.	Observed activity	Predicted activity			
		*Model 1	Model 2	Model 3	Model 4
1^	6.20	6.00	5.82	6.08	5.85
2	5.20	5.38	5.47	5.64	5.42
3	6.02	5.98	5.91	5.88	5.81
4	5.87	5.60	5.66	5.72	5.73
5	5.46	5.67	5.39	5.52	5.44
6	5.97	5.99	6.06	6.16	6.16
7^	7.05	6.53	6.38	6.44	6.25
8	5.84	6.07	5.85	5.97	5.66
9^	6.24	6.07	5.58	5.97	5.88
11	6.08	6.21	6.16	6.15	6.19
12^	5.84	5.68	5.50	5.53	5.43
13	6.35	6.21	6.16	6.15	6.25
14	6.13	5.82	5.81	5.71	5.79
15	5.41	5.35	5.20	5.34	5.43
16	4.66	4.84	4.71	4.83	4.73
17	4.38	4.41	4.57	4.42	4.68
18	5.94	6.19	6.34	6.14	6.29
20	5.69	5.59	5.74	5.78	5.91
21	5.74	5.61	5.65	5.63	5.62
22	5.68	5.52	5.75	5.41	5.31

NB. The numbers with ^ represent the test set.

*= Best of the selected model

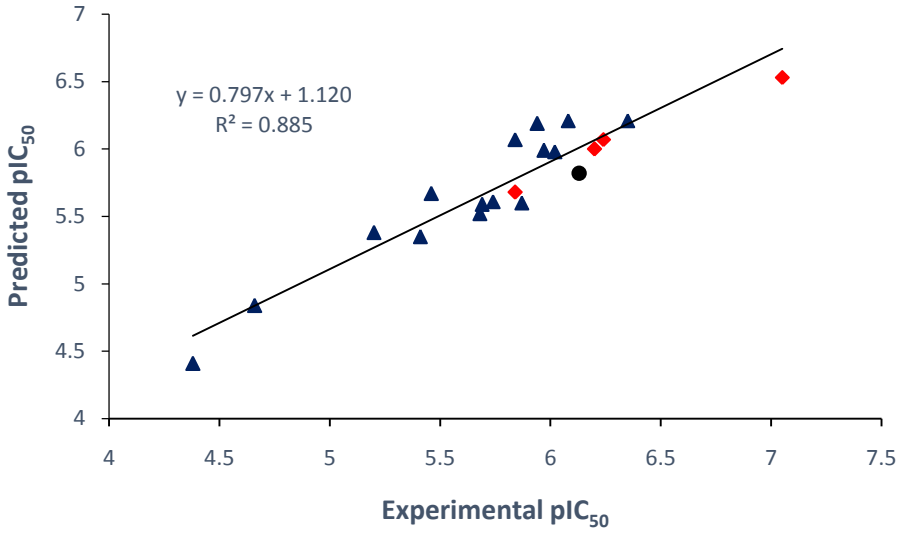


Figure 4.3 Plots of the observed and predicted pIC_{50}values for the training and test sets for activities of bisbenzylisoquinoline alkaloids.

NB: ▲ represents training sets while ◆ represents test sets.

Table 4.7Y-randomisation results of QSAR analysis of the activity of Bisbenzylisoquinoline derivatives.

SN	Correlation	R^2	Q^2
1	0.3672	0.1867	0
2	0.0026	0.0528	0
3	0.1405	0.1848	0
4	0.2010	0.0180	0
5	0.4773	0.3290	0.0928
6	0.0633	0.2227	0.1334
7	0.0015	0.4282	0.1533
8	0.1711	0.1680	0
9	0.1014	0.2121	0
10	0.1586	0.1395	0

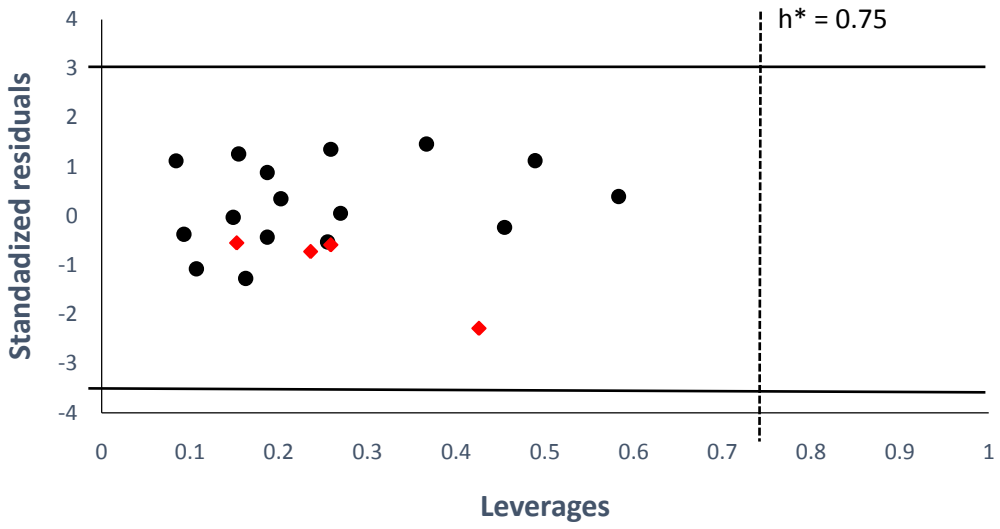


Figure 4.4 Williams plot for an externally validated for activities of bisbenzylisoquinoline alkaloids. Cut-off value h*, = 0.75.

NB: ● represents training sets while represents test sets.

Table 4.8Correlation matrix of the descriptors selected for the activity of Bisbenzylisoquinoline derivatives model generation.

	pIC_{50}	BCUTw-1h	BCUTp-1h	CrippenLogP	maxssO	MLFER_L	geomRadius
pIC_{50}	1						
BCUTw-1h	0.7086	1					
BCUTp-1h	0.7874	0.7890	1				
CrippenLogP	-0.3339	-0.1157	-0.0240	1			
maxssO	0.6123	0.3288	0.6503	0.3729	1		
MLFER_L	0.3283	0.6428	0.7098	0.2105	0.1965	1	
geomRadius	-0.3359	-0.0890	-0.0847	0.2671	-0.0190	-0.0611	1

4.6 Results of Cytotoxicity (pIC_{50}) Modellingof Naphthylisoquinoline Derivatives against L6 Cell Line.

From the GA-MLRA models generated in the BuildQSAR program, four models were selected as the best models based on theirhigh squared correlation coefficient R^2, low standard error *s* and high Fisher coefficient *F*. Statistical parameters of the internal validation parameters for each of the models are presented for each of the four models:

MODEL 4.6.1

pIC50 = + 3.9934 (± 0.0761) + 0.2810 (± 0.0785) MDEC-33

(R^2= 0.6256, R^2_{adj} = 0.6139,F = 53.4729, s = 0.2120, Q^2_{LOO} = 0.7865)

MODEL 4.6.2

pIC50 = + 3.9969 (± 0.0764) + 0.0414 (± 0.0829) SaasC + 0.2644 (± 0.0852) MDEC-33

(R^2= 0.6378, R^2_{adj} = 0.6144,F = 27.2910, s = 0.2119, Q^2_{LOO} = 0.5641)

MODEL 4.6.3

pIC50 = + 4.0397 (± 0.0940) - 0.1084 (± 0.0879) ETA_Eta_B + 0.1413 (± 0.0932) XLogP - 0.1147 (± 0.1134) Weta3.unity

(R^2= 0.6861, R^2_{adj} = 0.6428,F = 15.8442, s = 0.2039, Q^2_{LOO} = 0.5738)

MODEL 4.6.4

pIC50 = + 4.0053 (± 0.0722) - 0.1047 (± 0.0663) ETA_Eta_B + 0.0765 (± 0.0755) XLogP - 0.0859 (± 0.0917) Weta3.unity - 0.0540 (± 0.0732) Weta1.polar - 0.1814 (± 0.0755) η

(R^2= 0.7098, R^2_{adj} = 0.6579,F = 13.6937, s = 0.1996, Q^2_{LOO} = 0.5935)

Looking at the parameters (R^2, R^2_{adj} and Q^2_{LOO}) for all the four models, it could be said that all the models are good and predictable models. But further analysis yield variation in the four models as seen in Table 4.9. Comparison of observed and calculated toxicity models are shown in Table 4.10. The correlation matrix of the descriptors used in the models are presented in Table 4.12.

Model 4.6.1 may not have the highest values of $R^2$and R^2_{pred}, but since it has the highest R^2_m (Overall) value, itwas selected as the best of the four models. Figure 4.5 is the plot of the observed and predicted pIC_{50} values for the training and validation sets of the best model. The Y-randomisation analyses results for model 4.6.1 are also presented in Table 4.11. The Williams Plot shown in Figure 4.6 gives the applicability domain ofmodel 4.6.1. Correlation matrix of the descriptors that occurred in model 4.6.1 are given in Table 4.12.

4.6.1 Evaluation of model 4.6.1

Model 4.6.1 is represented by the following equation:

MODEL 4.6.1

pIC_{50} = + 3.9934 (± 0.0761) + 0.2810 (± 0.0785) MDEC-33

(R^2= 0.6256, R^2_{adj} = 0.6139,F = 53.4729, s = 0.2120, Q^2_{LOO} = 0.7865)

Model 4.6.1 is a one-parameter equation, which has MDEC-33 as the only descriptor that influence the cytotoxicity of naphthylisoquinoline alkaloids. The descriptor has a positive correlation coefficient which indicates that decreasing the value of the descriptor will decrease the cytotoxicity value. Squared correlation coefficient (R^2) of 0.6256 explains about 62.6% variance(adjusted coefficient of variation) ofthe L6 cell lines of the tested compounds. The leave-one-out (LOO) cross-validation result (Q^2_{LOO} = 0.7865) shows how robust and stable the model was. The Y-randomisation test, as shown in Table 4.11, reveals that the model did not occur by chance. The Williams plot of Figure 4.6 suggest a compound lying just on the cut-off value, below which all other compound falls within the applicability region. Hence the compounds contain no outlier.

4.7 Results of Cytotoxicity (pIC_{50})Modelling of Bisbenzylisoquinoline Derivatives against KB Cell Line.

Based on theirhigh squared correlation coefficient R^2, low standard error *s* and high Fisher coefficient *F*, four of the models generated by GA-MLRA using Build QSAR program selected as the best models were presented below for bisbenzylisoquinoline derivatives against KB cell line along with their statistical as well as internal validation parameters:

Table 4.9Ravichandran*et al*.,parameters for the cytotoxicity models of Naphthylisoquinoline Alkaloids.

Model	Q^2>0.5	R^2_{pred}> 0.6	$\|R_o^2 - R'^2_o\|$< 0.3	$[(R^2 - R_o^2)/R^2]$< 0.1	0.85<k>1.15	$R^2_m \geq 0.5$
1*	0.5729	0.7865	0.0217	0.0056	0.9738	0.6339
2	0.5641	0.7775	0.0325	0.0036	0.9712	0.6200
3	0.5738	0.6796	0.0422	0.0203	0.9718	0.5995
4	0.5935	0.7044	0.0697	0.1042	1.0247	0.6236

*= Best of the selected model

Table 4.10 Comparison of observed toxicity to predicted toxicity of Naphthylisoquinoline Alkaloids Models.

Comp. No.	Observedcytotoxicity	Predictedcytotoxicity			
		*Model 1	Model 2	Model 3	Model 4
2	4.00	3.60	3.58	3.81	3.90
3	4.36	4.15	4.18	4.21	4.30
4	4.46	4.15	4.18	4.26	4.38
5	4.03	4.15	4.11	4.04	4.27
6	4.00	4.34	4.36	4.08	4.17
7	4.57	4.34	4.36	4.22	4.43
8	4.00	4.34	4.30	3.95	4.27
9	4.01	4.34	4.30	4.02	4.38
10	4.43	4.34	4.31	3.92	4.08
11	4.40	4.34	4.32	3.95	4.18
12	4.00	3.89	3.89	3.95	3.96
13	4.23	4.05	4.06	4.30	4.17
14	4.28	4.05	4.02	4.24	4.11
15	4.00	4.09	4.14	4.29	4.17
16	4.00	4.09	4.11	4.21	4.09
17	4.00	4.09	4.10	4.10	4.13
18	4.00	4.09	4.06	4.02	4.06
19	4.25	4.09	4.15	4.21	4.07
20	4.00	4.09	4.11	4.15	4.15
21^	4.00	4.09	4.03	3.97	3.96
22	4.01	4.09	3.99	4.00	3.95
23	4.42	4.25	4.25	4.40	4.45
24	4.29	4.13	4.16	4.11	4.09
25	4.46	4.36	4.35	4.37	4.38
26	4.24	4.36	4.32	4.33	4.22
27	4.14	4.25	4.25	4.40	4.33
28^	4.14	4.25	4.22	4.31	4.27
29	3.41	3.72	3.69	3.65	3.63
30^	3.73	3.93	3.95	3.75	3.79
31^	3.89	4.15	4.19	4.07	4.13
32^	3.39	3.57	3.55	3.64	3.66
33	3.39	3.57	3.56	3.60	3.28
34^	3.52	3.76	3.80	3.45	3.48
35	3.66	3.77	3.85	4.14	3.85
36^	4.05	3.96	4.00	3.90	3.88
37	4.02	3.76	3.80	3.77	3.77
38^	3.82	3.77	3.85	4.08	4.01
39	4.64	4.39	4.40	4.09	4.21
40^	3.54	3.59	3.62	3.76	3.70
41^	3.57	3.58	3.56	3.79	3.73
42	3.68	3.59	3.51	3.52	3.70
43	3.33	3.48	3.46	3.82	3.47
44	3.58	3.60	3.63	3.87	3.59
46	3.54	3.94	3.97	3.84	3.68

Numbers with ^ represent the test sets.

*= Best of the selected model

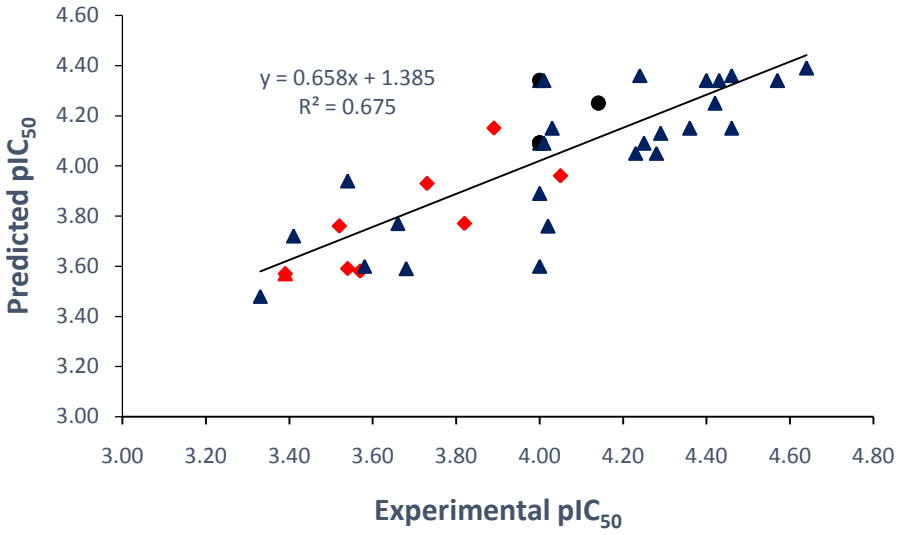


Figure 4.5 Plots of the observed and predicted pIC_{50} values for the training and test sets for cytotoxicity of naphthylisoquinoline alkaloids.

NB: ▲ represents training sets while ◆ represents test sets.

Table 4.11 Y-randomisation results of the QSAR analysis of the cytotoxicity of Naphthylisoquinoline derivatives.

SN	Correlation	R^2	Q^2
1	0.0466	3.0571E-5	0
2	0.1873	0.0164	0
3	0.1137	0.0476	0
4	0.1118	0.0129	0
5	0.1248	0.0073	0
6	0.1766	0.1534	0.0499
7	0.1854	0.0023	0
8	0.0814	0.0173	0
9	0.0779	0.1122	0.0221
10	0.0983	0.0034	0

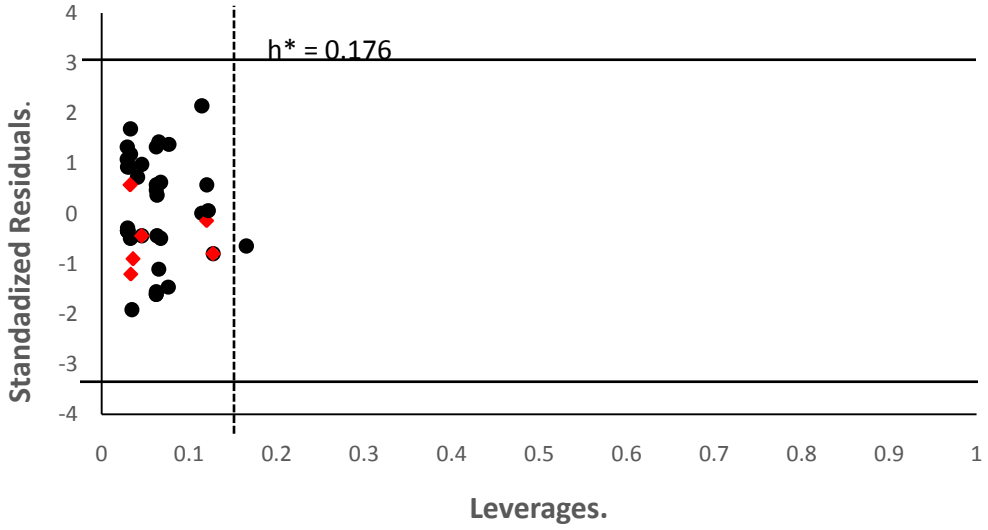


Figure 4.6 Williams Plot for an externally validation for cytotoxicity of naphthylisoquinoline alkaloids. Cut-off value h*= 0.176

NB: ● represents training sets while represents test sets.

Table 4.12Correlation matrix of the descriptors selected for the cytotoxicity of Bisbenzylisoquinoline derivatives model generation.

	pIC_{50}	SaasC	ETA_Eta_B	MDEC-33	XLogP	Weta3.unity	Wlambda3.eneg	Weta1.polar	η
pIC_{50}	1								
SaasC	0.3440	1							
ETA_Eta_B	-0.3541	-0.2778	1						
MDEC-33	0.8222	0.3253	-0.2474	1					
XLogP	0.5143	0.3574	-0.0786	0.54517	1				
Weta3.unity	-0.3094	-0.2499	-0.1427	-0.30852	-0.1989	1			
Wlambda3.eneg	0.3269	0.3004	0.1461	0.38576	0.0938	-0.6377	1		
Weta1.polar	-0.3132	-0.1501	-0.0425	-0.37774	-0.2037	0.3834	-0.2869	1	
η	-0.6636	-0.0743	0.1044	-0.72066	-0.3409	-0.0287	-0.1305	0.0172	1

MODEL 4.7.1

pIC_{50} = + 4.1639 (± 0.1281) + 0.4041 (± 0.1629) ATSc4 + 0.3876 (± 0.1374) VCH-6 + 0.1682 (± 0.1630) SHBint9

(R^2= 0.8742, R^2_{adj} = 0.8323,F = 20.8550, s = 0.1886, Q^2_{LOO} = 0.7115)

MODEL 4.7.2

pIC_{50} = + 4.1341 (± 0.1417) + 0.3638 (± 0.1869) ATSc4 + 0.4048 (± 0.1671) VCH-6 + 0.0828 (± 0.1378) SC-3

(R^2= 0.8325, R^2_{adj} = 0.7766,F = 14.9056, s = 0.2177, Q^2_{LOO} = 0.7112)

MODEL 4.7.3

pIC_{50} = + 4.1539 (± 0.1431) + 0.3369 (± 0.1885) ATSc4 + 0.1089 (± 0.1539) BCUTp-1l + 0.4172 (± 0.1658) VCH-6

(R^2= 0.8428, R^2_{adj} = 0.7904,F = 16.0838, s = 0.2109, Q^2_{LOO} = 0.6944)

MODEL 4.7.4

pIC_{50} = + 4.1311 (± 0.1342) + 0.3501 (± 0.1802) ATSc4 + 0.3908 (± 0.1521) VCH-6 - 0.0950 (± 0.1247) SP-6

(R^2= 0.8482, R^2_{adj} = 0.7976,F = 16.7588, s = 0.2073, Q^2_{LOO} = 0.7221)

In all the models, the value of R^2 shows how fit the points are on the regression line. The robustness and predictive power of the models were evident in both the Q^2 and R^2_{pred} values as shown in Table 4.13. The models were analysed further for R^2_m value which strike a balance between low Q^2 and high R^2_{pred} value. The predicted toxicity values by the models were compared to the experimental values as shown in Table 4.14. The correlation matrix of the descriptors used in the models are likewise presented in Table 4.15.

Based on the value R^2_m, model 4.7.1 was selected as the best model. The relationship between the predicted and experimental value of this model is shown in Figure 4.7. Table 4.16 and Figure 4.8show the Y-randomisation analysis and the Williams plot obtained for the model, respectively.

4.7.1 Evaluation of model 4.7.1

Model 4.7.1 is represented by the following equation:

MODEL 4.7.1

pIC_{50} = + 4.1639 (± 0.1281) + 0.4041 (± 0.1629) ATSc4 + 0.3876 (± 0.1374) VCH-6 + 0.1682 (± 0.1630) SHBint9

(R^2= 0.8742, R^2_{adj} = 0.8323,F = 20.8550, s = 0.1886, Q^2_{LOO} = 0.7115)

Model 4.7.1 is a three parameter model which contain ATSc4, VCH-6 and SHBint9 as the influential descriptors. The positive regression coefficient imply that decreasing the values of the descriptors will decrease the value of cytotoxicity.

Squared correlation coefficient (R^2) of 0.874 explains about 87.4% variance in biological activity of the tested compounds. The leave-one-out (LOO) cross-validation result (Q^2_{LOO} = 0.7115) shows how robust and stable the model was. The Y-randomisation test, as shown in Table 4.15, indicates that the model did not occur by chance. The analysis of the applicability domain for the training and validation sets using the Williams plots shows that only one compound of the validation set was identified as a typical *X*-outlier ($h > h^*$ =0.92). This is seen in Figure 4.8.

Table 4.13Ravichandran*et al*.,parameters for thecytotoxicity models of Bisbenzylisoquinoline Alkaloids.

Model	Q^2>0.5	R^2_{pred}> 0.6	$\|R_o^2 - R'^2_o\|$< 0.3	$[(R^2 - R_o^2)/R^2]$< 0.1	0.85<k>1.15	$R^2_m \geq 0.5$
1*	0.7115	0.9304	0.0517	0.0833	1.0127	0.6552
2	0.7112	0.9387	0.0430	0.0254	1.0504	0.6465
3	0.6944	0.8730	0.0313	0.0017	1.0405	0.6233
4	0.7221	0.7295	0.1509	0.0038	1.0586	0.5802

*= Best of the selected model

Table 4.14 Comparison of observed to predicted cytotoxicity of Bisbenzylisoquinoline Alkaloids Models.

Comp. No.	Observed toxicity	Predicted toxicity			
		*Model 1	Model 2	Model 3	Model 4
1^	3.97	3.92	3.81	3.85	3.68
2	3.39	3.42	3.41	3.31	3.31
3	4.33	4.21	4.03	4.14	4.29
4	3.98	3.75	3.92	3.86	3.97
5	4.09	4.03	4.11	4.00	4.15
6	4.13	4.24	4.32	4.41	4.35
7	4.39	4.17	4.33	4.40	4.21
8	4.36	4.52	4.51	4.41	4.53
11^	4.50	4.61	4.32	4.41	4.35
12	3.98	4.24	4.11	4.00	4.15
13	4.75	4.61	4.32	4.40	4.35
14^	4.26	4.32	3.92	3.85	3.97
15^	4.26	3.99	3.97	4.08	3.75
16^	3.38	3.21	3.39	3.39	3.49
17	3.35	3.48	3.37	3.48	3.55
20	4.73	4.58	4.68	4.64	4.60
21	4.42	4.61	4.61	4.63	4.53
22	3.53	3.58	3.70	3.77	3.45

NB. The numbers with ^ represent the test set.

*= Best of the selected model

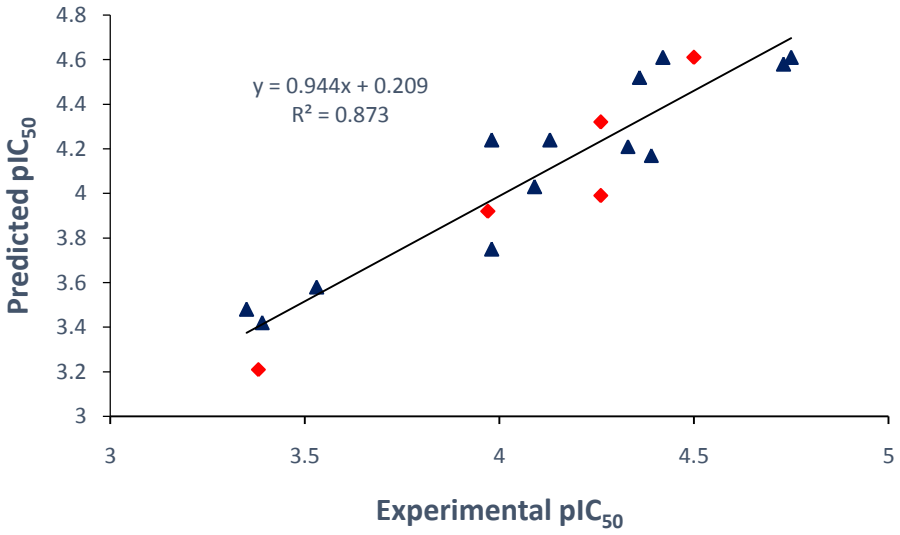


Figure 4.7 Plots of the observed and predicted pIC_{50} values for the training and test sets for cytotoxicity of bisbenzylisoquinoline alkaloids.

NB: ▲ represents training sets while ◆ represents test sets.

Table 4.15 Y-randomisation results of QSAR analysis of the cytotoxicity of Bisbenzylisoquinoline derivatives.

SN	Correlation	R^2	Q^2
1	0.3330	0.1162	0
2	0.0510	0.1109	0
3	0.6735	0.4532	0
4	0.2183	0.2596	0
5	0.6712	0.2876	0
6	0.2911	0.2649	0
7	0.0174	0.0031	0
8	0.1838	0.2534	0
9	0.2643	0.0478	0
10	0.2160	0.5574	0

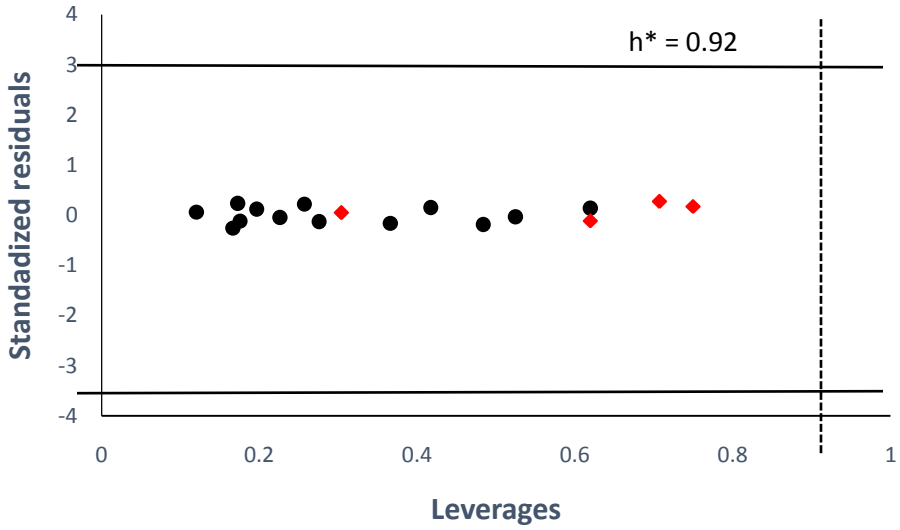


Figure 4.8 Williams Plot for an externally validated for cytotoxicity of bisbenzylisoquinoline alkaloids. Cut-off value: h* = 0.92.

NB: ● represents training sets while ◆ represents test sets.

Table 4.16 Correlation matrix of the descriptors selected for the cytotoxicity of Bisbenzylisoquinoline derivatives model generation.

	pIC50	ATSc4	BCUTp-1l	VCH-6	SC-3	SP-6	SHBint9
pIC50	1						
ATSc4	0.5204	1					
BCUTp-1l	0.1667	0.2079	1				
VCH-6	0.5404	-0.2484	-0.3326	1			
SC-3	0.0306	-0.0084	0.6015	-0.2991	1		
SP-6	-0.1048	-0.1841	0.0752	0.2502	-0.0505	1	
SHBint9	0.2581	-0.0614	0.2499	-0.0877	0.3017	-0.2465	1

CHAPTER FIVE

5.0 DISCUSSION

5.1 Interpretation of the Descriptors in Model 4.4.4

BCUTc-1lis known as number of lowest partial charge weighted BCUTS. The BCUT descriptors encode connectivity information and atomicproperties of the molecule. BCUTc-1l descriptorwas found to correlate with the antimalarial activity. The positive sign of the correlation is an indication thatthe two variables are positively (more *X* means more *Y*) or negatively (more *X* means less *Y*) related. As could be seen from the model equation, both the pIC_{50} and BCUTc-1l have a good positive correlation coefficient (0.3934) as indicated in Table 4.4, meaning that the descriptor BCUTc-1l influences the antimalarial activity of naphthylisoquinoline derivatives against K1 strain of *Plasmodium falciparum*.Various descriptors among which is BCUTc-1l descriptor have been used (Pereira*et al*., 2015) in QSAR-Assisted virtual screening of lead-like molecules from marine and microbial natural sources for antitumor and antibiotic drug discovery.

SssCH2 (Sum of atom-type E-State: -CH2-) and minHsOH (Minimum atom-type H E-State: -OH) descriptors belongs to Electrotopological State Atom (E-State) Index. The E-state index refers to an atom/sub-moleculardescriptor encoding both electronic and topologicalinformation. The E-state parameter takes intoconsideration the effect of changing electronicenvironment on the activity pattern of the molecules. It also help to determine those regions within the molecules thatare most influential in their biological activity variation. The intrinsic values of E-state parameter increases with an increase in electronegativity or electron density of anatom while it decreases with an increase in the extent ofbranching (substitution) of the atoms. For example, theintrinsic value of the primary carbon ($-CH_3$) is greater thanthe secondary carbon ($-CH_2-$) which is followed by thetertiary (-CH<) and quaternary (>C<) carbon atoms. It thusexplains that an increase in the degree of branchingdecreases the intrinsic value of

the atoms. The descriptors representing the E- state indices are positively correlated to the antimalarial, hence the activity of the antimalarial compound will increase with increase in the values of both $SssCH_2$ and minHsOH descriptors.E-state indices have found many applications in the fields of quantitative structureactivity/property/ toxicity studies, some of which include antithyroid agents(Abou-Shaaban *et al*., 1996), antiadrenergic activity(Roy *et al*., 1999), antimalarial activity(Roy *et al*., 2000), antineoplastic activity(Roy *et al*., 2001), blood-brain partitioning data(Rose and Hall, 2002), anti-Inflammatory activity(De*et al.*, 2002).

Wlambda2.unity which stands for directional WHIM,(weight holisticinvariant molecular) descriptors based on statisticalindices and are built in such a way as to capture relevantmolecular 3D information regarding molecular size, shape,symmetry, and atom distribution with respect to invariantreference frames. The negative sign of the coefficient of determination points to the fact that decreasing the value of the descriptor, increases the value of the activity of the antimalarial compounds.Singh*et al*.(2014) developed QSTR modelling for qualitative and quantitative toxicity predictions of diverse chemical pesticides in honey bee for regulatory purposes using probabilistic neural network (PNN) and generalised regression neural network (GRNN) QSTR models and reported R^2 values of 0.841. In that study, Wlambda2.unity and BCUTw-1h descriptors played major roles.

μis known as dipole moment. The dipole moment is often considered as the direct characteristics of the polarity of a molecule. It indicates the strength and orientation behavior of a molecule in an electrostatic field. In the regression equation, the dipole moment is positively correlated with the biological activity, thus indicating that increasing the polarity of the molecule by substituting the groups with more polar groups would account for an increase in the biological activity. This descriptor has been reported elsewhere to be important in determining the activity of antimalarial compounds (Paliwal *et al.,* 2012).

LogP, which referred to the logarithm of the ratio of the concentration of solute between two solvents, is a measure of lipophilicity or hydrophobicity of compounds. The positive sign of the descriptor in the regression model indicates that increasing the hydrophobicity of the substituents will account for an increase in the biological activity of the compound. Log P was reported as the determinate of biological activity of which antimalarial compounds (Paliwal *et al.,* 2012).

5.2 Interpretation of the Descriptors in Model 4.5.1

BCUTw-1h known as *number of highest atom weighted BCUTS* is a BCUT (Burden-CAS-University of Texas eigenvalues) descriptor. BCUT descriptorswere designed to encode atomic properties relevant to intermolecular interactions. BCUTw-1h descriptor shows a high value of correlation (*r*= 0.7086) between the descriptor and the activity of the antimalarial compounds. Since the sign of the coefficient of determination is positive, the value of the activity increases with increase in the value of the descriptor.Singh*et al*.(2014) in their QSTR Modelling for Qualitative and Quantitative Toxicity Predictions of Diverse Chemical Pesticides in Honey Bee for Regulatory Purposesfound the BCUTw-1h descriptor to be one of the structural features responsible for the toxicity of chemical pesticides.

CrippenLogP is a Crippen's descriptor and is related tohydrophobicity. The negative value of the coefficient of determination implies that reducing the value of the CrippenlogP increases the activity of the compounds. Sun*et al*.(2015) reported the role played by CrippenlogP, in which bothlocal and global models were built for the prediction of chemicalaquatic toxicityof differentfish species.After feature reduction andvariable selection by the F-score and Pearson correlation coefficientmethods, four descriptors including CrippenLogP feature in their model.

maxssO stands for *maximum atom-type E-state -O-*, which is an electrotopological state atom type descriptor. The regression coefficient has a positive sign indicating that increase in the value of the

descriptor, will increase the activities of the compounds.Although the use of the descriptor in QSAR analysis was not clearly reported, electrotopological state atom type descriptor (a group in which maxssO, together with SssCH2 and minHsOH belongs) was discussed in section 5.1.

5.3 Interpretation of the Descriptors in Model 4.6.1

MDEC-33 is the only descriptor in the model which stand for *molecular distance edge between all tertiary and quaternary carbons*. It is an MDE(Molecular Distance Edge)Descriptor. The good positive correlation (*r*=0.8222) of MDEC-33 with cytotoxicity implies that the descriptor MDEC-33strongly influences the cytotoxicity value of the antimalarial compounds. The sign of the coefficient of determination shows a direct relationship, in which reducing the value of the descriptor, reduces the cytotoxicity value of the compounds.MDE descriptor has been reported elsewhere to have been used in QSAR analysis in anti-infective drugconcentrations prediction (Valitalo*et al*., 2016), and anti-cancer activity (Kashyap*et al*., 2016).

5.4 Interpretation of the Descriptors in Model 4.7.1

ATSc4**,** known as ATS autocorrelation descriptor, weighted by charges is an autocorrelationdescriptorcharge which was found to be correlated to the cytotoxicity. The significant and positive correlation (*r* = 0.5204) indicates that the descriptor ATSc4 strongly influences the cytotoxicity value of bisbenzylisoquinoline derivatives as an antimalarial agent. The positive sign of the coefficient of determination of the descriptor implies that reduction in the value of the descriptor reduces the cytotoxicity of the compounds. In a study of antimalarial activity of Prodiginines (Masand*et al*., 2013) against *Plasmodium falciparum*, the best model was shown to contain autocorrelation

descriptors (ATS7v, ATS5v) which determined activity of the Prodiginines. The positive sign in front of their coefficient of determination indicates that increase in the values of the descriptors increase the values of the activity.

VCH-6,which stands forvalence chain, order 6, is a topological descriptors. This descriptor describes the overall topology of the molecules whose correlation coefficient value is equally great, indicating how the descriptor can affect the cytotoxicity of the compounds. The coefficient of determination of the descriptor is positive which indicates a direct relation between the descriptor and cytotoxicity. Thus, decreasing this descriptor value, reduces cytotoxicity. A similar investigation conducted by Patel*et al.*(2014), on the quantitative structure-activity relationship (2d and 3d-QSAR) analysis of tylophorine derivatives as efficacious in antimalarial drug design, the valence chain, order 6 (chiV6chain) was reported in the best model as the most significant descriptor responsible for the activity of tylophorine derivatives.The chichaindescriptor (VCH.7) was also reported to have some influence on the anticancer activity of 255 gossypol acetic acid derivatives (Kashyap*et al.,* 2016).

SHBint9 is an E-State descriptors of strength for potential hydrogen bonds of path length 9,which is an electrotopologicalstateatomtypedescriptor. Although the descriptor is correlated with the cytotoxicity of the compounds, but not as significant as the other descriptors. This is evidencedfrom the value of it correlation coefficient (r = 0.2581), but the positive sign of the correlation coefficient ensures that decreasing the value of the descriptor, reduces the value of cytotoxicity of the compounds.Verma(2014)investigated the *in vitro*vero cell cytotoxicity of some antitubercular agents containing chiral pentaamines, bis–cyclic guanidines, bis cyclic thioureas and bis-cyclic piperazines moieties using 2D descriptors and found that the hydrogen bonding parameters(maxHBint2 maxHBint8 SHBint9 minHbd SHBd,SHBint2, SHBint8, SHdNH) were responsible for the cytotoxicity of the series to the compounds they studied. The models obtained indicate less hydrogenbonding potentials of the compounds to be responsible forgreater cytotoxicity towards the vero cell. Furthermore, Yang and Morris

(2010) modelled the dose of nitroxantrone in rats and found thatSHBint9, Acnt and SHBint3 played a significant role in determination of the dose excreted in the bile (PD_b) and the biliary clearance (CL_b), respectively, where PD_b and CL_bwere found to be positively correlated with the SHBint9 descriptor.

CHAPTER SIX

6.0 SUMMARY, CONCLUSION AND RECOMMENDATION

6.1 Summary

This researchwas designed to develop quantum structure activity-relationship models that could explain:

i. Whether there is any relationship between the structural feature of naphthylisoquinoline derivatives and their anti-malarial activities and cytotoxicity.
ii. Whether there is any relationship between the structural feature of bisbenzylisoquinoline derivatives and their anti-malarial activities and cytotoxicity.
iii. Whether the generated models could be use to predict the activity/cytotoxicity of all the derivatives of naphthylisoquinoline/bisbenzylisoquinoline alkaloids.

The work has established structural features of naphthylisoquinoline and bisbenzylisoquinolinederivatives that could be useful in future design of new but highly potent and safer antimalarial compounds from the naphthylisoquinoline and bisbenzylisoquinoline series that might be potential drugs for malarial treatment in the future.

6.2 Conclusion

These studies-quantum modelling of some potent, non-toxic anti-malarial compounds-werecentred at developing valid QSAR models that will be able to predict the activity as well as the cytotoxicity of derivatives of naphthylisoquinoline and bisbenzylisoquinoline alkaloids.Severaldescriptors were found to influence the activities and cytotoxicity of naphthylisoquinoline and bisbenzylisoquinoline

alkaloids. For the activities of naphthylisoquinoline alkaloids, the following descriptors were also found to be important BCUTc-1l, SssCH2, minHsOH, Wlambda2.unity, μ and LogP,in modelling the antimalarial activity of these class of compounds. While descriptor MDEC-33 is responsible for the cytotoxicity of the compounds. The activities of bisbenzylisoquinoline alkaloids are functions of BCUTw-1h, CrippenLogP and maxssO descriptors, while the cytotoxicity is affected by ATSc4, VCH-6 and SHBint9 descriptors.

6.3 Recommendation

In the future design of novel antimalarial drugs using naphthylisoquinoline and bisbenzylisoquinoline derivatives, it is recommended that

i) Various substituents attached at various positions should be given consideration.

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

iii) It is also recommended that the BCUTc-1l, (number of high lowest partial charge weighted BCUTS), SssCH2, (Sum of atom-type E-State: -CH2-), minHsOH, (Minimum atom-type H E-State: -OH), Wlambda2.unity (Directional WHIM, weighted by unit weights), μ (dipole moment) and LogP (hydrophobicity). BCUTw-1h (number of high lowest atom weighted BCUTS), CrippenLogP (Crippen's hydrophobicity), and maxssO (Maximum atom-type E-State: -O-) descriptors for activity and MDEC-33 (Molecular distance edge between all tertiary carbons), ATSc4 (ATS autocorrelation descriptor, weighted by charges), VCH-6 (Valence chain, order) and SHBint9 (Sum of E-State descriptors of strength for potential Hydrogen Bonds of path length 9) for cytotoxicity, should be given priority in designing anti-malarial drug with lower cytotoxic effect.

REFERENCES

Abou-Shaaban, R. R. A., Al-Khamees, H. A., Abou-Auda, H.S. and Simonelli, A.P. (1996**).** Atom level electrotopological state indexes in QSAR: Designing and testing antithyroid agents. *Pharmaceutical Research*,**13**: 129-136.

Afantitis, A., Melagraki, G., Sarimveis, H., Igglessi-Makorpoulon, O. and Kollias, G. (2009). A novel QSAR model for predicting the inhibition of CXCR3 receptor by 4-*N*-aryl-[1, 4] diazepane ureas. *European Journal of Medicinal Chemistry*,**44**: 877-884.

Afantitis, A., Melagraki, G., Sarimveis, H., Koutentis, P. A., Markopoulos, J. and Igglessi-Markopoulou, O. (2006). A novel simple QSAR model for the prediction of anti-HIV activity using multiple linear regression analysis. *Molecular Diversity*,**10**(3): 405-414.

Afantitis, A., Melagraki, G., Sarimveis, H., Koutentis, P. A., Markopoulos, J. and Igglessi-Markopoulou, O. (2006).A novel QSAR model for predicting induction of apoptosis by 4-aryl-4H-chromenes. *Bioorganic and Medicinal Chemistry*,**14** (19): 6686-6694.

Ambure, P., Gajewicz, A., Puzyn, T. and Roy, K. (2015). 'NanoBRIDGES' sofwares: open access tools to perform QSAR and nano-QSAR modelling. *Chemometrics and Intelligent Laboratory Systems*,**147**:1-3.

Anscombe, F. J., and Tukey, J. W. (1963). The examination and analysis of residuals. *Technometrics*,**5**: 141-160.

Baumann, K. (2003). Cross-validation as the objective function for variable-selection techniques.*Trends in Analytical Chemistry*,**22** (6): 395-406.

Becke, A.D. (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A*,**38**:3098–3100.

Benigni, R. and Giuliani, A. (2003). Putting the predictive toxicology challenges into perspective: reflections on the result.*Bioinformatics* (Oxford, England), **19**(10): 1194–1200.

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

Bernazzani, L., Duce, C., Micheli, A., Mollica, V., Sperduti, A. Starita, A. and Tine, M. R. (2006). Predicting physical-chemical properties of compounds from molecular structures by recursive neural networks.*Journal of Chemical Information Model*,**46**: 2030−2042.

Bitonti, A. J., Sjoerdsma, A., McCann, P. P., Kyle, D. E., Oduola,A. M., Rossan, R. N., Milhous, W. K. and Davidson, D. E. Jr. (1988). Reversal of chloroquine resistance in malaria parasite Plasmodium falciparum by desipramine. *Science*,**242**(4883): 1301–1303.

Branco, L. C., Carrera, G. V. S. M., Aires-de-Sousa1, J., Martin, I. L., Frade, R. and Afonso, C. A. M. (2015). Physico-Chemical Properties of Task-Specific Ionic Liquids. Retrieved January 09, 2016, from http://cdn.intechopen.com/pdfs/13913.pdf.

Brazil, D. F. (2006). Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. A ções de control de malária: manual para profissionais de Saúde básica. Brasilia, Editora Ministério da Saúde, 52 pages.

Brazil, D. F. (2007). Ministério da Saúde. Secretaria de Vigilância em Saúde. Sifuação epitdemiolõgica da malária no Brazil. 12 pages.

Breman, J. G. (2001). The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden.*American Journal of Tropical Medicine and Hygiene*, **64**: 1-11.

Bringmann, G. and Pokorny, F.(1995). The naphthylisoquinoline alkaloids. In: G.A. Cordell (Ed.). The alkaloids: Chemistry and Pharmacology. Vol. 46. Academic Press, New York, pp.127-271.

Bringmann, G., Dreyer, M., Faber, J. H., Dalsgaard, P. W., Staerk, D., Jaroszewski, J. W., Ndangalasi, H., Mbago, F., Brun, R., Reichert, M., Maksimenka, K. and Christensen, S. B. (2003).Ancistrotanzanine A, the first 5,3'-coupled naphthylisoquinoline alkaloid, and two further, 5,8'-linked related compounds from the newly described species Ancistrocladus tanzaniensis.*Journal of Natural Products*,**66**(9): 1159-1165.

Bringmann, G., François, G., Aké Assi, L. and Peters, Schlauer, J. (1998).The alkaloids of Triphyophyllum peltatum (Dioncophyllaceae). *Chimia*,**52**: 18–28.

Bringmann, G., Hamm, A., Günther, C., Michel, M., Brun, R. and Mudogo, V. (2000).Ancistroealaines A and B, two new bioactive naphthylisoquinolines, and related naphthoic acids from Ancistrocladus ealaensis.*Journal of Natural Products*, **63**(11): 1465–1470.

Bringmann, G., Hinrichs, J., Peters, K. and Peters, E.-M. (2001). Synthesis of a chiral aryl-ferocenyl ligands by intramolecular coupling to a biaryl-related lactone. *Journal of Organic Chemistry*, **66**(2): 629–632.

Bringmann, G., Pfeifer, R.-M., Schreiber, P., Hartner, K., Kocher, N., Brun, R., Peters, K., Peters, E.-M. and Breuning, M. (2004). Synthesis and antitrypanosomal activity of 2-aminomethyl-1-(2-oxyphenyl)naphthalenes. *Tetrahedron*, **60**: 6335–6344.

Bringmann, G., Pfeifer, R.-M., Schreiber, P., Hartner, K., Schraut, M. and Breuning, M. (2004).The 'lactone method': enantioselective preparation of novel P,N-biaryl ligands and their use in the synthesis of the biarylic alkaloids, ancistrotanzanine B and ancistroealaine A.*Tetrahedron*, **60**: 4349–4360.

Canfield, C. J., Pudney, M. and Gutteridge, W. E. (1995).Interactions of atovaquone with other antimalarial drugs against Plasmodium falciparum in vitro.*Experimental Parasitology*, **80**: 373-381.

Clough, B. and Wilson, R. J. M. (2001). Antibiotics and the plasmodial plastid organelle. In: *Antimalarial Chemotherapy: Mechanisms of Action, Resistance, and New Directions in Drug Discovery* (ed. P. J. Rosenthal). Totawa, NJ: Humana Press, pp.265-286.

David, A. W. (2002). The role of QSAR in biomolecular discovery. *Briefings in bioinformatics*,**3**(1): 73-86.

Davis, L. (1991). *Handbook of Genetic Algorithms*. Van Nostrand Reinhold: New York, USA, p. 21.

De, K., Roy, K., Saha, A. and Sengupta, C. (2002). QSAR with electrotopological state atom index. Part V. Anti-inflammatory activity of 7 α-halogenocorticosteroids and their derivatives. *Indian Journal of Chemical Society*,**79**: 513-519.

Deoliveira, D.B. and 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 Modelling*. Academic Press Ltd. London. (*J.* Devillers, Ed.). Academic Press, London, pp. 35-66.

Dwuma-Badu, D., Ayim, J. S. K., Tackie, A. N., El Sohly, M. A., Knapp, J. E., Slatkin, D. J. and Schiff, P. L. (1975). Trigilletimine: a new bisbenzylisoquinoline alkaloid from Triclisia species. *Experientia*,**31**:1251-1252.

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 forApplicabilit**y** Evaluations of Classification- and Regression-Based QSARs.*Environmental Health Perspectives*,**111**(10): 361-372.

Ferreira, J. E. V., Figueiredo, A. F., Barbosa, J. P., Cristino, M. G. G., Macedo, W. J. C., Silva, O. P. P., Malheiros, B. V., Serra, R. T. A. and Ciriaco-Pinheiro, J. (2010). A study of new antimalarial artemisinins through molecular modelling and multivariate analysis. *Journal of the Serbian Chemical Society*, **75**(11): 1533–1548.

Fidock, D. A. (2010). Drug discovery – priming the antimalarial pipeline. *Nature*,**465**: 297-298.

Fidock, D.A., Rosenthal, P. J., Croft, S. L., Brun, R. and Nwaka, S. (2004). Antimalarial drug discovery: efficacy models for compound screening.*Nature Reviews Drug Discovery*,**3**: 509-520.

Filler, S., Causer, L.M., Newman, R. D., Barber, A. M., Roberts, J. M., Mac Arthur, J., Parise, M. E. and Steketee, R. W. (2003). Malaria surveillance –United States, 2001. Centers for Disease Control and Prevention. Atlanta, Georgia. U.S. Department of Health, Education, and Welfare. MMWR,**52**(5):1-14.

Fournet, A., Mufioz, V., Manjon, A. M., Angelo, A., Hocquemiller, R., Cortes, D., Cave, A. and Bruneton, J. (1988). Activité antiparasitaire d'alcaloïdes bisbenzylisoquinolinoleiques. a. Activité in vitro sur des promastigotes de trois souches de Leishmania. *Journal of Ethnopharmacology*,**24**:327-335.

François, G., Bringmann, G., Phillipson, J. D., Aké Assi, L., Dochez, C., Rübenacker, C., Schneider, M., Wéry, M., Warhurst, D. C. and Kirby, G. C. (1994). Activity of extracts and naphthylisoquinoline alkaloids from Triphyophyllum peltatum, Ancistrocladus abbreviates and A. barteri against Plasmodium falciparum in vitro. *Phytochemistry*,**35**: 1461-1464.

François, G., Timperman, G., Eling, W., Ake Assi, L., Holenz, J. and Bringmann, G. (1997). Naphthylisoquinoline alkaloids against malaria: evaluation of the curative potentials of dioncophylline C and dioncopeltine A against Plasmodium berghei in vivo. *Antimicrobial Agents and Chemotherapy*,**41**: 2533–2539.

Gamo, F. J., Sanz, L.M., Vidal, J., de Cozar, C., Alvarez, E., Lavandera, J. L., Vanderwall, D. E., Green, D. V. S., Kumar, V., Hasan, S., Brown, J. R., Peishoff, C. E., Cardon, L. R. and Garcia – Bustos, J. F. (2010). Thousands of chemical starting point for antimalarial lead identification. *Nature*,**465**: 305–310.

Golbraikh, A, Tropsha, A. (2002). Beware of q^2! *Journal of Molecular Graphics and Modelling*,**20**:269-276.

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

Greenwood, B. M., Bojang, K., Whitty, C. J. and Targett, G. A. (2005). Malaria. *Lancet*,**365**: 1487-1498.

Guerin, P.J., Olliaro, P., Nosten, F., Druilhe, P., Lax-Minarayam, R., Binka, F., Kilama, W., Ford, N. and White, N. J. (2002). Malaria: current states of control, diagnosis, treatment, and a proposed agenda for research and development. *Lancet Infectious Diseases*,**2**: 564-573.

Hadanu, R., Idris, S. and Sutapa, I. W. (2015). Qsar analysis of benzothiazole derivatives of antimalarial compounds based on am1 semi-empirical method. *Indonesian Journal of Chemistry*, **15**(1): 86–92.

Hallock, Y. F., Manfredi, K. P., Blunt, K. P., Cardellina II, J. H., Schäffer, M., Gulden, K. P., Bringmann, G., Lee, A.Y., Clardy, J., François, G. and Boyd, M. R. (1994). Korupensamines A-D, novel antimalarial alkaloids from Ancistrocladus korupensis *Journal of Organic Chemistry*,**59**: 6349-6355.

Hammett, L., P. (1935). Some Relations between Reaction Rates and Equilibrium Constants. *Chemical Reviews*,**17**:125-136.

Hammett, L., P. (1970). *Physical Organic Chemistry*, 2nd Edn, McGraw-Hill,New York, pp.164-167.

Hansch, C., Maloney, P., P., Fujita, T. and Muir, M., R. (1962). Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature*,**194**:178-180.

Hoaglin, D.C. and R.E. Welsch, R. E. (1978). The hat matrix in regression and ANOVA. *Journal of the American Statistical Association*,**32**: 17-22.

Hou, X., Chen, X., Zhang, M. and Yan, A. (2016). QSAR study on the antimalarial activity of Plasmodium falciparum dihydroorotate dehydrogenase (PfDHODH) inhibitors. *SAR and QSAR in Environmental Research*,**27**(2):101-24.

Jalali-Heravi, M., Asadollahi-Baboli, M. and Shahbazikhah, P. (2008). QSAR study of heparanase inhibitors activity using artificial neural networks and Levenberg—Marquardt algorithm. *European Journal of Medicinal Chemistry*,**43**: 548-556.

Kaschula, C. H., Egan, T. J., Hunter, R., Basilico, N., Parapini, S., Taramelli, D., Pasini, E. and Monti, D. (2002). Structure-activity relationships in 4-aminoquinoline antiplasmodials. The role of the group at the 7- position. *European Journal of Medicinal Chemistry*,**45**: 3531-3539.

Kashyap, V. K., Pandey, R., Saksena, R. and Khan, F. (2016). An Artificial Neural Network Modelling Approach for Development of QSAR Model for Anticancer Activity of Gossypol Acetic Acid against Anticancer Target BCL2. *European Academic Research*,**4**(1): 102-111.

Krettli, A.U., Andrade-Neto, V.F., Brandao, M.G.L. and Ferrari, W.M.S. (2001). The search for new antimalarial drugs from plants randomly selected: a review. *Memórias do Instituto Oswaldo Cruz*,**96**: 1033-1042.

Kuroda, H., Nakazawa, S., Katagiri, K., Shiratori, O., Kozuka, M., Fujitani, K. and Tomita, M. (1976). Antitumour effect of bisbenzylisoquinoline alkaloids. *ChemicalandPharmaceutical Bulletin*, **24**: 2413-2420.

Lameira, J., Medeiros, I.G., Reis, M., Santos, A. S. and Alves, C.N. (2006). Structure–activity relationship study of flavone compounds with anti-HIV-1 integrase activity: A density functional theory study. *Bioorganic and Medicinal Chemistry*,**14**: 7105–7112.

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

Liu, C. and Xiao, P. (1983). Biological activities and domestic resources of bis (benzylisoquinoline) alkaloids. *Chinese Pharmaceutical Bulletin*, **18**:287-292.

Mambu, L. and Grellier, P. (2008). Antimalarial compounds from traditionally used medicinal plants. In: Colegate SM, Molyneux RJ, eds. Bioactive Natural Products. Detection, Isolation and structural determination, 2nd edn. Boca Raton, USA: CRC press, pp.491-529.

Marini-Bettolo, G. B. (1980). Present aspects of the use of plants in traditional medicine. *Journal of Ethnopharmacology*, **2**:5-7.

Marshall, S. J. (1991). *Ph.D. Thesis*. Brighton Polytechnic, Brighton, United Kingdom.

Marshall, S. J., Russell, P. F., Wright, C. W., Anderson, M. M., Phillipson, J. D., Kirby, G. C., Warchurst, D. C. and Schiff, P. L. Jr. (1994). *In Vitro* antiplasmodial, antiamoebic, and cytotoxic activities of a series of bisbenzylisoquinoline alkaloids. *Antimicrobial Agents and Chemotherapy*,**38**: 96-103.

Martin, S.K., Oduola, A. M. and Milhous, W. K. (1987). Reversal of chloroquine resistance in *Plasmodium falciparum* by verapamil. *Science*, **235**: 899-901.

Masand, H. V., Mahajan, T. D., Patil, N. K., Hadda, B. T., Youssoufi, H. M., Jawarkar, D. R. and Shibi, I. G. (2013). Optimisation of Antimalarial Activity of Synthetic Prodiginines: QSAR, GUSAR, and CoMFA analyses. *Chemical Biology* and *Drug Design*,**81**: 527–536.

Meneguzzi, C. V., Biral dos Santos, C., Pinto, I., Feitoza, R. L., Feitoza, N. H., Falqueto, A. (2009). Use of geoprocessing to define malaria risk areas and evaluation of the vectorial importance of anopheline mosquitoes (Diptera: Culicidae) in Espírito Santo, Brazil. *Memórias do Instituto Oswaldo Cruz*, Rio de Janeiro,**104**(4): 570-575.

Meshnick, S. R. (2001). Artemisinin and its derivatives. In; *Antimalarial Chemotherapy: Mechanisms of Action, Resistance, and New Directions in Drug Discovery* (ed. P. J. Rosenthal). Totowa, NJ: Humana Press, pp.191-201.

Minovski, N., Župerl, S., Drgan, V. and Novič, M. (2013). Assessment of applicability domain for multivariate counter-propagation artificial neural network predictive models by minimum Euclidean distance space analysis: A case study. *Analytica Chimica Acta*,**759**: 28–42.

Molfetta, F. A., Bruni, A. T., Rosseli, F. P. and Silva, A. B. F. (2007). A partial least squares and principal component regression study of quinone compounds with trypanocidal activity. *Journal of Structural Chemistry*,**18**: 49-57.

Mundy, S. B., White, A. C. Jr., Hines, J. S., Marino, B. J. and Young, E. J. (1996). *Southern Medical Journal*, **89**: 616-618.

Najafi, A., Sobhanardakani, S. and Marjani, M. (2013). Exploring QSAR for antimalarial activities and drug distribution within blood of a series of 4-aminoquinoline drugs using genetic-MLR. *Journal of Chemistry*. http://dx.doi.org/10.1155/2013/560415.

Oliver-Bever, B. (1986). *Medicinal Plants in Tropical West Africa*. Cambridge University Press, Cambridge, United Kingdom, p. 375.

Organization 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.(Date of access: 15.02.2015).

Paliwal, S., Sharma, J. and Paliwal, S. (2012). Quantitative structure activity relationship analysis of bisbenzofuran cations as antimalarial agents employing multivariate statistical approach. *Indian Journal of Chemistry*,**51B**: 617-630.

Parr, R. G. and Pearson, R. G. (1983). 'Absolute Hardness: Companion Parameter to Absolute Electronegativity'.*Journal of the American Chemical Society*, **105**(26): 7512-7516.

Parr, R. G., Szentpaly, L. V. and Liu, S. (1999). 'Electrophilicity Index'.*Journal of the American Chemical Society*,**121**(9): 1922-1924.

Partridge, S. J., Russell, P. F., Kirby, G. C., Bray, D. H., Warhurst, D. C., Phillipson, J. D., O'Neill, M. J. and Schiff, P. L. (1988). *In vitro* antimalarial activity of *Triclisia patens* and some of its constituent alkaloids. *Journal of Pharmacy and Pharmacology*, Suppliment **40**: 53-56.

Partridge, S. J., Russell, P. F., Kirby, G. C., Warhurst, D. C., Phillipson, J. D., O'Neill, M. J. and Schiff, P. L. (1989). *In vitro* antimalarial activity of *Triclisia dictyophylla* and some of its constituent alkaloids. *Journal of Pharmacy and Pharmacology*, Suppliment **41**: 92-97.

Patel, S. K., George, L., Khedkar, V. M., Lone, M. Y., Jha, P. C., Jasrai, Y. T., Pandya, H. A. and Highland, H. N. (2014). Metaphorical quantitative structure-activity relationship (2d & 3d-qsar) analysis of tylophorine derivatives as efficacious in antimalarial drug design. *International Journal of Pharmaceutical Sciences and Research*,**5**(10): 4325-4338.

Pavanand, K., Webster, H. K., Yongvanitchit, K. and Dechati-wongse, T. (1981). Antimalarial activity of Tiliacora triandra Diels against *Plasmodium falciparumin vitro*. *Phytotherapy Research*,**3**(5): 215-217.

Pearson, R. G. (1988). 'Absolute Electronegativity and Hardness: Application to Inorganic Chemistry'.*Inorganic Chemistry*,**27**(4): 734-740.

Pereira, F., Latino, D. A. and Gaudêncio, S. P. (2015). QSAR-assisted virtual screening of lead-like molecules from marine and microbial natural sources for antitumor and antibiotic drug discovery. *Molecules*,**20**(3): 4848-4873.

Pramod, C., Nair, M. and Sobhia, E. (2008).Quantitative structure activity relationship studies on thiourea analogues as influenza virus neuraminidase inhibitors.*European Journal of Medicinal Chemistry*,**43**: 293-299.

Puzyn, T., Leszczynski, J. and Cronin, T. D. (2010). Recent Advances in QSAR Studies. Methods and Applications. Springer Dordrecht Heidelberg London New York. **8**. pp. 3-11.

Ravichandran, V., Harish, R., 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

Raynes, K. J., Stocks, P. A., O'Neill, P. M., Park, B. K. and Ward, S. A. (1999). New 4- aminoquinoline Mannich base antimalarias. 1. Effect of an alkyl substituent in the 5'- position of the 4'-hydroxyanilino side chain. *Journal of Medicinal Chemistry*,**42**: 2747-2751.

Ringwald, P., Bickil, J. and Basco, L. (1996). Randomised trial of pyronaridine *versus* chloroquine for acute uncomplicated falciparum malaria in Africa. *Lancet*,**347**: 24–28.

Rogers, R. D. and Seddon, K. R. (eds). (2003). *Ionic Liquids as Green Solvents:* Progress and Prospects, ACS Symposium Series 856, American Chemical Society, Washington D.C, pp. 215-219.

Rose, K. and Hall, L.H. (2002). Modelling blood-brain barrier partitioning using the electrotopological state. *Journal of Chemical Information and Computer Science*,**42**: 651-666.

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

Roy, K., De, A.U. and Sengupta, C. (1999**).** QSAR with electrotopological state atom index: Antiadrenergic activity of N, N-dimethyl-2- bromo-2-phenylethylamines. *Indian Journal of Chemistry*,**38B**: 942-949.

Roy, K., Pal, D. K. and Sengupta, C. (2000). Hansch analysis of antimalarial cyclic peroxy ketals with physicochemical and electrotopological parameters. *Drug Design and Discovery*,**17**: 183-190.

Roy, K., Pal, D.K. and Sengupta, C. (2001). QSAR of antineoplastics V: exploration of receptor interaction sites of antitumor N-(7- Indolyl) benzenesulfonamides targeting G1 phase using electrotopological state atom index. *Drug Design and Discovery*,**17**: 207-218.

Sahu, N. S., Sharma, M., Mourya, V. and D. V. Kohli, D. V. (2012). Qsar study of some substituted 4-quinolinyl and 9-acridinyl hydrazones as antimalarial agents. *Acta Poloniae Pharmaceutica ñ Drug Research*,**69**(6): 1153-1165.

Santos, B., C., R., Vieira, J. B., Lobato, C. C., Hage-Melim, L. I. S., Souto, R. N. P., Lima, C. S., Costa, E. V. M., Brasil, D. S. B., Macêdo, W. J. C. and Carvalho, J. C. T. (2014). A SAR and QSAR Study of New Artemisinin Compounds with Antimalarial Activity. *Molecules*,**19**: 367-399.

Schüürmann, G.,Ebert, R. U.,Chen, J., Wang, B., Kühne, R. (2008). External Validation and Prediction Employing the PredictiveSquared Correlation Coefficient – Test set Activity Mean.*Chemical Information* and *Modelling*,**48**(11): 2140-2145.

Shetty P. (2012). The numbers game. *Nature*,**484**:14–15.

Singh, K. P., Gupta, S., Basant, N. and Mohan, D. (2014). QSTR Modelling for Qualitative and Quantitative Toxicity Predictions of Diverse Chemical Pesticides in Honey Bee for Regulatory Purposes. *Chemical Research in Toxicology*,**27**: 1504−1515.

Stocks, P. A., Raynes, K. J. and Ward, S. A. (2001). *Novel quinoline antimalarial chemotherapy: Mechanisms of Action, Resistance, and New Directions in Drug Discovery* (ed. P. J. Rosenthal). Totowa, NJ: Humana Press,pp.235–253.

Sun, L., Zhang, C., Chen, Y., Li, X., Zhuang, S., Li, W., Liu, G., Leea, P. W. and Tang, Y. (2015). In silico prediction of chemical aquatic toxicity with chemical category approaches and substructural alerts. *Toxicology Research*,**4**: 452-463.

Tackie, A. N., Dwuma-Badu, D., Okarter, T., Knapp, J. E., Slatkin, D. J. and Schiff, P. L. (1974). Constituents of West African medicinal plants. II. The isolation of alkaloids from selected *Triclisia* species. *Lloydia*,**37**:1-5.

Taft, W. R. (1952). Polar and steric substituent constants for aliphatic and o- benzoate groups from rates of esterification and hydrolysis of esters. *Journal of the AmericanChemical Society*,**74**:3120-3128.

Thangapandian, S., John, S. and Lee, K. W. (2012). Classical and 3D QSAR studies on inverse agonists of human histamine H1 receptor.*Molecular Simulation*,**38**(13):1143–1156.

Todd, M., Paul, F.H., Douglas, Y. and Alexander, T. (2012). Does rational selection of training and test sets improve the outcome of QSAR modelling? *Journal of Chemical Information and Modelling*,52(10): 2570-8. Doi:10.1021/ci300338w.

Todeschini, R. and Consonni, V. (2000). Handbook of molecular descriptors. Wiley-VCH, Weinheim, Germany, pp.667-669.

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.

Tropsha A, Golbraikh A. (2007). Predictive QSAR modelling workflow, model applicability domains, and virtual screening. *Current pharmaceutical design*,**13**: 3494–3504.

Tropsha, A. (2010). Best practices for QSAR model development, validation and exploitation. *Molecular Informatics*,**29**(6-7): 476-488.

Tropsha, A., Gramatica, P. and Gombar, V. K. (2003).The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models.*QSAR and Combinatorial Science*,**22**(1): 69–77.

Valerio, L.G. Jr. (2009). In silico toxicology for the pharmaceutical sciences.*Toxicology and Applied Pharmacology*,**241**: 356–370.

Valitalo*, P.,* Griffioen*,* K.*,* Rizk*, M.,* Visser*, S.,* Danhof*,* M.*,* Rao*,* G.*,* van der Graaf*, P.,* van Hasselt, J.*(2016).Structure-Base*d Prediction of Anti-infective Drug Concentrations in the Human Lung Epithelial Lining Fluid. *Pharmaceutical Research*, **33**: 856–867.

Van Dyke, K. June 1991. U.S. patent 5,025,020.

Van Schalkwyk, D. A., Waiden, J. C. and Smith, P. J. (2001). Reversal of chloroquine resistance in *Plasmodium falciparum* using combination of chemosensitizers. *Antimicrobial Agents* and *Chemotherapy*,**45**: 3171-3174.

Verma, S. K. (2014). QSAR Modelling of *in vitro*vero cell cytotoxicity of some antitubercular agents containing chiral pentaamines, bis–cyclic guanidines, bis cyclic thioureas and bis-cyclic piperazines moieties using 2D descriptors. *Indian Journal of Applied Research*,**4**(9): 462-467.

Walker, J. D., Jaworska, J., Comber, M. H. I., Schultz, T. W. and Dearden, J. C. (2003). Guidelines for developing and using quantitative structure-activity relationship.*Enviromental Toxicology and Chemistry*,**22**(8): 1653–1665.

Wells, T. N. C., Alonso, P. L. and Gutteridge, W. E. (2009). New medicines to improve control and contribute to the eradication of malaria. *Nature Reviews Drug Discovery*,**8**: 879–891.

WHO. (2008). *World Malaria Report*. World Health Organization: Geneva, Switzerland. http://apps.who.int/malaria/wmr2008/.(Date of access: 23.04.2015).

WHO.(2013). World Malaria Report 2013.Geneva: World Health Organization. Available:http://www.who.int/malaria/publications/world_malaria_report_2013/report/en/(Date of access: 19.03.2014).

Wiesner, J., Ortmann, R., Jomaa, H. and Schlitzer, M. (2003). New antimalarial drugs. *Angewandte Chemie International Edition in English,**42:*** 5274-5293.

Winter, R. W., Kelly, J. X., Smilkstein, M. J., Dodean, R., Bagby, G. C., Rathbun, R. K., Levin, J. I., Hinrichs, D. and Riscoe, M. K. (2006). Evaluation and lead optimisation of anti-malarial acridones. *Experimental Parasitology*,**114**: 47-56.

Yang, S. Y. (2010). Pharmacophore modelling and applications in drug discovery: challenges and recent advances. *Drug Discovery Today*,**15**(11–12):444–450.

Yang, X. and Morris, E., M. (2010).Pharmacokinetics and biliary excretion of mitoxantrone in rats. *Journal of Pharmaceutical Sciences*,**99**(5): 2502-2509.

Yap, C.W., Xue, Y., Li, H., Li, Z.R., Ung, C.Y., Han, L.Y., Zheng, C.J., Cao, Z.W. and Chen, Y.Z. (2006). Prediction of compounds with specific pharmacodynamic, pharmacokinetic or toxicological property by statistical learning methods.*Mini Reviews in Medicinal Chemistry*,**6**(4): 449–459.