QUANTUM MODELING OF SOME COUMARIN AND NEOLIGNANS DERIVATIVES AS ANTIMICROBIAL COMPOUNDS AGAINST *Candida albicans and Epidermophyton floccosum*

BY

ABDULJELIL AJALA
(M.Sc/SCI/21465/2012-2013)

DEPARTMENT OF CHEMISTRY
AHMADU BELLO UNIVERSITY,
ZARIA, NIGERIA

JULY, 2017

QUANTUM MODELING OF SOME COUMARIN AND NEOLIGNANS DERIVATIVES AS ANTIMICROBIAL COMPOUNDS AGAINST *Candida albicans and Epidermophyton floccosum*

BY

Abduljelil AJALA, B.Sc. (Hons) Chemistry (ABU) 2008
(MSc/SCIEN/21465/2012-2013)

A DISSERTATION SUBMITTED TO THE SCHOOL OF POSTGRADUATE STUDIES, AHMADU BELLO UNIVERSITY, ZARIA NIGERIA

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

DEPARTMENT OF CHEMISTRY
FACULTY OF PHYSICAL SCIENCE
AHMADU BELLO UNIVERSITY,
ZARIA, NIGERIA

JULY, 2017

Declaration

I declare that the work in this Dissertation entitled, "**Quantum Modeling of Some Coumarin and Neolignans Derivatives as Antimicrobial Compounds Against *Candida albicans and Epidermophyton floccosum***" has been carried out by me in the Department of Chemistry.The information derived from the literature has been duly acknowledged in the text anda list of references provided. No part of this project was previously presented for anotherdegree or diploma in any institution.

AbduljelilAJALA	Signature	Date

Certification

ThisDissertation entitled, "**Quantum Modeling of Some Coumarin and Neolignans Derivatives as Antimicrobial Compound Against *Candida albicans and Epidermophyton floccosum***" by Abduljelil AJALA meets the regulations governing the award of the degree of M.Sc. Physical Chemistry of Ahmadu Bello University, and is approved for itscontribution to knowledge and literally presentation.

---------- ----------

Prof. A. Uzairu Date

Chairman, Supervisory Committee

---------- ----------

Prof. S.O. Idris Date

Member, Supervisory Committee

---------- ----------

Prof. A.O. Oyewale Date

Head of Department

---------- ----------

Prof. S.Z. Abubakar Date

Dean, School of Postgraduate Studies

Dedication

Tomy parents, Mr and Mrs Ajala Abdulsalam and my wife.

Acknowledgments

I wish to express my profound thanks to Almighty Allah for His grace,mercy,love, favours and protection.

I lack words to express my appreciation to my supervisors Prof. A.Uzairu and Prof. S.O.Idris, I must confess I am humbled by the confidence and trust bestowed on me.I am particularly greatful for their corrections, which helped me to straighten the work.I thank them for being patient with me.May the Almighty continue to order their steps.I also thank the Head of Department, academic and technical staff and students of the Department of Chemistry for all their favours.

To my colleagues and friends Alfa Taofeek (Shooting Range), Alfa Tajudeen Olowonjoyin, Seye Adedirin, David Ebuka, Abdulhakeem Olayiwola, Hassan Samuel, Israel Edache, Philip, Zakari Ibrahim, Eric, Tajudeen Jimoh, Abdullateef Jimoh, Jerry Gambo, Muhydeen Olaniyan, Ibrahim Abdulrazak, Taofeeq Adebisi, Zaks, Elijah Agbele and my wife. May the Almighty continue to order your steps and bless you all, I am grateful. And all my family members, may the Almighty reward you abundantly.

Abstract

Quantitative structure activity relationship (QSAR) study was carried out using B3LYP (6-31G) variant of density functional theory and genetic function approximation (GFA) techniques to correlate the chemical structures of coumarin and neolignams derivatives and their biological activities against resistant strains of *Candida albicans*and *Epidermophyton floccosum*.QSAR models generated were evaluated using internal as well as external test set predictions and applicability domain.The results showed that thermodynamic, dimensional, steric, geometrical, quantum-chemical, WHIM, electrotopological state, moment of inertia, lipophilicity and electronic descriptors were responsible for the biological activities of thecompounds.The proposed models provided a good understanding of the anti-*Candida albicans* andanti-*Epidermophyton floccosum* activity of Coumarin and Neolignans derivatives and could be used as guidance for proposition of more potent and safer chemopreventive agents,within the four series of chemical compounds.

Table of Contents

List of Tables

List of Figures

List of Plates

Abbreviations

0D, 1D, 2D, 3D	Zero, one, two and three dimensional descriptor respectively
ACE	angiotensin converting enzyme
ADMET	Absorption Distribution Metabolism and Excretion and Toxicity
AECC	average eccentricity
B3LYP	Becke's three parameter hybrid functional
CA	Candida albicans
DFT	Density functional theorem
GFA	genetic function approximation
HCV	hepatitis C virus
HQSAR	hologram quantitative structure activity relationship
LUMO	lowest unoccupy molecular orbital
MIC	Minimum inhibitory concentration
MMFF	Molecular mechanics force field
MR	molar refractivity
MV	molar volume
NAC	non-albicans Candida
PM3	Parametric 3
pMIC	Logarithmic to base 10 of the Minimum inhibitory concentration
QSAR	quantitative structure activity relationship
RAS	renin-angiotensin system
SAR	Structure activity relationship

ST	surface tension
TE	total energy
tPSA	topological polar surface area
TS	thymidylate synthase

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background of the Study

Candidiasis (Plate 1.1) is a diverse group of infections caused by members of the genus Candida, especially *Candida albicans*(Havlickova *et al*., 2008). These organisms typically infect the skin, nails, mucous membranes and gastrointestinal tract when natural resistance to its overgrowthis reduced. Cutaneous and mucosal manifestations of candidiasis can be divided into several distinct clinical syndromes. Thirty years ago, superficial fungal infections were common, but systemic fungal infections were not as frequent as today (Achkar and Fries, 2010). The incidence of superficial and systemic fungal infection has been increasing because of the increasing incidence of severe diseases (e.g. malignancies or HIV-infection) or immunosuppressant therapies (systemic steroids or chemotherapy) (Havlickova *et al*., 2008).

Epidermophyton floccosum (Plate 1.2) on the other hand contain a number of species from three different genera (Durdu *et al.*, 2017) which invade the keratinized layers of skin, hair and nail causing infections known as dermatophytosis (Peter *et al*, 2010). Dermatophyte infections are frequently encountered and are responsible for 90% of the dermatophyte infections in Britain (Peter *et al.*,2010). In experiments designed to extend the original finding of (Longbottom and Pepys, 1964)with the dermatophyte, many skin diseases such as tinea and ringworm caused by dermatophytes exist in tropical and semitropical areas.

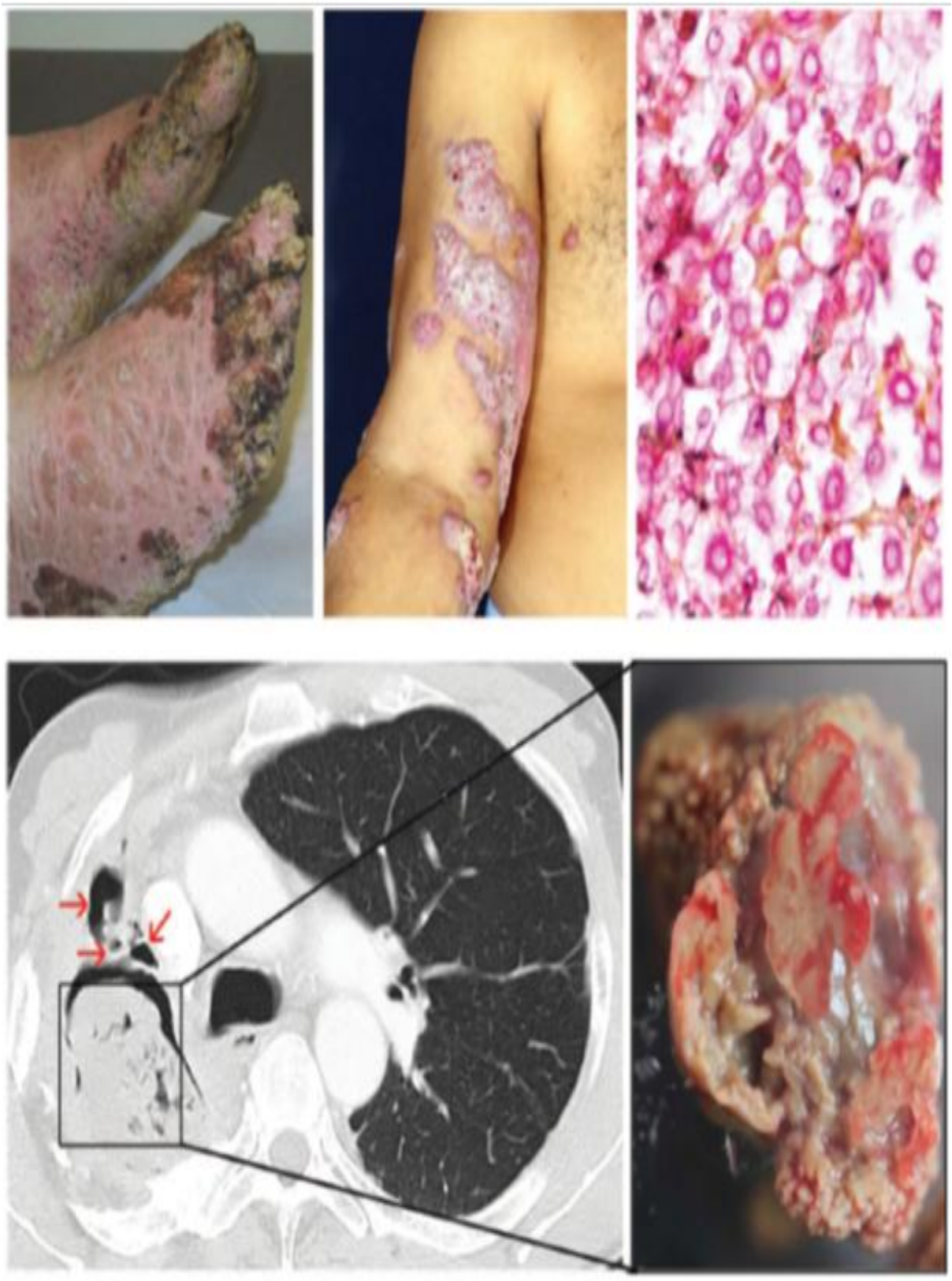


Plate 1.1: *Candida albicans infection* (public health library)

Plate 1.2: *Epidermophyton floccosum* (CDC Public image Library)

1.1.1Occurences of *Candida albicans* and *Epidermophyton floccosum*

Candida albicans is the most common type of yeast infection found in the mouth, intestinal tract and vagina, and it may affect skin and other mucous membranes. If the immune system is functioning optimally, this type of yeast infection is rarely serious. However, if the immune system is not functioning properly, the candida infection can migrate to other areas of the body, including the blood and membranes around the heart or brain.The organism is part of the normal microbial flora in human beings and domestic animals, and is associated with the mucous surfaces of the oral cavity, gastrointestinal tract and vagina. Immune dysfunction can allow this organism to switch from a commensal to a pathogenic organism capable of infecting a variety of tissues and causing a possibly fatal systemic disease (Noverr *et al*., 2004; Noverr *et al*., 2001; Traynor and Huffnagle, 2001). The clinical spectrum of this organism infectionsranges from mucocutaneous to systemic life-threatening infections. The main risk factors that predispose to severe Candida infections are congenital or acquired defects of cell-mediated immunity (CMI), including quantitative and qualitative defects in neutrophils and dysregulated Th-cell reactivity (Blanco and Garcia, 2008).

The *Epidermophyton floccosum* infections are found commonly in scalp, non-hairy, glaborous region of the body, athletes foot, foot, nail, hands, barbers' itch"; bearded region of face and neck, steroid modified (White *et al*., 2008) as shown in Plate 1.1. In general, these fungi live in the dead, top layer of skin cells in moist areas of the body, such as between the toes,the groin, and under the breasts. These fungal infections causeonly a minor irritation. Other types of fungal infections could be more serious. They can penetrate into the cells and cause itching, swelling, blistering and scaling.

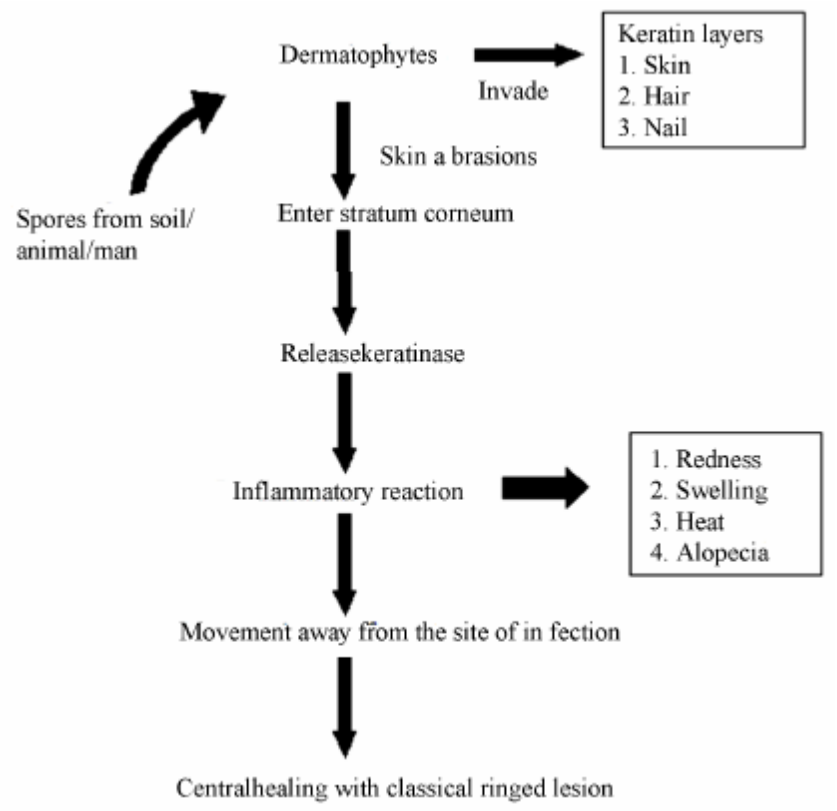


Fig 1.1: The schematic route of entry of dermatophytes into the host system. (Lakshmipathy and Kannabiran, 2010)

1.1.2 Toxicities of *Candida albicans* and *Epidermophyton floccosum*

Fungal infections are notable for their chronicity (e.g.,dermatophyte infections) and nonprotective or injurious inflammatory responses (e.g., dermatophytoses, subcutaneous mycoses, and vaginal candidiasis). In mammals, prostaglandins and leukotrienes can play a dual role in the pathogenesis of inflammatory diseases, both promoting and counteracting inflammatory processes (Nicosia *et al*., 2001).

Dermatophytosis can affect all keratinized areas of the body (hair, skin and nails)(Degreef, 2008). Depending on the region that is affected, the symptoms may vary. If hair is infected (tinea capitis, tinea barbea), there may be hair loss (ectotrix) or breakage (endotrix). On the skin, lesions may look circular or annular and elevated, producing a ringworm infection form. Zoophilic dermatophyte infections are more inflammatory (vesicle, pustules and blisters) than those caused by antropophilic dermatophytes. Infection of human nails may be present as discoloration, dystrophy, hyperkeratosis and occasionally onycholysis. The disease is not fatal. The main effects are aesthetic and will persist until treated with the appropriate medication (Degreef, 2008).

1.1.3 Epidemiology and economic impacts of *Candida albicans* and *Epidermophyton floccosum*

Since the early 1980s, fungi have emerged as major causes ofhuman disease, especially among the immuno-compromised and those hospitalized with serious underlying disease (Blumberg *et al*., 2001; Pfaller *et al*., 2007; Wilson *et al*., 2002). A recent study of the epidemiology of sepsis found that the annual number of cases of sepsis caused by fungal organisms in the United States increased by 207% between 1979 and 2000 (Moss, 2003). The morbidity and mortality associated with these infections are substantial (Golmohammadi and Dashtbozorgi, 2010; Gudlaugsson *et al*., 2003; Lin *et al*., 2001) and it is clear that fungal diseases have emerged as important public health

problems (Mcneil *et al*., 2001). An analysis of trends in infectious disease mortality in the United States (Wu *et al*., 2009),found a dramatic increase in multiple-cause mortality due to mycoses, from 1,557 deaths in 1980 to 6,534 deaths in 1997; the majority of these mycoses-related deaths were associated with *Candida, Aspergillus,* and *Cryptococcus* sp. infections.

Candida species are the fourth leading cause of nosocomialbloodstream infection (BSI) in the United States, accounting for 8% to 10% of all BSIs acquired in the hospital (Wisplinghoff *et al*., 2004). The estimation of the number of cases of nosocomial *candidemia* in the United States (Ibrahim *et al*., 2000; Wisplinghoff *et al*., 2004) as shown in Table 1.1, by working backward from estimates that (i) 2.5% to 10% of all patients admitted to U.S. hospitals will develop a nosocomial infection, (ii) BSIs represent 10% of all nosocomial infections, and (iii) 8% of nosocomial BSIs are caused by *Candida*species. Given these assumptions, the absolute number of cases of nosocomial candidemia ranges from 7,000 to 28,000 annually Table 1.1, If the crude mortality rate of *Candida* BSI is 40%, then 2,800 to 11,200 deaths each year may be associated with nosocomial candidemia. Given that approximately two-thirds of all *Candida* BSIs are nosocomial (Hajjeh *et al*., 2004), the total annual burden of candidemia in the United States is approximately 10,500 to 42,000 infections. These estimates are comparable to those obtained from National Hospital Discharge Survey (NHDS) statistics.

Fungal infections have become increasingly important in hospitalized patients and are a major cause of morbidity and mortality. The costs associated with treating an episode of *Candidemia* have been estimated at $34123– 44536, and the prolongation of hospital stay may average as much as 34 days(Rentz *et al*., 1998).

Table 1.1 : Estimated annual number of cases and associated deaths due to nosocomial BSIs with *Candida* species in the United States[a](Wisplinghoff *et al.*, 2004)

.

Total nosocomial infection rate (%)	Total number of:			
	Nosocomial infections[b]	Nosocomial BSIs[c]	Nosocomial BSIs due to candida[d]	Deaths due to Candida BSIs[e]
2.5	875,000	87,500	7,000	2,800
5	1750,000	175,000	14,0000	5600
10	3500,000	350,000	28,0000	11,200

[b]Assumes that 35 million patients are hospitalized in the United States each year in critical care settings. [c]Assumes that 10% of all nosocomial infections are BSIs.[d]Assumes that 8% of all nosocomial BSIs are due to *Candida* species. [e]Assumes a 40% mortality rate associated with *Candida* BSIs

Candida spp. are the fourth most common cause of bloodstream infections in the US (Wisplinghoff *et al*., 2004) and *Candida albicans* remains the most commonly isolated species(Abi-said *et al*., 1997). This rising prevalence of *Candida* infections has been well characterized in immuno compromised patients(Fraser *et al*., 1992). In recent years, there has been a shift toward invasive *Candidiasis* and *Candidemia* due to non-albicans Candida spp (NAC) in hospitalized patients (Pappas *et al*., 2016). This is significant since many of the NAC species have variable susceptibilities to some commonly prescribed antifungal agents (Marco *et al*., 1998)

This pathogen occurs worldwide and infections are relatively frequent(Seebacher *et al.*, 2008).Historically, *Candida albicans* accounted for 70 to 90% of the isolates recovered from infected patients while other *Candida* species rarely isolated from clinical specimens (Dhib *et al*., 2013). Table 1.2 represents epidemiological data of *Candida* spp. for the last 10 years. In the last few decades, there have been numerous reports of *Candida* infections in India (Basu *et al*, 2003), documented that out of sixty three isolates of yeasts (66.6%) were *Candida albicans*. In another study, *Candida albicans* is the major cause of serious fungal infections in the United States and *Candida* species are the fourth most commonly cultured microbe from blood(Morrell *et al*., 2005).

In addition to differences in the fungal ecology of the different continents, the large use of azoles antifungal agents may have contributed to this progressive shift of the epidemiology of candidemia(Eggimann *et al*., 2005). These infections are associated with a high mortality rate that ranges from 46 to 75%, reflecting the severity of this illness (Pfaller and Diekema, 2007).

Table 1.2: Distribution of *Candida* species in epidemiological surveys of clinical isolates since the last 10 years.

Period of study	Region	Candida albicans (%)	Candida tropicalis (%)	Candida glabrata (%)	Candida parapsilosis (%)	Other Candida species (%)
1991-2000	India	14	38	3	2	26
2000-2003	Switzerland	64	9	15	1	2-9
	Norway	70	7	13	6	1-3
2001	Us	55	9	21	11	2
2001-2002	Paris ICU	54	9	17	14	2-4
2001-2005	India	21.5	35.3	17.5	20	1-3.3
2003	India	45.8	24.7	1	10.5	10-38.4
2004	Japan	41	12	18	23	2
2004	France	55	5	19	13	4
2004-2008	North America	46	8	26	16	1-3
2004-2008	Denmark	57	5	21	4	9
2007-2010	Brazil	34	15	10	24	17
2007-2011	India	30	50	20	-	-
	World	48	11	18	17	4

Table 1.2: Distribution of *Candida* species in epidemiological surveys of clinical isolates since the last 10 years (continuation).

2008-2009	Europe	55	75	6	14	3-4
	North America	43	11	24	17	2-4
	Asia	57	12	14	14	2
2009	Portugal	90	3	5	2	2
2010	Turkey	36	1	-	12	2
2010-2011	Worldwide	43	7	31	20	7

As a result of high mortality rates and prolonged length of hospital stays, *Candida* infections also have their impact on health care costs, which is estimated at one billion dollar per year in the US alone (Rentz *et al*., 1997).

Candida species uncommonly cause vertebral osteomyelitis. A case of lumbar vertebral osteomyelitis was caused by *Candida albicans* and review 59 cases of *Candidal* vertebral osteomyelitis reported in the literature. The mean age was 50 years, and the lower thoracic or lumbar spine was involved in 95% of patients. 83% of patients had back pain for 11 month, 32% presented with fever, and 19% had neurological deficits. The erythrocyte sedimentation rate was elevated in 87% of patients, and blood culture yielded *Candida* species for 51%. *Candida albicans* was responsible for 62% of cases, *Candida tropicalis* for 19%, and *Candida glabrata* for 14% (Miller and Mejicano, 2001).

Invasive infection due to *Candida* species is largely a condition associated with medical progress, and is widely recognized as a major cause of morbidity and mortality in the healthcare environment. There are at least 15 distinct Candida species that cause human disease, but >90% of invasive disease is caused by the 5 most common pathogens, *Candida albicans*, *Candida* glabrata, *Candida* tropicalis, *Candida* parapsilosis, and *Candida* krusei. Each of these organisms has unique virulence potential, antifungal susceptibility, and epidemiology (Pappas *et al*., 2016).

Recurrent vulvovaginal candidiasis (RVVC) can affect approximately 8% of women of reproductive age be defined as four or more attacks of symptomatic candidal vaginitis in a 12-month period (Sobel, 2003). The true incidence of RVVC remains unknown. Estimates over many years suggest that the inci coldence is approximately 5% of women during their reproductive age. A recent study by Foxman et al., which included interviewing 2000 women, determined that the incidence of RVVC in the US is

approximately 8% of women of reproductive age (Sobel, 2003). This almost certainly represents an overestimation. The availability of over the counter (OTC) an timycotic agents precludes any prospective data collection determining the incidence of candidal viginitis in North America (Sobel, 2003).

Dermatophytes occurs worldwide and infections are relatively frequent (Clayton and Hospital,1992).The main pathogen in northern Europe and North America is *Trichophyton rubrum*, and *Microsporum canis*. Zoophilic dermatophytes are more common in southern Europe and Arabic countries. Farm worker are more susceptible to infection with tinea barbea (Guo *et al*., 1996). In the case of tinea cruris, men are more susceptible than women (Marchisio *et al*., 1996).

The distribution of dermatophytes (in Germany), isolated from skin and nail lesions has changed significantly within the last 100 years *Epidermophyton floccosum* and *Microsporum audouinii*dominated among human pathogenic dermatophytoses in the twenties, whereasTrichophyton rubrum was almost insignificant (Ghannoum and Rice, 1999) . A completely different spectrum of dermatophyte species has been reported from Islamic states like Isfahan, Iran, Crete and Greece among others. A total of 1,213 patients from the Tehran area, Iran,suspected to have dermatophytic lesions were examinedover a 3-year period (1999–2001), in this study, *Epidermophyton floccosum* was the most frequent dermatophyte (31.4%) (Falahati *et al*., 2003). In another study encompassing 2,203 patients with dermatomycoses in Isfahan, Iran, the prevalence of clinical forms and of the causative agents of dermatophytoses were determined, Epidermophyton floccosum was 17.6% (Dehghan, 1997).

1.1.4 Historical efforts in curing and managing *Candida albicans* and *Epidermophyton floccosum*

The burden of Invasive Candidiasis (IC) is tremendous in terms of morbidity, mortality and cost, and it is clear that we must do more than seek better therapeutic agents if we

are to impact this burden (Diekema and Pfaller, 2004; Fridkin, 2005b; Wisplinghoff *et al*., 2004). Fungal BSIs have been shown to have some of the highest rates of inappropriate initial therapy and hospital mortality among all etiologic agents of BSI examined (Ibrahim. *et al*, 2000; Morrell *et al*., 2005). The most common cause of inappropriate therapy for fungal BSI is the omission of initial empirical therapy (Morrell *et al*., 2005), such omission or delay in therapy has been linked directly to mortality (Blot *et al*, 2002; Morrell *et al*., 2005; Pfaller *et al*., 2007). Thus, despite an impressive array of new, potent, and nontoxic antifungal agents, we are failing in the management of these infections (Miller and Mejicano, 2001; Morrell *et al*., 2005; Puzyn *et al*., 2010). Lack of specific clinical findings and slow, insensitive diagnostic testing complicate the early recognition and treatment of IC (Pfaller *et al*., 2006).

Given the substantial excess mortality due to *candidemia* and the difficulties encountered in administering early and effective antifungal therapy (Puzyn *et al*., 2010), better methods of prevention will decrease candidemia-associated mortality more effectively than will advances in therapy (Diekema and Pfaller, 2004; Fridkin, 2005b). Prevention of nosocomial candidemia is similar to that of many other nosocomial infections and should involve three "low-tech" strategies (Diekema and Pfaller, 2004), the first being the implementation of intensive programs to maximize compliance with current hand hygiene recommendations. Such programs are essential; although seemingly simple, compliance with hand washing recommendations among health care providers occurs only 40% of the time in settings where it is indicated (Laverdiere *et al*., 2002). Both alcohol and chlorhexidine are effective in killing *Candida* species on the hands of health care workers (Apic and Didier, 2002) and will decrease the risk of patients acquiring *Candida* colonization and subsequent infection in the health care setting. Second, strategies to improve adherence to current recommendations for

placement and care of central venous catheters (Fridkin, 2005b) are also necessary. An educational program emphasizing essential components of these guidelines successfully reduced catheter-related BSIs due to both bacteria and *Candida* in an SICU (Coopersmith *et al*., 2002): catheter-related candidemia was reduced from 12% of all BSIs to 0% in the 18 months following the educational program. Finally, given the importance of antibiotic exposure as a risk factor for candidemia, control of antimicrobial use, especially those with antianaerobic activity (Blumberg *et al*., 2001), as well as piperacillintazobactam and vancomycin (Lin *et al*., 2005), is an important component of candidemia prevention. These three strategies improved hand hygiene, optimal catheter placement and care, and prudent antimicrobial use should be primary in the approach to prevention of morbidity and mortality resulting from nosocomial candidemia (Diekema and Pfaller, 2004; Fridkin, 2005b). These efforts may also reduce candidemia-related costs by reducing the need to treat central line-related candidemia (Fridkin, 2005a, 2005b). The remaining preventive strategy, antifungal prophylaxis, must always be considered secondary to the three low-tech approaches and should be applied only when the rate of candidemia remains elevated despite assiduous application of these measures, preferably in a subpopulation within the ICU with a cumulative incidence of IC approaching or exceeding 10% (Diekema and Pfaller, 2004). Antifungal prophylaxis has proven to be effective in decreasing mucosal candidiasis and IC in neutropenic patients (Marr *et al*., 1997). Administration of fluconazole (400 mg/day) during neutropenia has proven effective in decreasing infections due to *Candidaalbicans, C. tropicalis,* and *C. parapsilosis*(Abi-said *et al*., 1997; Antoniadou *et al*., 2003). In contrast, the relative benefits, harms and the cost-effectiveness of antifungal prophylaxis in nonneutropenic critically ill patients remain incompletely defined (Calandra and Marchetti, 2002; Sobel and Rex, 2001), with resultant wide

variations in clinical practice (Eggimann *et al*., 2005; Pappas *et al*., 2004). Nonetheless, the implementation of targeted antifungal prophylaxis has been shown to be effective in certain ICU settings (Blumberg *et al*., 2001; Calandra and Marchetti, 2002; Lipsett, 2004; Vincent *et al*., 1998), although the generalizability of these findings has been questioned (Diekema and Pfaller, 2004; Sobel and Rex, 2001).Management of patients with systemic candidosis might include measures to prevent cross infection and handwashing with disinfectants that are active against candida (Lee *et al*., 1985). The cutaneous mycoses are superficial fungal infections of the skin, hair or nails. Essentially no living tissue is invaded, however a variety of pathological changes occur in the host because of the presence of the infectious agent and its metabolic products. The usual approach to the management of cutaneous infections is to treat with topical agents if possible, but nail and hair infections, widespread dermatophytosis and chronic non-responsive yeast infections are best treated with oral antifungal agent (Ellis, *et al* 2006).

1.1.5 Risk factors of *Candida albicans and Epidermophyton floccosum*

Candidasis is an acute or chronic infection produced by *Candid*a, generally limited to the skin and mucous membranes, but it could produce a serious systemic disease (Gamboa *et al*., 2006). The risk factors associated with candidemia and *invasive candidiasis* (IC) have been well established and have not changed substantially in the past 2 decades (Diekema and Pfaller, 2004; Mun *et al*., 2000; Pfaller *et al*., 2007; Wey *et al*., 2015). Those determined to be independent risk factors for IC on the basis of multivariate analysis include exposure to broad-spectrum antimicrobial agents, cancer chemotherapy, mucosal colonization by *Candida* spp., indwelling vascular catheter (especially central venous catheter), total parenteral nutrition (TPN), neutropenia, prior surgery (especially gastrointestinal), and renal failure or hemodialysis (Blumberg *et al*.,

2001; Diekema and Pfaller, 2004; Morrell *et al*., 2005; Mun *et al*., 2000; Piarroux *et al*., 2004; Sobel and Rex, 2001; Wey *et al*., 2015)

Epidermophyton floccosum on the other hand, are fungi obtaining a mid-transmittable disease which are acquired from infected animals or birds and fomites. Detection of dermatophyte Texas is correlated to epidemiological apprehension. These are important to manage infection and public health issues associated with types of Dermatophytosis. Traditionally, the dermatophytosis is normally referred to as "tinea" or "ring-worm" infections (Lakshmipathy *et al*., 2010). In humans, pruritus is a widespread symptom. The skin lesion is usually characterized by inflammation with erythema, scaling and occasionally blister formation. The habitual signs of inflammatory reactions such as redness, swelling, heat and alopecia are distinguishing at the infection position (Laksmipathy *et al*., 2010). The identification of dermatophytes is based on methods that focus on morphological, physiological, ecological and genetic features. Anthropophilic and zoophilic dermatophytes has mostly been recognized through internal transcribed spacer (Sharma *et al*., 2015).

Despite the availabilty of the aforementioned drugs for combating *Candida albicans* and *Epidermophyton floccosum*, there have been increasing cases of infections attributed to these organisms. Even more worrisome is the pace of resistance to the existing antibiotics by these pathogenic organisms pushing researchers to ask if humanity have reached the hypothetical post anti-biotic era (Alanis, 2005). This has necesitated an urgent need for new anti-*Candida albicans* and *Epidermophyton floccosum*.

Certainly, rational prediction of the biological activities of potential drugs candidates will undoubtedly helps in quick , safe and economical discovery and development of *Candida albicans* and *Epidermophyton floccosum* drugs by relegating to the

background the old try and error methods which is costly, time consuming and constituting serious problem to the environment due to large waste usually released.This can only be achieved via Quantitative Structural Activity Relationship (QSAR) modelling. This has the knock-on effect of reducing cost, a major consideration in all commercial companies. Structure activity relationship (SAR) studies are usually carried out by making minor changes to the structure of a lead to produce analogues and assessing the effect that these structural changes have on biological activity. The investigation of numerous lead compounds and their analogues has made it possible to make some broad generalizations about the biological effects of specific types of structural changes. The success of the SAR approach to drug design depends not only on the knowledge and experience of the design team but also a great deal of luck (Puzyn *et al*., 2010).

1.2 Statement of the Research Problem

The development of resistance to existing anti-biotics by *Candida albicans* and *Epidermophyton floccosum* is still unacceptably high adding a serious threat in modern healthcare and contributing heavily to the world disease burden in the form of higher treatment costs, increased morbidity and mortality and in some cases permanent loss of specific drug therapies. Skin diseases causing fungus are therefore confronted with some major problems:

a. Increase resistance in strains of *Candida albicans* and *Epidermophyton floccosum* to classical drugs, the treatment costs and the fact that most antifungal drugs have only fungistatic activity, justify the search for new strategies (Jose *et al*., 2006).

b. Globally significant increase in the prevalence of systemic fungal infections during the past decades which is due to greater use of broad-spectrum antibiotics (Graybill, 1992).

c. The number of antifungal drugs available at present is very small, with much greater difficulty in production, with many side-effects, and with the possibility of the appearance of resistance (Blanco and Garcia, 2008)

d. Diagnosis of these diseases can be problematic because of the difficulty of interpreting the very different clinical pictures in individuals in the presence of colonization, infection and/or disease (Blanco and Garcia, 2008).

However, the increasing problem of multi-drug resistant strains is the major challenge for the investigation and designs of novel drug candidates which are not only active against stable drug resistant *Candida albicans* and *Epidermophyton floccosum* but also shorten the length of therapy. Also the intimidating cost of drug development both in terms of money and time has contributed to fungi drug development.

1.3 Significance of the Study

This study will provide information on the safety and efficacy of Coumarin and Neolignans derivatives with anti-*Candida albicans* and *Epidermophyton floccosum* activity. This studies identifies the activity of these compounds that can be used in drug discovery and could be sources of lead compounds for new anti-*Candida albicans* and *Epidermophyton floccosum*drug development.

1.4 The Research Questions

What structural and electronic properties of Coumarin and Neolignans derivatives determine its activity and what can be altered to improve this activity?

1.5 Research Hypothesis

The alternatively hypothesis to this research includes :

The size of the observed anti-Candida albicans and Epidermophyton floccosum inhibitory abiliities of Coumarin and Neolignans and their activities are functions of the

empirical poperties which makes the descriptor of the total chemical structure of the compounds under study.

The null hypothesis to this research includes:

The observed activity and inhibitory activities of Coumarin and Neolignan compounds against *Candida albicans* and *Epidermophyton floccosum* is independent of the descriptors of the total chemical structure of the compounds.

1.6 The Research Design

QSAR methodologies have the potential of decreasing substantially the time and effort required for the discovery of the new medicines (Tong *et al*., 2005). A major step in constructing the QSAR models is to find a set of molecular descriptors that represents variation of the structural properties of the molecules (He and Jurs, 2005). The QSAR analysis employs statistical methods to drive quantitative mathematical relationships between chemical structure and biological activity (Cronin, 2005). Thus, the use of the QSAR in the development of a theoretical model to predict the biological activity of a set of compounds is very important. The strategy used in the QSAR methodology includes the following steps:

a. selection of a data set;

b. generation of the molecular structures;

c. optimization of the geometry of the molecular structures by appropriate method;

d. generation of several structural descriptors;

e. application of variable selection or/and methods data reduction of the calculated descriptors;

f. regression analysis; and finally

g. evaluation of the validity and predictability of the developed QSAR models (Kiralj and Ferreira, 2009).

1.7 Theoretical Framework of the Study

Computational chemistry represents molecular structures as numerical models and simulates their behavior with the equations of quantum and classical physics. Available programs enable scientists to easily generate and present molecular data including geometries, energies, and associated properties (electronic, spectroscopic, and bulk). The usual paradigm for displaying and manipulating these data is a table in which compounds are defined by individual rows and molecular properties (or descriptors) by the associated columns. A QSAR attempts to find consistent relationships between the variations in the values of molecular properties and the biological activity for a series of compounds so that these ‘‘rules’’ can be used to evaluate new chemical entities (Patel *et al*., 2014). A QSAR generally takes the form of a linear equation

$$Biological\ Activity = Constant + (C_1.P_1) + (C_2.P_2) + (C_3.P_3) + \cdots \qquad 1.1$$

where the parameters P_1 through P_n are computed for each molecule in the series and the coefficients C_1 through C_n are calculated by fitting variations in the parameters and the biological activity (Patel *et al*., 2014).

While searching, one finds numerous hits for lead candidates, and thus lead optimization is hindered. To get more target structural information, high-through put proteincrystallization can be explored (Khan *et al*., 2011). Cheminformatics methods must be applied while generating data using highthrough put techniques in order to assure that good absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties are achieved while making and screening compounds, this approach is called a multi-parametric optimization strategy (Khan *et al*., 2011). Several physicochemical properties of drug molecules such as aqueous solubility, partition coefficient (log P),

distribution coefficient (log D), ionization constant (pKa), and topological polar surface area (tPSA) play an important role in majority of the processes.

Statistical or chemometric techniques form the mathematic foundation for building a QSAR model. Most easilyinterpretable method was found to be linear regressionanalysis among various statistical methods for QSAR . These regressions represent direct correlation of independentvariables (x) with a dependent variable (y). Thismodel can be considered for prediction of **y** from the dataof **x** variables. This can either belong to qualitative orquantitative set of system (Khan *et a*l., 2011).

1.8 Aim and Objectives of the Study

The aim of this study is to correlate the molecular properties of Coumarin and Neolignans derivatives with experimental data from biological activity of *anti-candida albicans and anti-Epidermophyton floccosum*

The aim can be achieved through the following objectives:

a. Selection of a data set.

b. Generation of the molecular structure

c. Optimization of the geometry of the molecular structure by appropriate method

d. Generation of several structural descriptors

e. Application of variable selection or / and methods data reduction of the calculated descriptors

f. Regression analysis; and finally

g. Evaluation of the validity and predictability of the developed QSAR models.

1.9 Scope and Limitation of the Study

Any QSAR method wouldn't be tried for a dataset unless the experimental expects that the study will provide useful three-dimensional structure–activity insights. Since scientists know that it is the properties of molecules that govern their biological properties, it is especially gratifying to see a summary of how changes in structure change biological properties. Methods that do not provide such a graphical result are often less attractive to the scientific community.

A major factor in the continuing enthusiasm for QSAR comes from the proven ability of several of the methods to forecast correctly the potency of compounds not used in their derivation (Eriksson *et al*., 2003).Validation by forecasting compounds not used in the derivation is usually included in QSAR reports, this ability to forecast affinity is gaining new respect as scientists realize that we are far away from the hoped-for fast and accurate forecast of affinity from the structure of a protein-ligand complex.Thus scientists whose primary focus is laboratory work can use the computer to gain insights into the structure–activity relationships of their compounds.

The following are the limitation of QSAR

a. Lack of sufficient number of training molecules.

b. Consideration of only two-dimensional structures.

c. Insufficient parameters for relating drug–receptor interactions such as

Hammett constant.

d. Unavailability of specific physiochemical parameters.

e. Unavailability of representation of stereochemistry.

f. No unique solution with high risk of failure and chance correlations.

g. Requirement of knowledge of substituent constants and chemistry utilized to

design a molecule.

h. Lack of suggestion to synthesize a new compound through classical QSAR equations with no graphical output (Puzyn *et al*., 2010).

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 *Candida albicans* and *Epidermophyton floccosum*

Candidiasis is a yeast infection cause by *Candida albicans.* It is a serious disease that affects over four billion people a year worldwide(Segal, 2005). Most Candida infections occur on epithelial surfaces such as in the mouth, nail, vaginal, and skin regions, these surface infections are easily treatable using topical or oral anti-fungal agents (Segal, 2005). In some cases, the Candida infection passes into the blood stream, becomes systemic and infects many vital organs. This systemic infection is commonly seen in patients that have compromised immune systems such as those with HIV, diabetes, cancer, or those who are recovering from an organ transplant or other major surgeries. Systemic Candidiasis is fatal with a 30%-60% mortality rate (Vincent *et al*., 1998). This increase in infectionsis associated with excessive morbidity and mortality (Gudlaugsson *et al*., 2003; Wilson *et al*., 2002) and is directly related to increasing patient populations at risk for the development of serious fungal infections. Some strains, specifically *Candida albicans*, are becoming increasingly resistant to conventional antifungal treatments (Pfaller, 2007). For example, in 2013 the United States Center for Disease Control (CDC) reported that fluconazole resistant *Candida albicans* poses a serious threat and is responsible for approximately 3,400 cases annually, including those that are superficial as well as those in the bloodstream (Murtaza and Mukhtar, 2015). The lifestyle and virulence strategies used by *Candida albicans* are the subject of intense research (Pfaller, 2007).

Dermatophytes on the other hand, are a group of morphologically and physiologically allied molds which have the hazard at the global plane as these are generally causative

agent of fungal infections. They hold two imperative properties: they are keratinophilic and keratinolytic agents (Sharma *et al*., 2015). They have an aptitude to digest keratin in their saprophytic circumstances and consume it as a substrate. The World Health Organization estimates universal occurrence of dermatomycoses to be about 20% (Marques *et al*., 2000). Infections which are caused by dermatophytes are known as dermatophytosis. Dermatophytosis is caused by the genera *Microsporum, Trichophyton* and *Epidermophyton*. *Microsporum* and *Trichophyton* are human and animal pathogens. *Epidermophyton* is a human pathogen. *Anthropophilic* species are responsible for the majority of human infections. The dermatophytes are a group of closely related fungi that have the capacity to invade keratinized tissue (skin, hair and nails) of humans(Maraki*et al*., 2007)

2.2 Treatment Protocols of *Candida albicans* and *Epidermophyton floccosum*

The burden of Candidiasis is tremendous in terms of morbidity, mortality, and cost(Bisno *et al*., 2005), and it is clear that we must do more than seek better therapeutic agents if we are to impact this burden (Diekema and Pfaller, 2004; Fridkin, 2005b) and in particular, side effects(Bissell *et al*., 1993; Cannon *et al*., 1995; Fotos and Paso, 1992; Odds, 2003).

Also, the confirmation of dermatophyte infections is very essential before prescribing suitable therapy. The spontaneous healing of dermatophytosis is uncommon which makes the implementation of treatment more important. A variety of antifungal agents are available in the market. However, the selection of the appropriate antifungal is very important for the effective treatment.

2.3 Antimycotic Drugs of *Candida albicans* and *Epidermophyton floccosum*

The classes of antimycotic drugs include those that are either fungicidal or fungistatic. The fungistatic agents include: fluorinated pyrimidine analogues (*e.g.* 5-fluorocytosine)

and azoles (*e.g.* ketoconazole). The fungicidal agents include polyenes (*e.g.* nystatin/amphotericin B), echinocandins (*e.g*. caspofungin) and allylamines (*e.g.*terbinafine). Also griseofulvin (From miscellaneous class) is one of the common classes of antifungal agents used for the treatment of dermatophytosis. The action of these agents was described below.

2.3.1 Pyrimidine analogues

Flucytosin or 5-Flurocytosine (5-FC) is a synthetic antimycotic compound and is administered intravenously for systemic candidal infections. 5-FC has no intrinsic antifungal capacity, but after uptake and transformation by susceptible fungal cells, its metabolites become incorporated into DNA and RNA molecules (Vermes *et al*., 2000).

2.3.2 Azoles

The target of azole based antimycotic drugs (*e.g.* ketoconazole and fluconazole) is the fungal enzyme P_{450} cytochrome lanosterol 14α-demethylase (Ghannoum and Rice, 1999). The function of this enzyme is the demethylation of the ergosterol precursor at the C4 position. Topical azoles are cheaper, potent having broad spectrum activity against dermatophytes, moulds and yeasts. In inflammatory tinea infections, the use of azoles in combination with corticosteroids has been recommended. A large number of topical azole derivatives are available in the market: bifonazole, butoconazole, clotrimazole, coroconazole, fenticonazole, flutrimazole, isoconazole, ketoconazole, oxiconazole, omoconazole, econazole, terconazole (Giusiano *et al*., 2008).

2.3.3 Polyenes

Polyene antifungal agents were the standard therapy for systemic fungal infections prior to the development of azoles. Polyenes interact with fungal membrane sterols resulting in the production of aqueous pores in the membrane. Of the two well-known polyenes, amphotericin B is more effective at killing fungal organisms thannystatin. However, the

former has a narrow therapeutic window, which limits its usefulness in the clinic (Ghannoum and Rice, 1999).

2.3.4 Echinocandins

Echinocandins (*e.g.* caspofungin and micafungin) are semi-synthetic derivatives of fungal fermentation products (Bellmann, 2007). They are cyclic lipopeptides that specifically inhibit the fungal ß-glucan synthase. They are administered intravenously for systemic candidal infections (Denning, 2003). ß-glucan inhibitors act as specific noncompetitive inhibitors of ß-(1,3)-glucan synthase.

2.3.5 Allylamines

Allylamines (*e.g.* terbinafine and naftifine), these agents are typically applied topically to dermal surfaces (skin/fingernails) and can also be used to treat oral candidiasis in conjunction with azoles. Both oral as well as topical formulations are used for treating tinea conditions. Terbinafine belonging to this class is the drug of choice which is given orally in nail infections as well as in tinea conditions that respond poorly with topical or first line of treatment. The use of oral terbinafine is restricted in children below the age of 2 years. It is also an effective drug for the treatment of tinea pedis in adults (Millikan, 2010). Naftifine, terbinafine and butinfine are topical derivatives of allylamines and benzylamines. The topical derivatives of allylamines possess inherent anti-inflammatory activity. The butinfine is an effective topical agent against dermatophytes as well as *Candida albicans* (Brennan and Leyden, 1997).

2.3.6 Griseofulvin

Thisis a drug of choice against dermatophytosis. This drug is more effective than fluconazole and is equally effective as compared to itraconazole and terbinafine. Griseofulvin has advantages over other antifungal agents in that, it is much cheaper and has no side effect even at higher doses. The efficacy of this drug can be increased by

simultaneous ingestion of fatty food. Both oral as well as topical forms of griseofulvin are available (González *et al*., 1996).

2.4Evaluation of Drug Resistance by *C.albicans* and *E.floccosum* to Antimycotics

Resistance to various classes of antifungal agents was reviewed recently (Niimi *et al*., 2010). Antifungal resistance has been well documented and the mechanisms have been elucidated through *in vitro* studies (Niimi *et al*., 2010).

2.5 Exploiting Compounds Active Against New Target

Traditional medical practice relied on the use of plants and their components as new sources for drugs. These include bronchodilators (theophylline from Chinese tea), pain inhibitors (morphine from poppies) and antineoplastics (vincristine from periwinkle)(Licciardi and Underwood, 2011).

2.5.1 Coumarin

Coumarins are a group of polyphenolic compounds isolatedfrom plant product Tonka bean, *coumarou* in 1820. Coumarins are colorless and crystalline phytochemical substances. It is an oxygen heterocycle, which occurs either in free or combined form with glucose. They belong to the family of benzopyrones, which consists of benzene ring joined by a pyrone ring as shown in Figure 2.1 (Lacy and Kennedy, 2004; Ojala, 2001). Benzopyrones are subdivided into benzo-α-pyrones and benzo-γ-pyrones of

[A] $\alpha - Benzopyrone.$ [B] $\gamma - Benzopyrone$

Fig. 2.1 : The chemical structures of benzopyrone subclasses with the basic Coumarin structure.

which coumarins and flavonoids are prime members of benzo-α-pyrones and benzo- γ – pyrones class respectively. Studies have shown that they function as blood diluting agent and also exhibits anti-fungicidal property. Researches have also proved the selective

cytotoxicity of coumarins for tumor cells and also the effect of coumarins in the regulation of immune response, cell growth and differentiation (Vianna *et al*., 2012).

Coumarins possess a wide range of pharmaceutical actions and biological functions and has great importance in the use of medicine and also used for treating various clinical conditions (Lacy and Kennedy, 2004). They exhibit significant pharmacokinetic activity due to its rapid absorption and metabolism (Lacy and Kennedy, 2004) in the body. Coumarins have a significant effect on physiological, bacteriostatic and anti-tumor activity. Table 2.1 shows the list of coumarins and its therapeutic functions.

There has been a dramatic increase in pathogen resistance to both pharmaceutical and agrochemical antimicrobial agents (Ojala, 2001). New prototype compounds are needed to address this situation. Successful discovery of novel natural product antimicrobials has necessitated the development of new bioassay techniques and protocols that allow for the detection of small amounts of biologically active chemicals, which should be selective enough to determine optimum target pathogens, and amenable to the analysis of complex mixtures (Ojala, 2001). Antimicrobial activities have been evaluated with diverse settings often difficult to compare. There are reports on efficacies of pure coumarins against Gram-positive and Gram-negative bacteria as well as fungi (Ojala, 2001), also extracts have shown activities, *e.g*. methanol extract from *Mitracarpusscaber* against *Staphylococcus aureus* and *Candida albicans*(Bisignano *et al*., 2000). Free 6-OH in the coumarin nucleus has been found to be important for antifungal activity,

Table 2.1: Coumarins and its therapeutic role (Rohini, and Srikumar, 2014)

Coumarin	Therapeutic Role
Warfarin 4-hydroxycoumarin	Malignant melanoma
Pyranocoumarins	Renal cell carcinoma
Coumarin 7-hydroxycoumarin	Leukemia
Benzopyrones	Breast cancer
Psoralens	Skin disorder like mycosis fungoides,psoriasis vitiligo
Warfarin	anticoagulant
Pyranocoumarins	prostate cancer

while the free hydroxyl group at position 7 is important for antibacterial activity (Sardari *et al*., 1999)

2.5.2 Neolignans

Neolignans are a large group of naturally occurringphenols which are widely spread within the plant kingdom and are derived from the shikimic acid biosynthetic pathway. These compounds show dimeric structures formed by a □, □’-linkage between two phenyl propane units with a different degree of oxidation in the side-chain and a different substitution pattern in the aromatic moieties(Bernini *et al*., 2009). For nomenclature purposes, the C_6–C_3 unit is treated as propylbenzene and numbered from 1 to 6 in the ring, starting from the propyl group, and with the propyl group numbered from 7 to 9, starting from the benzene ring. With the second C_6–C_3 unit, the numbers are primed. When thetwo C_6–C_3 units are linked by a bond between positions 8 and 8’,the compound is named as a “lignan”. In the absence of the C–8 to C–8’ bond, the dimer formed from the two C_6–C_3 units is a “neolignan”. Compounds in which an ether oxygen atom provides the linkage between the two C_6–C_3 units, are also classifed under neolignans and are called oxyneolignans as shown in Figure 2.2. Because of the biological activities of Lignans, such as antioxidant(Johan and Sj, 2005; Yamauchi *et al*., 2005), antimicrobial(Kiyama *et al*., 2007)antitumour(Shan *et al*, 2007), anti-inflammatory(Jeng and Hou, 2005)and antiviral properties of lignans have been used for a long time both in the ethnical and conventional medicine (Yamauchi *et al*., 2005).

2.6 Need for an Alternative Tool

In order to reduce the costs, in 1980, high throughput screening technique advent and made possible robotized screening. Through this process, hundreds of thousands of

Fig. 2.2 :backbone of a Neolignane molecule

individualcompounds were screened per drug target per year(Damme, 2009). In responseto the increasing demand for new compounds, researchers started using combinatorial chemical technologies (Khan *et al*., 2011). The beauty of this modern drug designing is

that now we can tailor the drug with desired combinations computationally before going for experimental laboratory work. It covers the available web tools/databases and in silico techniques used in virtual screening and drug discovery processes to reduce the wet lab economy and time.However, apart from in vitro assays and in vivo experiments, application of in silico QSAR and Absorption Distribution Metabolism and Excretion and Toxicity(ADMET) in the prediction of biological activity and bioavailability are proving beneficial in drug discovery process (Khan *et al*., 2011).

Drug designing is an approach of finding drugs by designs, based on what the drug is targeting. In the real work, the researchers exploit all of the possible approaches to design or find good candidates for drug. Drug designing process starts with the scanning of a compound library, which is a collection of hundreds of thousand compounds from various sources, like nature, previous drug data or from traditional remedies. These compounds are searched for the drug like properties. The first step in this process is the high- throughput screening, followed by the other one called virtual screening. Virtual screening is the technique to analyze large chemical databases in order to identify potent drug candidates. Approximately 100 potent drug candidates reaches this stage, which are subjected for Quantitative Structure Activity Relationships (QSAR) decrease the chances of selecting a wrong candidate by establishing the relationship between the structure of a drug and its activity. At the end of this step approximately 20 drugs like compounds are only left, which are further screened for their Absorption, Distribution, Metabolism and Excretion (ADME) properties (Puzyn *et al*., 2010).

Further optimization of these candidates finally leads into the discovery of a new drug. Figs. 2.3, 2.4 and 2.5 shows how different approaches carried out during drug designing process. The first step in drug discovery is the search of a suitable lead molecule whose structure binds well with the target protein and serves the first and foremost proposal of

a drug. Ideally, the lead structure binds very well with the target protein and the compounds used as potential leads could come from many sources (Khan *et al*., 2011).

A majority of leads discovered in recent programs are derived from a collection that is now referred to as a 'Library'. These may take the form of natural product libraries, peptides libraries, carbohydrates libraries and/or small molecule libraries based on a variety of different molecular scaffolds. The use of combinatorial chemistry to produce small molecule libraries and the growing ability to master more complex chemical conversions on the solid phase have placed increased attention on the generation of more drug like libraries(Lipinski *et al*., 2001).

A key requirement for a successful virtual screening is access to a large and diverse library or a database. It has been argued that a good database is one that is the most diverse when it comes to its sampling of chemical space. Among all possible libraries, natural products collections arguably represent the highest degree of chemical diversity. Screening of chemical compounds library, to search the potential leads against a specific target with the help of computer is called as virtual screening. Databases of compounds with millions of small molecules are available commercially as well as on the web for public use. Virtual screening technology uses high throughput docking, homology searching and pharmacophores searches of 3D databases

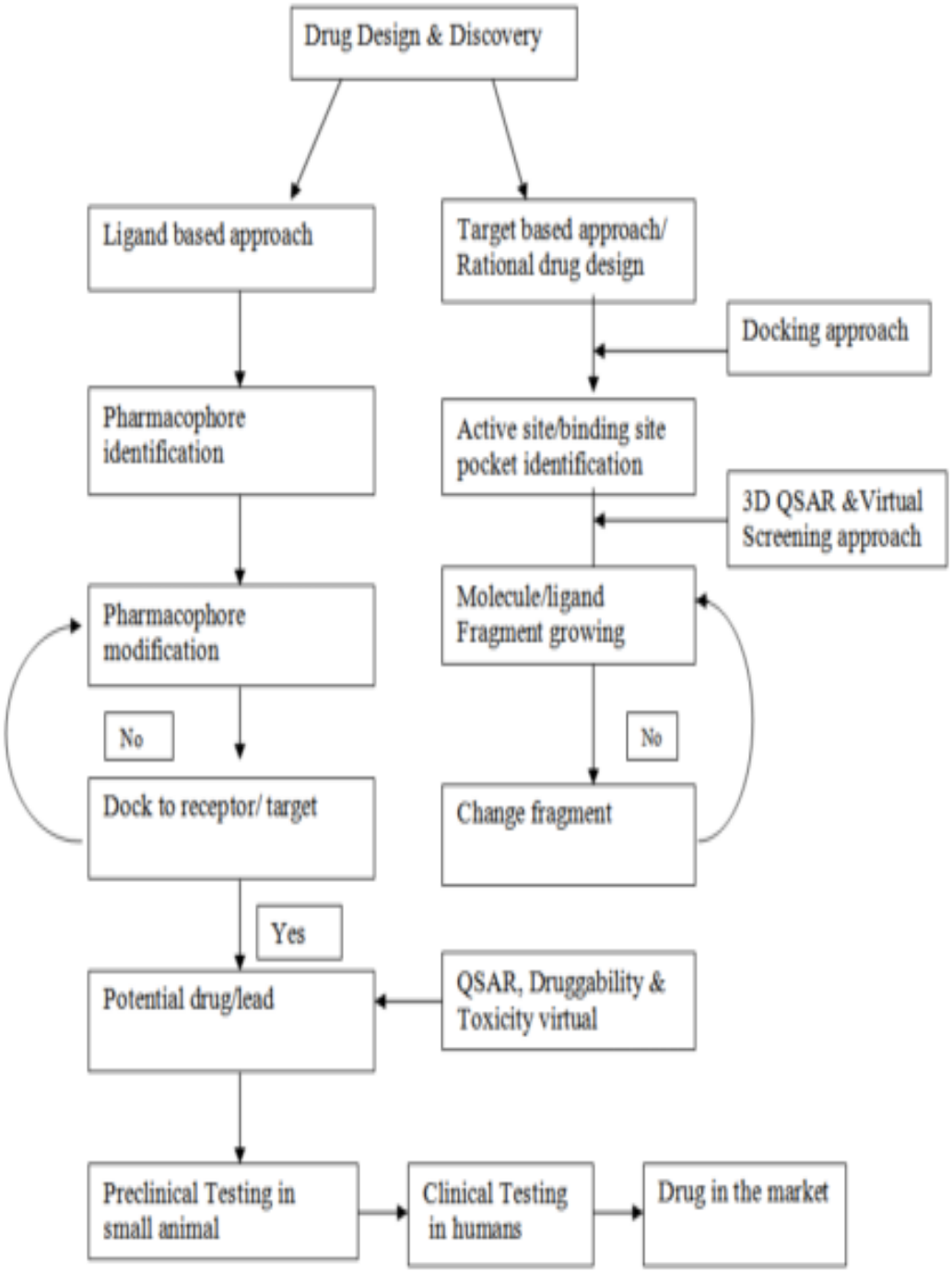


Fig 2.3: Different steps of drug design and discovery process (Khan *et al*.,2011)

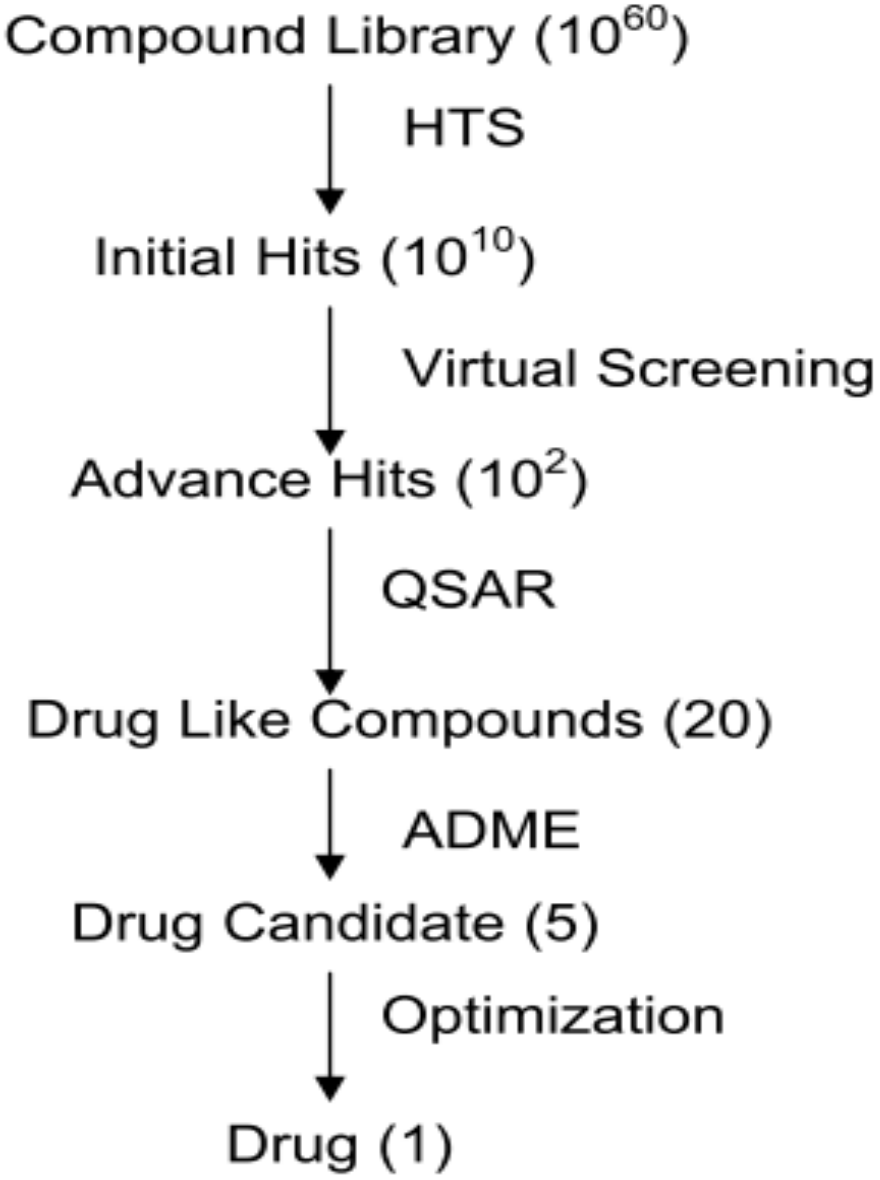


Fig 2.4: Virtual screening methods in drug discovery (Khan *et al*.,2011)

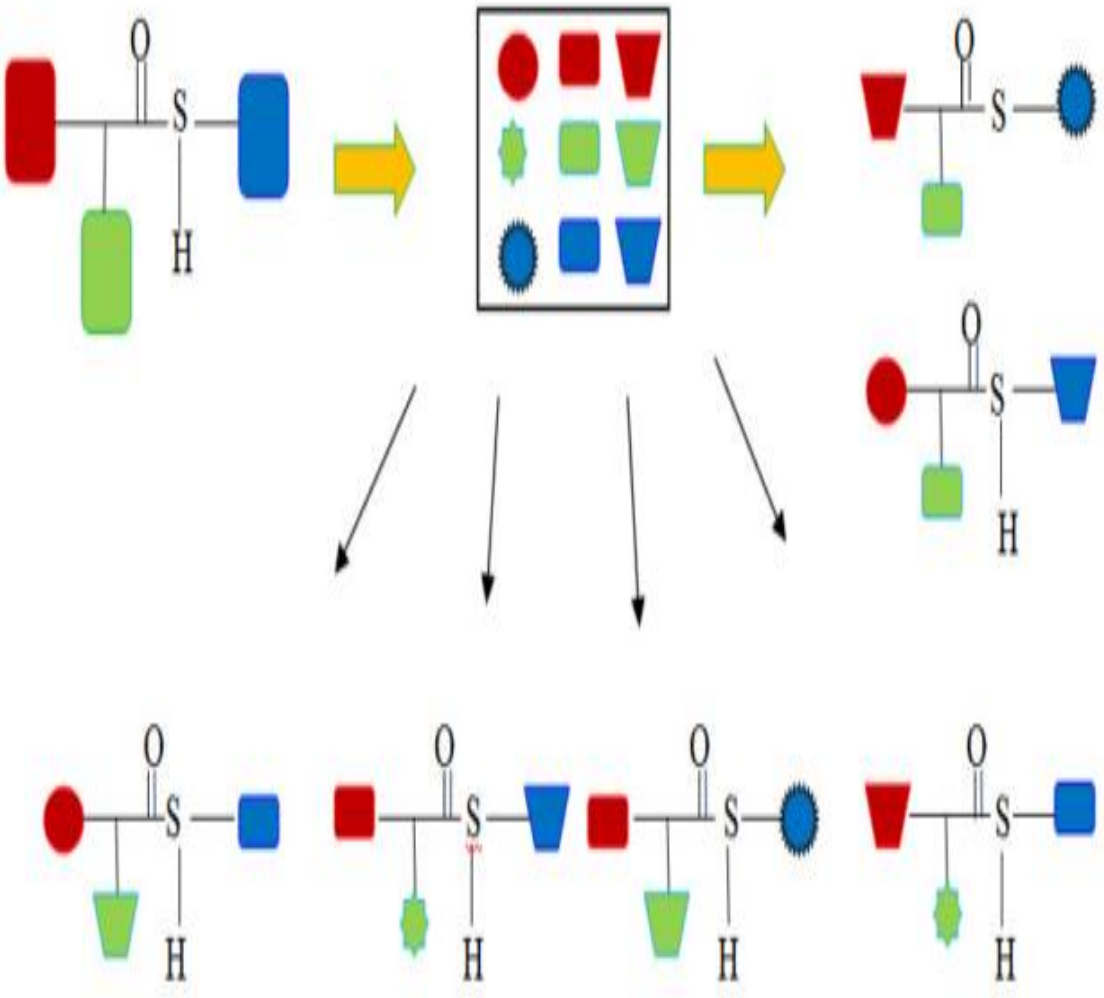


Fig 2.5: Combinatorial chemistry scheme for compound library development (Khan *et al*.,2011)

(Oprea and Matter, 2004; Seifert *et al*., 2003).This is perhaps the cheapest way to identify a lead and several cases have already proven successful using this method. Some important features to consider during virtual screening approach are:

a. knowledge of compounds,

b. knowledge of receptor structure and receptor-ligand interactions, and

c. standard knowledge about drugs and drugs characteristics.

Combinatorial chemistry is a synthetic tool that enables chemists to rapidly generate thousands of lead compound derivatives for testing. As shown in Fig. (2.3) above, a scaf- fold is employed that contains a portion of the ligand that remains constant. Sub site groups (shown in red, green, and blue) are potential sites for derivatization. These sub sites are then reacted with combinatorial libraries to generate a multitude of derivative structures, each with different substituent groups. One can see how a vast number of compounds can be generated as a result of the combinatorial process. By carefully selecting libraries based upon the study of the active site, we can target the derivatization process towards optimizing ligand receptor interactions. Combinatorial chemistry is a very powerful technique that chemists can employ to aid in the refinement of the lead compound (Puzyn *et al*., 2010).

To move from a lead to a drug candidate, the lead is used as a base structure for the synthesis of sometimes hundreds of derivatives in a process known as "Lead Optimization". In this step, medicinal chemists make many changes to the original lead molecule and determine effect of changes on activity. The most promising leads are prioritized based on a set of criteria. Starting with intuitive structural modification to the development of structure-activity relationship SAR and QSAR one can gain tremendous information. Chemical diversity integrates its wealth of organic, medicinal and

computational chemistry and its technological capabilities to offer a platform for the most efficient lead optimization solutions. For new drug discovery, a synthetic chemical route is usually preferable in order to conserve natural resources and to control purity and yield (Puzyn *et al*., 2010). Once a number of lead compounds have been found, computational and laboratory techniques have been very successful in refining the molecular structures to give a higher drug activity and fewer side effects. This is done both in the laboratory and computationally by examining the molecular structures to determine which aspects are more responsible for both the drug activity and the side effects (Bruce, 2010). One of the techniques most often used to refine drugs is QSAR.

Many important receptors are membrane-bound proteins, which are notoriously difficult to crystallize. In such cases, a lead compound or active ligand must be found, and then the structure of the ligand guides the drug design process. Schematic representation of ligand library formation is shown in Fig. (2.3).QSAR allows the use of statistics to lower the trial and error factor in drug design (Wang *et al*., 2006). The computers help analyze variables in the numerous mathematical equations. The underlying premise of QSAR is that there is a relationship between the biological or pharmacological activity of a compound and its structural, physical and chemical properties. These physicochemical descriptors, which include parameters to account for hydrophobicity, topology, electronic properties and steric effects are determined empirically or, more recently, by computational methods. Activity is a function of structure and an equation can be determined relating activity to parameters which can be determined. An important advantage of QSAR is that it models the in vivo situation, since it is based on experimental biological activity data. This approach allows important structural requirements for activity to be identified and hence narrows the search for the optimum molecule. Since it is a quantitative measure, how much each aspect alters activity can

also be determined. QSAR equations, once determined, also allow new compound's activity to be predicted based on structural data and hence saves the time in synthesizing molecules unlikely to have good activity (Singh *et al*., 2009).

In other words, QSAR is a mathematical models relating measured biological activity of series of structurally related chemical compounds/pharmacological agents to the variation in their chemical structure. QSAR models are nowadays regarded as a scientifically credible tool for predicting and classifying biological activities of untested chemicals (Votano *et al*., 2004) QSAR has become inexorably embedded as an essential tool in the pharmaceutical industry, from lead discovery, optimization to lead development and computer aided drug designing. A growing trend is to use QSAR early in the drug discovery process as a screening and enrichment tool to estimate from further development those chemicals lacking drug like properties or those chemicals predicted to elicit a toxic response.

2.7 Quantitative Structure Activity Relationships Study in Computer Aided Drug Design

Quantitative structure activity relationship (QSAR) has served as a valuable predictive tool in the design of pharmaceuticals and agrochemicals. From decades to recent research, QSAR methods have been applied in the development of relationship between properties of chemical substances and their biological activities to obtain a reliable statistical model for prediction of the activities of new chemical entities. Classical QSAR studies include ligands with their binding sites (Fig 2.3 to 2.5), inhibition constants, rate constants, and other biological end points, in addition molecular to properties such as lipophilicity, polarizability, electronic, and steric properties or with certain structural features. 3D-QSAR has emerged as a natural extension to the classical Hansch and Free–Wilson approaches, which exploit the three-dimensional properties of

the ligands to predict their biological activities using robust chemometric techniques (Patel *et al*., 2014).

The fundamental assumption of QSAR is that variations in the biological activity of a series of chemicals that target a common mechanism of action are correlated with variations in their structural, physical and chemical properties. Since presumably these structurally related properties of a chemical can be determined by experimental or computational mean much more efficiently that its biological activity using in-vivo or in-vitro approaches, a statistically validated QSAR model is capable of predicting the biological activity of a new chemical within the same series in lieu of the time consuming and lab our-intensive processes of chemical synthesis and biological evaluation. Applied judiciously, QSAR can save substantial amount of time, money and human resources (Kar and Roy, 2011; Matthews and Contrera, 2007).

2.7.1 Steps involve in building QSAR models

QSAR Modelling process consists of 5 main steps as shown in Fig 2.6

a. Begins with the selection of molecules to be used

b. Selection of descriptors; numerical representer of molecular features

(e.g. number of carbon)

c. Original descriptor pool must be reduced in size

d. Model building

e. The reliability of the model should be tested

2.8 Previous QSAR Works on Anti-*Candida albicans* and *Epidermophyton floccosum* Molecules

In recent years, substantial progress has been made in the application of *in silico* computational methods to predict *Candida albicans* and *Epidermophyton floccosum* inhibitory activities of some other chemicals.

2.8.1 2D-QSAR study on 1-acetyl-3-Aryl-5- (4-Methoxyphenyl) pyrazole analogues as an antifungal agents

A series of 1-acetyl-3, 5-diarylpyrazole were screened for their antifungal activity against fungi *Candida albicans.* These compounds have showed moderate and very good antifungal activity. The Quantitative Structure Activity-Relationships (QSAR) study on the pyrazole series was made using lipophilic, electronic and steric parameters. Several statistical expressions were developed and best models were validated. The studies confirm that the antifungal activity is dependent on selected lipophilic and electronic parameters (Zheng and Tropsha, 2000).

2.8.2 QSAR and docking studies of Coumarin derivatives as potent HIV-1 integrase inhibitors:

Human immunodeficiency virus integrase (HIV-1IN) is an emerging and potential drug target for anti-HIV therapy. It is an enzyme essential for 30 processing and integration step in the life cycle of HIV. In the present study a series of coumarin derivatives (containing 26 compounds) as HIV-1IN inhibitors was subjected to quantitative structure–activity relationship (QSAR) analysis. For building the regression models two different variable selection approaches namely, genetic function approximation (GFA) and sequential multiple linear regression (SQ-MLR) were used and compared to predict the HIV-1IN inhibition activity. Based on prediction, the best validation model for 30 processing inhibition activity with squared correlation coefficient (r^2) = 0.8965, cross validated correlation coefficient (Q^2) = 0.8307 and external prediction ability pred_r^2= 0.5400 showed that Henry's law Constant (HLC), Partition Coefficient (PC) and Dipole moment-Z com- ponent (D^3) were the positive contributors, whereas for integration inhibition activity, parameters r^2 = 0.8904,Q^2 = 0.8174 and pred_r^2 = 0.7159 showed HLC, Logarithm of Partition Coefficient (LogP) and Dipole moment-Y component (D^2)

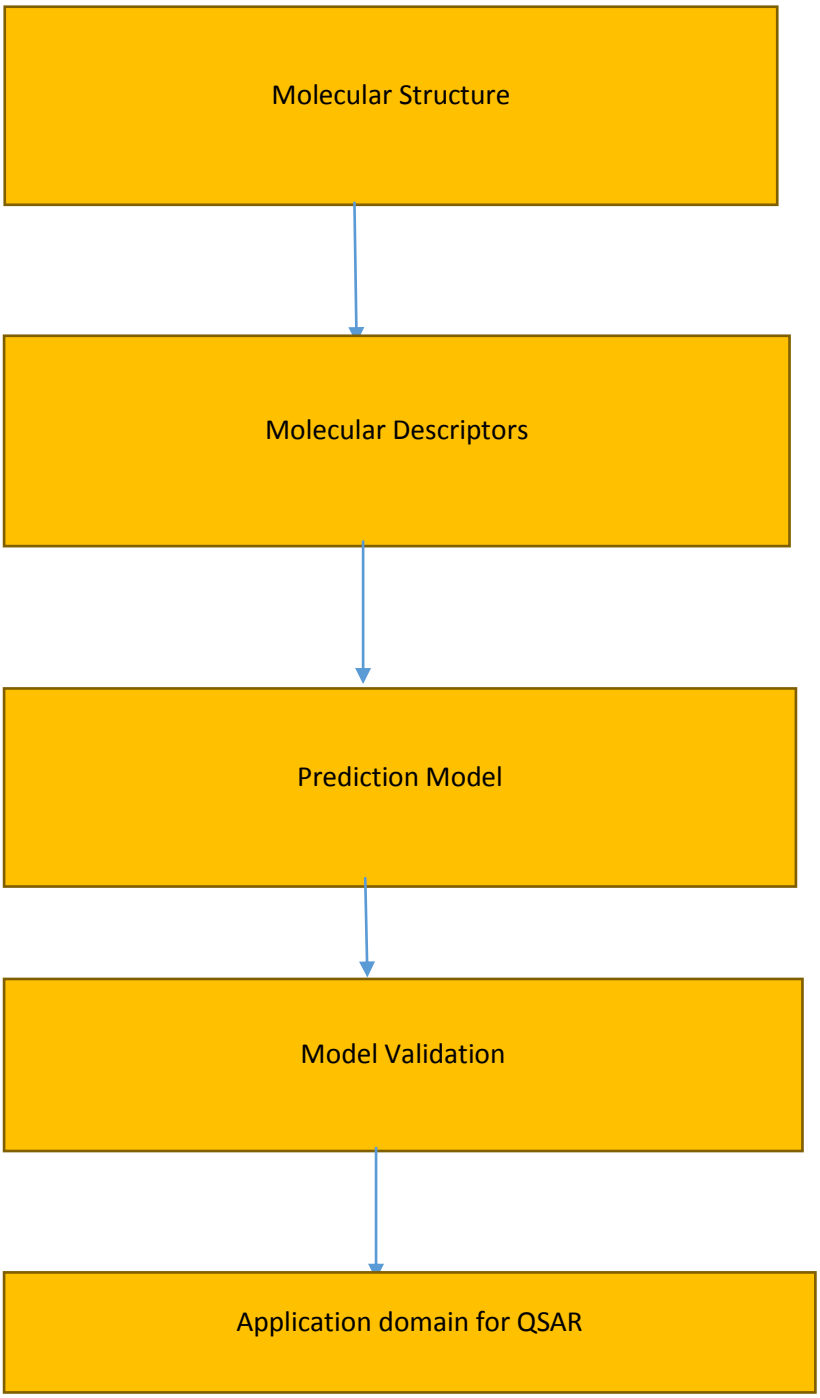


Fig 2.6: Steps involve in building QSAR models

contributed positively to the activity. The binding mode pattern of the compounds to the binding site of integrase enzyme was confirmed by docking studies. The results of the present study may be useful for designing more potent HIV-1IN inhibitors.(Ma *et al*., 2011)**.**

2.8.3 2D and 3D QSAR studies of flavonoids , biflavones and chalcones: anti-viral antibacterial , antifungal and anti-mycobacterial activities

Flavonoids comprise a group of naturally occurring compounds that are known by their biological properties. Antifungal activities were analyzed. The results were discussed in terms of the bioactivity tested, and the structure-activity relationship (SAR), bidimensional (2D)- and three-dimensional (3D)-QSAR studies carried out on flavones, flavanones, chalcones, and some biflavonoids(Mercader and Pomilio, 2012).

2.8.4 Antifungal and antibacterial activities of imidazolylpyrimidines derivatives and their QSAR studies

Quantitative structure-activity relationship (QSAR) models of newly synthesized substituted diphenyl imidazoly l-pyrimidines were established by using the molecular descriptors surface tension (ST), molar volume (MV), Xindex, molar refractivity (MR), total energy (TE), lowest unoccupy mplecular orbital (LUMO), Log P, volume (V), AR, average eccentricity (AECC). The logarithm of zone of inhibition of micro-organism i.e. *Candida albicans* strain are used as key properties to evaluate the QSAR models. The Predictive ability and accuracy of the model is determined by a cross validation method(Rathod, 2011).

2.8.5 QSAR studies, design, synthesis and antimicrobial evaluation of azole derivatives

QSAR analysis of a set of previously synthesized azole derivatives tested for growth inhibitory activity against *Candida albicans* was performed by using Associative Neural Network. To overcome the problem of overfitting due to descriptor selection, 5-fold

cross-validation with variable selection in each step of the analysis was used. The predictive ability of the models was tested through leave-one-out cross-validation, giving a Q^2 = 0.77 - 0.79 for regression models. Predictions for the external evaluation sets obtained accuracies in the range of 0.70 - 0.80 for regressions. Biological testing of compounds was performed by disco-diffusion method on solid medium culture versus strain *C. albicans* ATCC 10231 M885. Most of the compounds demonstrated high antifungal activity. Five synthesized compounds also showed activity against clinical isolate strain of *C. albicans* received from a biological material and resistant to fluconazole(Kovalishyn *et al*., 2014).

2.8.6 QSAR studies of some pyrazolones as antimicrobial agent.

Quantitative Structure Activity Relationship (QSAR) studies were performed on selected pyrazolone compounds as antimicrobial agent. The compounds has been established on the basis of analytical methods and advanced spectroscopic techniques. Three compounds *viz*. 4-Amino antipyrine thiosemicarbazone (C-8), 1-(2-Chloro-5-sulfophenyl)-3-methyl-5-pyrazolone (C-9) and 1-(4-Sulfoamido phenyl)-3-methyl-5-pyrazolone (C-10) recorded significant antifungal activity against *Macrophomina phaseolina*(Arora, 2013).

2.8.7 3D-QSAR of protein tyrosine phosphatase 1b inhibitors by genetic function approximation,

Three dimensional quantitative structure activity relationship (3D-QSAR) study performed by genetic function approximation (GFA) technique on a series of benzofuran/benzothiophene biphenyls.The QSAR models were generated using 92 compounds, and the predictive ability of the resulting model was evaluated against a test set of 26 compounds. The internal and external consistency of the final QSAR model was 0.694 and 0.672. Analyses of results from the present QSAR study indicate

that electronic, structural and shape descriptors govern the inhibitory activity (Vadlamudi and Kulkarni, 2003).

2.8.8 Chemistry by computer: an overview of the applications of computers in chemistry.

The molecular structures of 83 diverse organic compounds are correlated by a quantitative structure-activity relation (QSAR) to their minimum inhibitor concentration (MIC expressed as log(1/MIC)), involving six (6) descriptors with R^2 = 0.788, F = 47.140 . S^2 = 0.130 . A novel QSAR development technique is utilized combining advantages of the two frequency applied methods. The topological , electronic , geometrical and hybrid type descriptors for the compounds were calculated by CODESSA PRO software.

2.8.9 Pharmacophore modeling and atom-based 3D-QSAR studies of antifungal benzofurans.

A statistically significant 3D-QSAR model for training set of 24 compounds was obtained using pharmacophore hypothesis with correlation coefficient (r^2 = 0.916) and high Fisher ratio (F =113.9). Also, the predictive power of generated model for test set of 5 compounds was found to be significant which was confirmed by the high value of cross validated correlation coefficient (q^2 = 0.804) and Pearson-R (0.917) (Pathare *et al*., 2015).

2.8.10 QSAR modeling of antifungal activity of some heterocyclic compounds

QSAR analysis of a set of benzoxazoles, benzimidazoles, oxazolo pyridines and benzothiazoles, showing growth inhibitory activity against *Candida albicans*, was performed using a multiple regression procedure. Topological indices (TIs) and principal component analysis (PCA) on TIs were used in modeling antifungal activity. Selection of TIs relevant to developing QSAR models was made using the largest PC

factor loading scores. Correlation coefficient 0.97 obtained in the validation procedure indicated the excellent quality of the derived QSAR models(Ursu *et al*., 2006).

2.8.11 Medicinal chemistry QSAR analysis of n-myristoyltransferase inhibitors : antifungal activity of benzofurans

Benzofurans are a class of antifungal agents reported to act by selective inhibition of the N-myristoyltransferase (Nmt) enzyme in fungal cells. A threedimensional quantitative structure–activity relationship (3D-QSAR) was performed on a series of 29 molecules to find correlation between various physicochemical descriptors and Nmt inhibition. QSAR equations were evaluated using a training set of 24 molecules and an external test set of 5 molecules. The statistical quality of the QSAR models was evaluated using the parameters r^2, r^2_{cv}, and the r^2_{pred} measure. The results obtained indicated that the enzyme inhibitory activity of benzofuran analogues strongly depends on structural factors as expressed by hydrogen-bond acceptors and spatial factors as expressed by the principle moment of inertia in the X (PMI_X) and Y directions (PMI_Y) (Puranik and Kulkarni, 2009)

2.8.12 Synthesis, anti-fungal activity evaluation and QSAR studies on podophyllotoxin derivatives

The antifungal and cytotoxic activites of podophyllotoxin and seven C-4 substituted podophyllotoxin ester derivatives, viz: trans-cinnamyl, cis-cinnamyl, o-methoxy cinnamyl, dimethyl acrylyl, p-methoxy phenyl acetyl, 3,4-dimethoxy phenyl acetyl and 2,5-dimethoxy phenyl acetyl esters were evaluated on four fungi, viz: *Macrophomina phaseolina, Fusarium oxysporum, Myrrotheciumverrucarria and Asperigillus candidus,* The podophyllotoxin derivatives were synthesised and their structures were elucidated. Quantitative structure activity relationships were developed between the activity of these compounds against the four fungi and molecular descriptors. The linear regression models developed had one to two descriptors. For all the cases the r^2 was in the range of

0.73 to 0.96, indicating good data fit and q^2 was in the range of 0.60 to 0.68, indicating that the predictive capabilities of the models were acceptable. Solvent accessible surface area (namely the partial positive solvent-accessible surface area), *A*log *P*, highest occupied molecular orbital and conformational energy were identified as important descriptors(Kumar*et al*., 2007).

2.8.13 2D QSAR studies on a series of bifonazole derivatives with antifungal activity

Candida albicans (CA) has been identified as the major opportunistic pathogen in immunosuppressed patients. Most of currently available drugs are either highly toxic or becoming ineffective against resistant strains. An approach to overcome this burden relies on azole derivatives with increased potency and selectivity. Aiming at shedding some light on structural and chemical features that are important for the antifungal activity of azole derivatives, classical 2D QSAR and hologram QSAR (HQSAR) studies were performed for a diverse set of 52 bifonazole derivatives with antifungal activity. Topological descriptors, employed in Classical QSAR studies, resulted in models with low correlation (r^2 = 0.38, q^2 = 0.27) and lack of predictive power (r^2_{pred} = −0.6). On the other hand molecular holograms afforded HQSAR models with good correlation coefficients (r^2= 0.92, q^2 = 0.65) and good predictive ability (r^2_{pred} = 0.79) (Mota and Castilho, 2009)

2.8.14 QSAR study on 2, 5-disubstituted 1,3,4-oxadiazoles as antifungal agents

QSAR study were performed on novel 2,5 disubstituted 1,3,4-oxadiazoles analogues. Stepwise multiple linear regression analysis was performed to derive QSAR models for better activity and lesser side effects. The best QSAR model was selected, having correlation coefficient (r) = 0.91 and cross validated squared correlation coefficient (q2) = 0.70. The predictive ability of the selected model was also confirmed by leave one out cross validation. The QSAR model indicate that the thermodynamic descriptors (Molar

refractivity) and principal moment inertia, play an important role for antifungal activities. The data obtained from this present quantitative structure activity relationship study may be useful in the design of more potent substituted oxadiazole derivatives as antifungal agents(Nawale and Ashtekar, 2014).

In most of the recent QSAR works on anti-*Candida albicans* and *Epidermophyton floccosum* molecules as reviewed above,most of the training set used in the model development was small size. Also the external validation of QSAR model reflects the predictive capability of the model on new data set. This validation was technique obviously lacking in recent QSAR studies on *Candida albicans* and *Epidermophyton floccosum* activities of compounds as reviewed. Thus the models lack sound quality assurance.

CHAPTER THREE

3.0 MATERIALS AND METHODS

3.1 Materials

The materials used in this studies include; HP pavilion notebook PC intel(R) core^{TM} i5 – 3210M CPU @ 5.00 GHz processor, 6GB ram size on Microsoft windows7 Ultimate operating system, Spartan 14 V.1.1.0 , Chm 3D Pro 12.0.1V, PaDEL-descriptor tool kit and Microsoft office Excel 2013 version and Material Studio V 7.0.

3.2 Methods

The strategy used in the QSAR methodology includes the following steps: (i) collection of Chemical data; (ii) optimization of the geometry of the molecular structures; (iii) descriptors calculation; (iv) normalization of data; (v) division of data into training and test set;(vi) learning process (QSAR models development); (vii) evaluation of the validity and predictability of the developed QSAR models (Cronin, 2005) .

3.2.1 Collection of chemical data

A literature search was made to collect compounds with biological activities against *Candida albicans*and *Epidermophyton floccosun*(Bag *et al*., 2009; Batovska *et al*., 2007; Jayakumar *et al*., 2013; Motta and Almeida, 2011; Pinheiro *et al*., 2004). This was converted to pMIC which indicated the biological activity of compounds experimentally determined necessary for the inhibition of the two micro-organisms. The –log MIC (molar) scale refers pMIC.Table 3.1 shows both the training and test sets. The structures, values and figures for each member of the series are presented in Table 3.2 (and figure 3.1), Table 3.3, Table 3.4a(and Figure 3.2a), Table 3.4b (and Figure

3.3b),Table 3.5a (and Figure 3.4a), Table 3.5b (and Figure 3.5b) and Table 3.5c (and Figure 3.6c).

Table 3.1 Training and test sets

Organism	Compounds	Set	pMIC	
Candida albicans	Coumarin derivatives		Training set	Test set
		1	15	5
		2	21	7
		3	32	8
Epidermophyton floccosum	Neolignans derivatives	4	13	5

Fig. 3.1: Parent structure of training and test set of Chalcone derivatives

Table 3.2: Set 1 experimental *pMIC* values of the molecules against *Candida albicans*

Compound	Ring A	Ring B	Substituent Z	pMIC
1	4-SCH_3	4-F	-	4.53
2	4-SCH_3	4-Cl	-	4.16
3	4-SCH_3	4-Br	-	3.52
4	4-SCH_3	2,4-Cl	-	3.21
5	4-SCH_3	4-NO_2	-	3.48
6	4-SCH_3	4-OCH_3	-	4.15
7	4-SCH_3	H	-	3.80
8	4-SCH_3	4-OH	-	3.83
9	4-SCH_3	2-OH	-	4.13
10	4-SCH_3	3-OH	-	3.83
11	4-SCH_3	4-phenyl	-	3.12
12	2,3-OCH_3	4-OCH_3	-	4.47
13	4-SCH_3	-	H	4.72
14	4-SCH_3	-	Br	3.83
15	3,4-OCH_3	-	H	4.14
16	3,4-OCH_3	-	Br	3.55
17	4-phenyl	-	H	3.06
18	4-phenyl	-	Br	3.17
19	4-OCH_3	-	H	4.09
20	4-OCH_3	-	Br	3.51

Table 3.3: Set 2 experimental *pMIC* values of the molecules against *Candida albicans*

Compound	Structures	MIC (μg/ml)	pMIC
1		31.25	1.495
2		250	2.398
3		62.5	1.796
4		62.5	1.796

Table 3.3: Set 2 experimental *pMIC* values of the molecules against *Candida albicans* (continuation)

5		62.5	1.796
6		500	2.699
7		62.5	1.796
8		125	2.097

Table 3.3: Set 2 experimental *pMIC* values of the molecules against *Candida albicans* (continuation)

9		62.5	1.796
10		62.5	1.796
11		500	2.699
12		500	2.699

Table 3.3: Set 2 experimental *pMIC* values of the molecules against *Candida albicans* (continuation)

13		250	2.398
14		125	2.097
15		500	2.699
16		500	2.699

Table 3.3: Set 2 experimental *pMIC* values of the molecules against *Candida albicans* (continuation)

17		500	2.699
18		500	2.699
19		1000	3.00
20		1000	3.00

Table 3.3: Set 2 experimental *pMIC* values of the molecules against *Candida albicans* (continuation)

21		1000	3.000
22		500	2.699
23		250	2.398

Table 3.3: Set 2 experimental *pMIC* values of the molecules against *Candida albicans* (continuation)

24		250	2.398
25		200	2.301

Table 3.3: Set 2 experimental *pMIC* values of the molecules against *Candida albicans* (continuation)

26		500	2.699
27		1000	3.000
28		1000	3.000

Fig. 3.2a: Parent structure of unsaturated Ketone (a)

Table 3.4a: Set 3a experimental *pMIC* values of the molecules against *Candida albicans*

	X	Y	Z	A	B	Activity (ATCC 10231) pMIC		Activity (NCIM 3446) pMIC
F							F	
1	F	H	H	SCH_3	H	1.739	21	1.535
2	Cl	H	H	SCH_3	H	1.464	22	1.163
3	Br	H	H	SCH_3	H	0.923	23	0.525
4	Cl	H	Cl	SCH_3	H	0.813	24	0.211
5	NO_2	H	H	SCH_3	H	0.780	25	0.479
6	OCH_3	H	H	SCH_3	H	1.457	26	1.156
7	H	H	H	SCH_3	H	1.083	27	0.782
8	OH	H	H	SCH_3	H	1.134	28	0.833
9	H	H	OH	SCH_3	H	1.134	29	1.134
10	H	OH	H	SCH_3	H	1.134	30	0.833
11	Phenyl	H	H	SCH_3	H	0.522	31	0.124
12	OCH_3	-	-	-	OCH_3	1.199	32	1.500

Fig. 3.3b: Parent structure of unsaturated Ketone

Table 3.4b: Set 3b experimental *pMIC* values of the molecules against *Candida albicans*

X	Y	Activity (ATCC 10231) pMIC		Activity(NCIM 3446) pMIC
13	H	1.739	33	1.739
14	Br	0.946	34	0.849
15	H	1.483	35	1.182
16	Br	0.885	36	0.584
17	H	0.483	37	0.086
18	Br	0.584	38	0.186
19	H	1.111	39	1.111
20	Br	0.528	40	0.829

(a)

Fig. 3.4a: Parent structure of Neolignans derivative

Table 3.5a: Set 4a experimental pMIC values of the molecules against *Epidermophyton floccosum*

Compound	R_1	R_2	R_3	R_4	MIC	pMIC
1	OCH_3	OCH_3	H	Allyl	0.103	0.99
2	OCH_3	OCH_3	H	Allyl	0.154	0.81
3	OCH_3	OCH_3	OCH_3	Allyl	1.120	0.92
4	OCH_3	OCH_3	OCH_3	Allyl	0.143	0.84
5	OCH_2O	-----	H	Allyl	0.013	1.89
6	OCH_2O	-----	H	Allyl	0.040	1.40

(b)

Fig. 3.4b: Parent structure of Neolignans derivative

Table 3.5b: Set 4b experimental pMIC values of the molecules against *Epidermophyton floccosum*

Compound	R_1	R_2	R_3	R_4	MIC	pMIC
7	OCH_3	OCH_3	H	Allyl	0.042	1.38
8	OCH_3	OCH_3	H	Allyl	0.670	0.17
9	OCH_3	OCH_3	H	t-propenyl	0.195	0.71
10	OCH_3	OCH_3	H	t-propenyl	0.279	0.55
11	OCH_3	OCH_3	OCH_3	t-propenyl	0.064	1.19
12	OCH_3	OCH_3	OCH_3	t-propenyl	0.129	0.89

(c)

Fig. 3.4c: Parent structure of Neolignans derivative

Table 3.5c: Set 4c experimental pMIC values of the molecules against *Epidermophyton floccosum*

Compound	R_1	R_2	R_3	R_4	R_5	MIC	pMIC
13	OCH_3	OCH_3	H	Allyl	OCH_3	0.647	0.19
14	OCH_3	OCH_3	OCH_3	Allyl	OCH_3	0.600	0.22
15	OCH_2O	-	H	Allyl	OCH_3	0.675	2.99
16	OCH_3	OCH_3	H	Allyl	H	0.701	0.15
17	OCH_3	OCH_3	H	t-propenyl	H	0.751	0.12
18	OCH_3	OCH_3	OCH_3	t-propenyl	H	0.647	0.19

3.2.2 Optimization of the geometry of the molecular structure

The chemical structure of each compound in the data set was drawn with Chemdraw ultra V12.0, named and save as *cdx file. The first step in performing any QSAR studies is to specify the structure of the molecules**.** The correct 3D geometry of the compounds is very important to find a possible relationship between molecular descriptors of the compounds and their biological activity(Guido *et al*., 2008) as well as for subsequent calculation of 3D molecular descriptors. Chemdraw ultra V12.0 software package contains tools to sketch organic molecules in 3D using these tools, the 3D structures of all the compounds were build according to the following procedure. The procedures for building compounds of Fig. 3.1 is as follows : The Chemdraw software package was launched by double clicking the Chemdraw ultra icon on the laptop, the graphical interface appeared (Plate 3.1). By clicking on view on the main bar → show main toolbar, the tools bar appeared on the Chemdraw ultra interface (Plate 3.2). From the tools menu, a benzene ring was selected by clicking on ( ), the screen was then clicked to place the hexane ring. Next, solid bond ( ) was selected from the tools menu and clicked on C_6. These operations resulted in the structure . Multiple bond (=) was selected on the tools menu, the cursor was positioned on the methyl group ( ), again solid bond ( ) was selected from the tools menu and attached to the ethene group ( ). Multiple bond (=) was selected on the tools menu, the cursor was positioned on the methyl group, the text (A) was clicked in the tools menu, the cursor was placed above double bond and type “O” for Oxygen ( ). Again solid bond ( ) was selected from the tools menu and attached to the aldehyde group, finally, a hexane ring was selected from the tools menu by clicking on ( ), and attached to the methyl group and

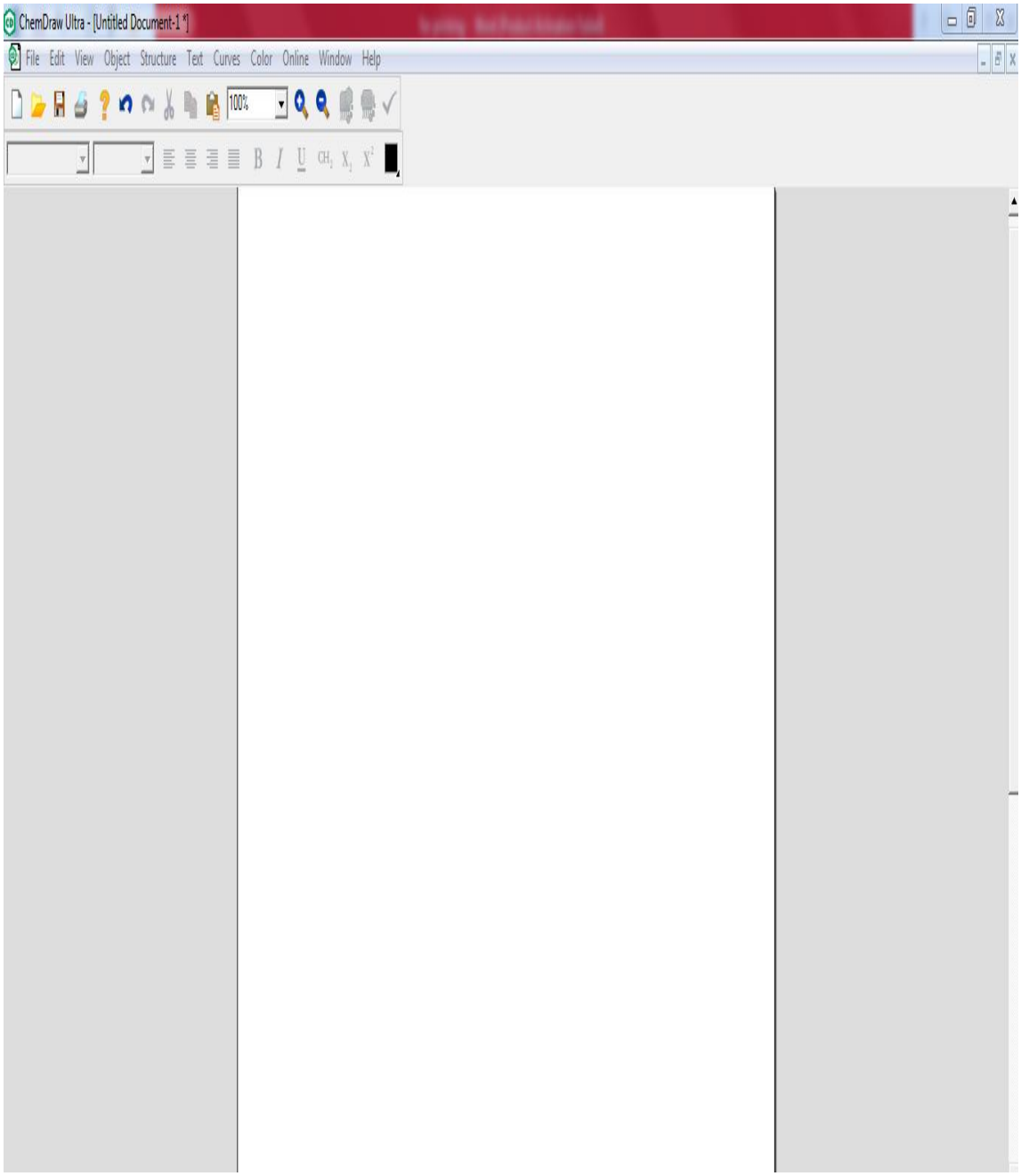


Plate 3.1 : Chemdraw graphical interface

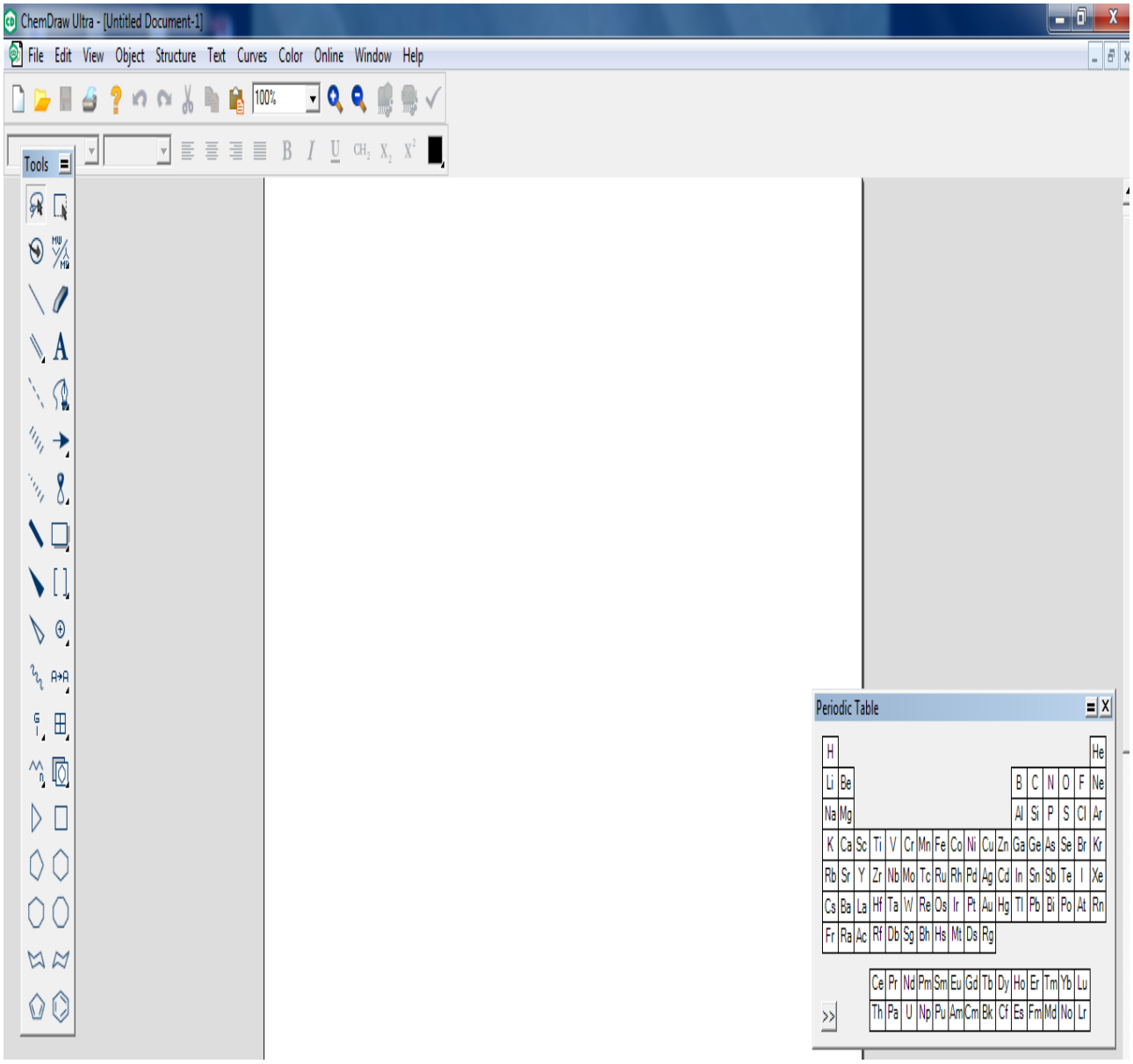


Plate 3.2 : Main toolbar

the operation resulted to the structure ( ). All other compounds were drawn as described.

Calculations were performed using the molecular modeling program SPARTAN '14 V1.1.0 on HP Pavilion g6 CORE^{TM}i5,5.0GHZ processor,6GB RAM size Operated system computer system Microsoft windows 7. The molecular geometry of all the structures were first optimized with molecular mechanics (MMFF) and semi-empirical (PM3) procedures included in Spartan '14 VI.1.0 software and the resulting geometries were further refined by means of Density functional theory (DFT) with Becke's three-parameter hybrid functional (Becke and Becke, 1993; Yanai *et al*., 2004) using LYP correlation functional. The standard Pople's 6-311G* basis set was used. The lowest energy structure was used for each molecules to calculate their physicochemical properties.

(a) Optimization with Spartan 14 V.1.1.0

The various procedures deployed in molecular optimization using Spartan 14 V.1.1.0 are highlighted thus: Spartan 14 software was lunched by double clicking on its icon on the desktop of the laptop specified in section 3.1 above creating its interface shown in Plate 3.3. By clicking on the ***file menu→Open***button on the menu bar, a file browser shown in Plate 3.4 popped up in the ***file name*** box of the window, the name of the molecule saved as *cdx* file as described in section 3.2.2 was indicated. ***All files*** option was selected from the *cdx* named file to allow the spartan software recognize the *cdx* file. By clicking ***Open*** on the window, the molecule was imported to the Spartan's interface. By clicking  option (Plate 3.3) on the menu bar for the interface, the structure of the molecule was completely imported unto the interface. The molecule was

represented using the ***ball and spoke***option in the *model* menu. The energy of the imported molecule

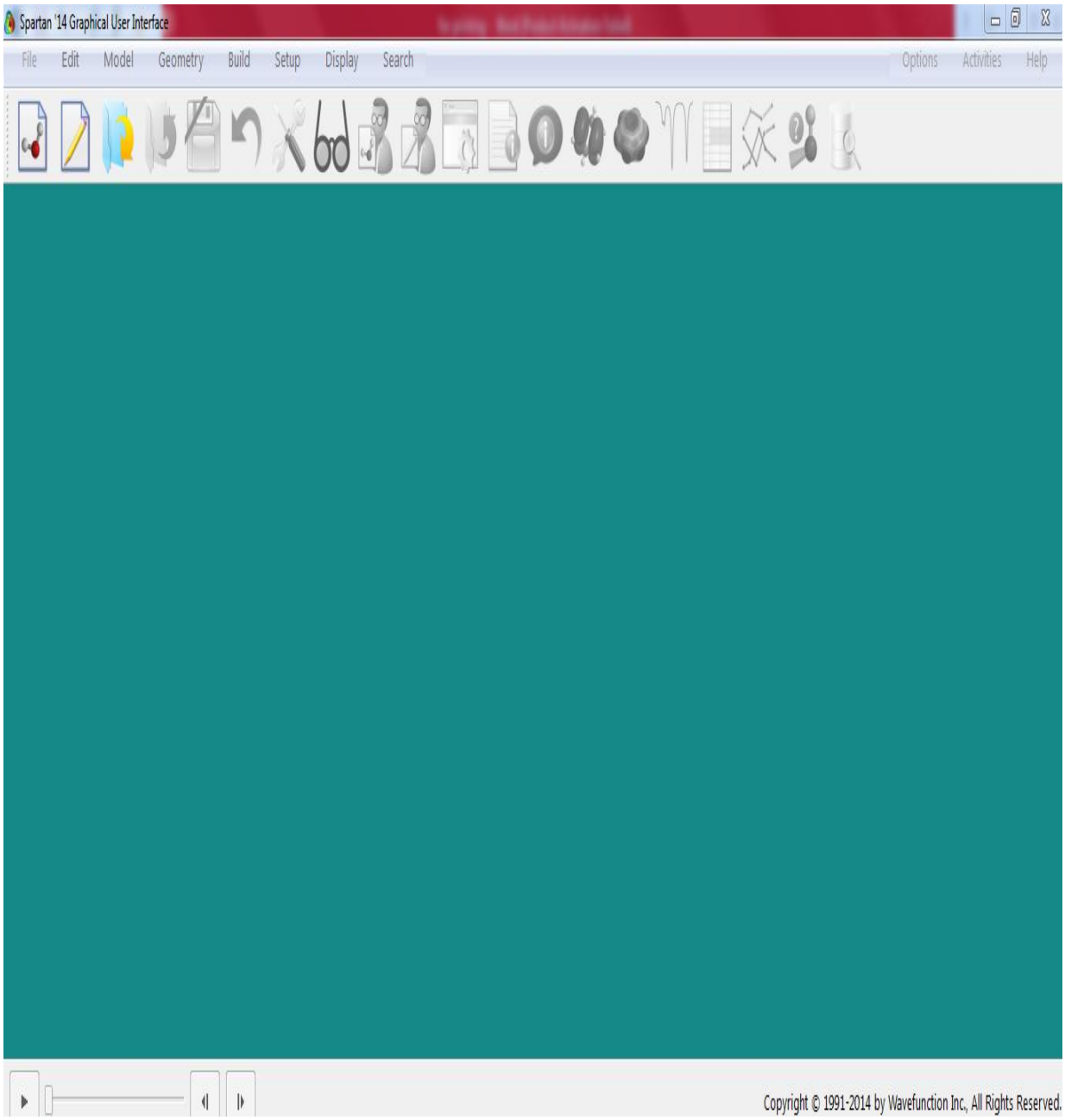


Plate 3.3: Spartan’14 graphic interface

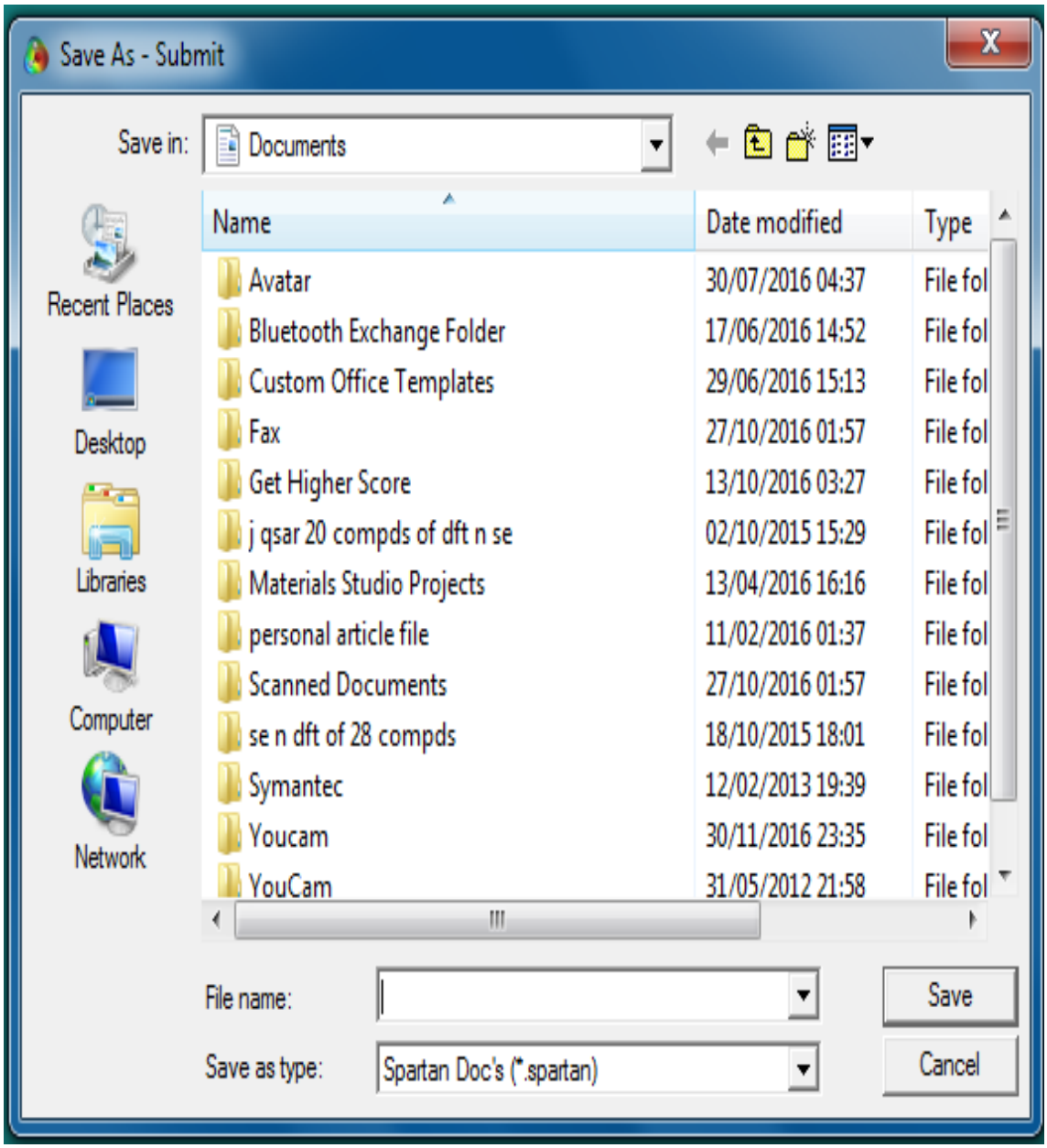


Plate 3.4: File browser

was minimized by selecting***build***(Plate 3.3) from the menu bar, ***minimize***was*clicked* as shown in Plate 3.5, this was done for the molecule to assume it form in test tube. The minimum energy geometry of the molecule was calculated by selecting ***set up***from the menu bar (Plate 3.3), ***calculations***was ***clicked***on the menu bar causing the appearance of a dialog box shown in Plate 3.6. On the dialog box, ***equilibrium geometry*** was selected from the leftmost checkbox to the right of ***calculate*** menu. This specifies optimization

of equilibrium geometry. Using the ***state*** checkbox in the dialog box,***ground*** state was selected. Through the leftmost checkbox on the second row, right of ***calculate*** menu, the type of calculation (molecular mechanics, semi-empirical, Hartree fock or DFT) was specified and that is ***DFT*** was clicked. Directly below the***state*** check box on the second row of ***calculate***menu is the check box for indicating the basis set is 6-31G* (B3LYP for DFT), next to it is the box for medium and ***vacuum*** medium was checked. The check box directly right of the ***start from*** menu was used to select the starting geometry of the molecule, ***mmff geometry*** was selected in all calculations. The ***compute***menu was used to specify the properties of the molecule to be computed; the check boxes for **QSAR** and **IR** were marked in this section. The ***print*** menu was used to mark the checkboxes for **orbital energies**, **thermodynamics**, **charges and bond order**. There were instances where the software runs out cycle during molecular optimization. This problem was remedied by increasing the number of cycles using the ***option*** menu by typing in the corresponding check box; OPTCYCLES = 1000000……***Calculations*** dialogue was exited by clicking on ***submit***at bottom right of the dialog causing the appearance of a file browser (Plate 3.4). In the ***file name*** check box of the browser, the name of the molecule saved as ***cdx*** file as described in section 3.2.2 abovewas typed in and***save*** option clicked.A notification that the calculation has been submitted appeared on the screen, clicking on ***Ok*** removed thescreen.Through the***file***menu on the Spartan interface

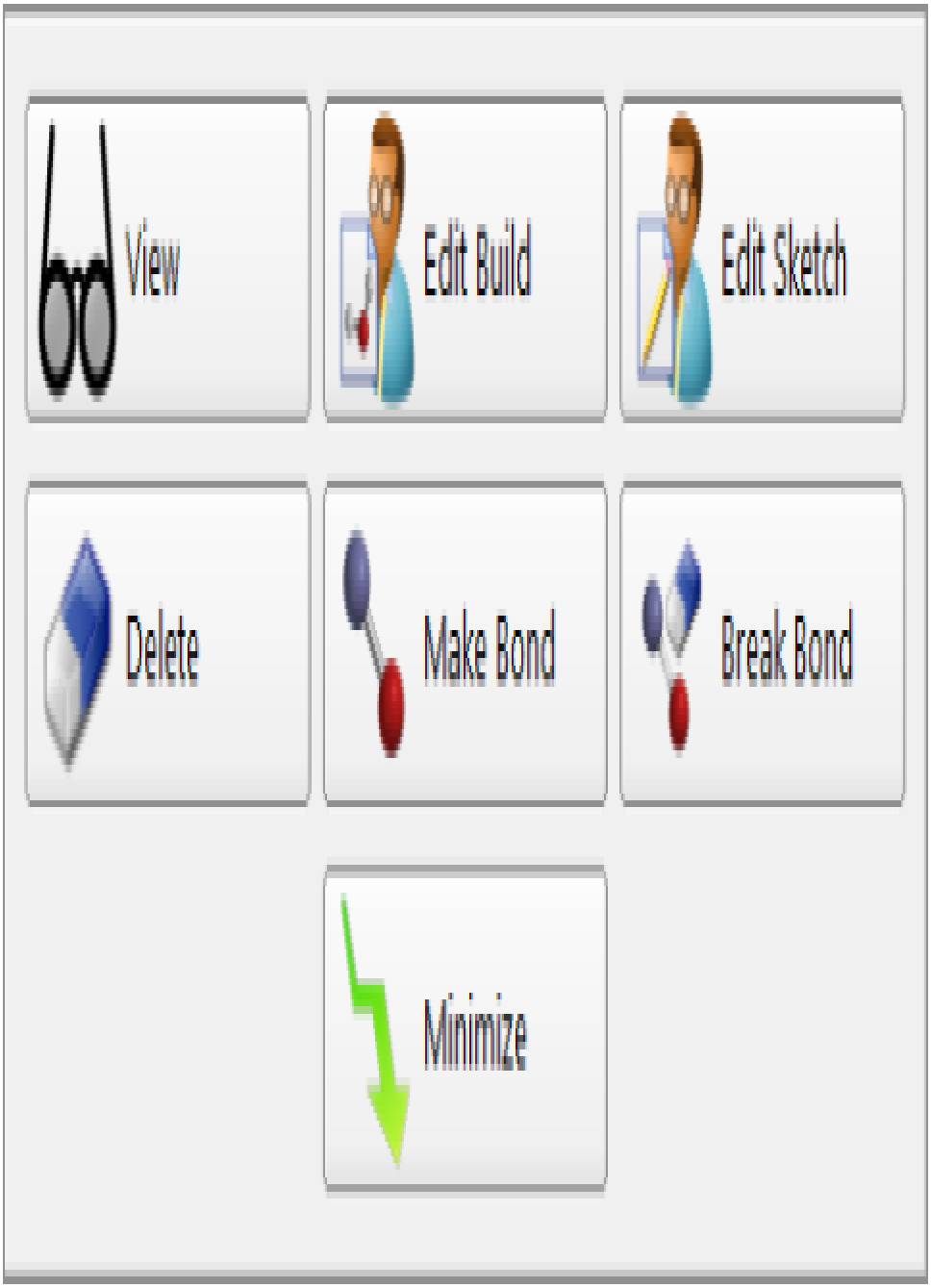


Plate 3.5 : interface for minimize energy

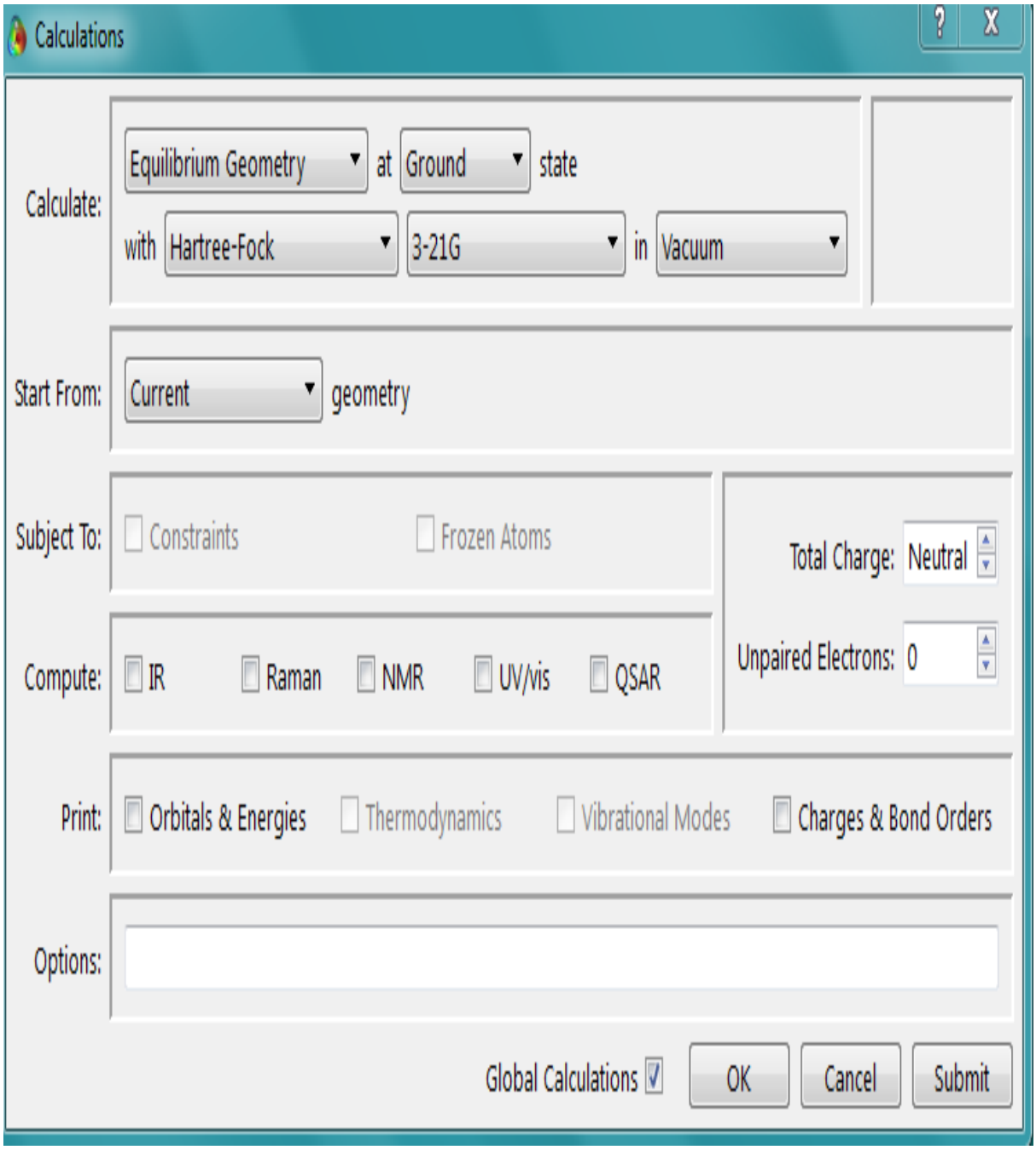


Plate 3.6: Spartan' 14 *Calculations* dialog box.

the optimized molecule was saved as ***sdf**** file in a folder. This procedure was applied to all other molecules in each of the data sets.

3.2.3 Descriptor calculation

In the last decades, several scientific researches have been focussed on studying how to catch and convert by a theoretical pathway the information encoded in the molecular structure into one or more numbers called molecular descriptors used to establish quantitative relationships between structures and properties, biological activities and other experimental properties(Braumann, 1986).The molecular descriptors used in the QSAR modelling were calculated using Padel descriptor tool kit, Spartan 14 software and ChmBio3D Pro 12.0.1V software. Table 3.6 shows the symbol and definition of descriptors in the models.

(a) Computing molecular descriptors using Padel descriptor tool kit.

The Padel software was launched by double clicking the jar file in the Padel folder to create its interface shown in Plate 3.7. On the interface, a directory containing the molecular structural files was selected from a prop up dialog box shown in Plate 3.8 by clicking the ***molecules directory/file*** option on the window. By clicking on the ***descriptor output file*** menu, the folder where the result of the calculations were saved as *CSV* file type was created from a file browser (Plate 3.7) that propped up. By clicking the

Table 3.6: Molecular descriptors symbols and definition in the models

S/NO	Symbol	Definition
1	MW	Molecular weight
2	CAA	Conolly accessible area
4	MR	Molar refractivity
5	MV	Molar volume
6	P	Polarizability
7	SHBint5	Sum of E-State descriptors of strength for potential Hydrogen Bonds of path length 5
8	WD.Unity	Non-directional WHIM, weighted by unit weights
9	Wlambda2.eneg	Directional WHIM, weighted by Mulliken atomic electronegativites
10	Wlambda3.polar	Directional WHIM, weighted by atomic polarizabilities
11	BCUTw_11	nhigh lowest atom weighted BCUTS
12	PNSA_3	Charge weighted partial negative surface area
13	MOMI-XZ	Moment of inertial x/z
14	PSA	Polar surface area

Table 3.6: Molecular descriptors symbols and definition in the models (continuation)

15	MEP	Maximum Electrostatic potential
16	cLogp	Octanol-water partition coefficient
17	SE	Solvation energy
18	H^o	Standard enthalpy
19	G^o	Standard Gibbs energy

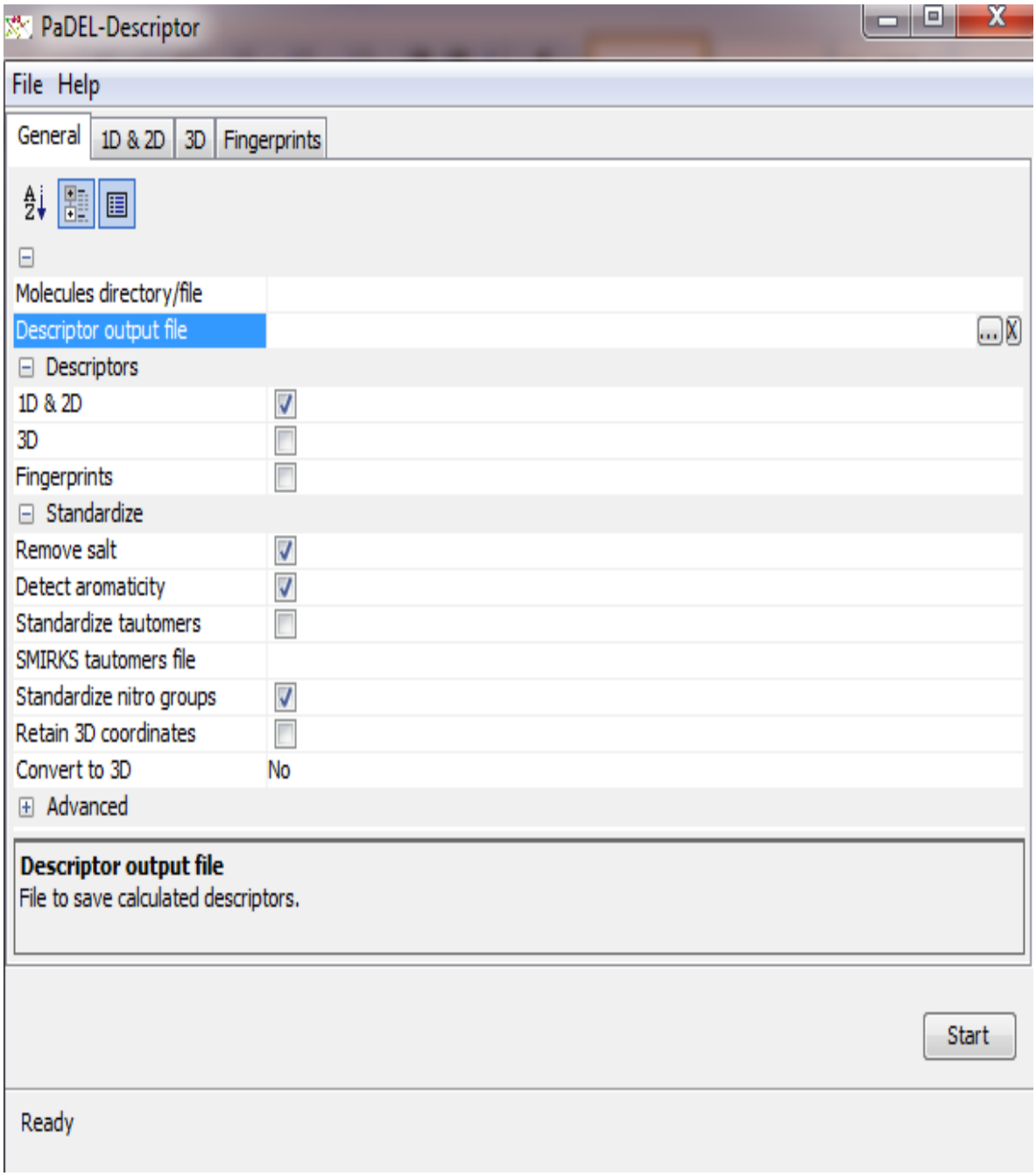


Plate 3.7: Padel-Descriptor interface

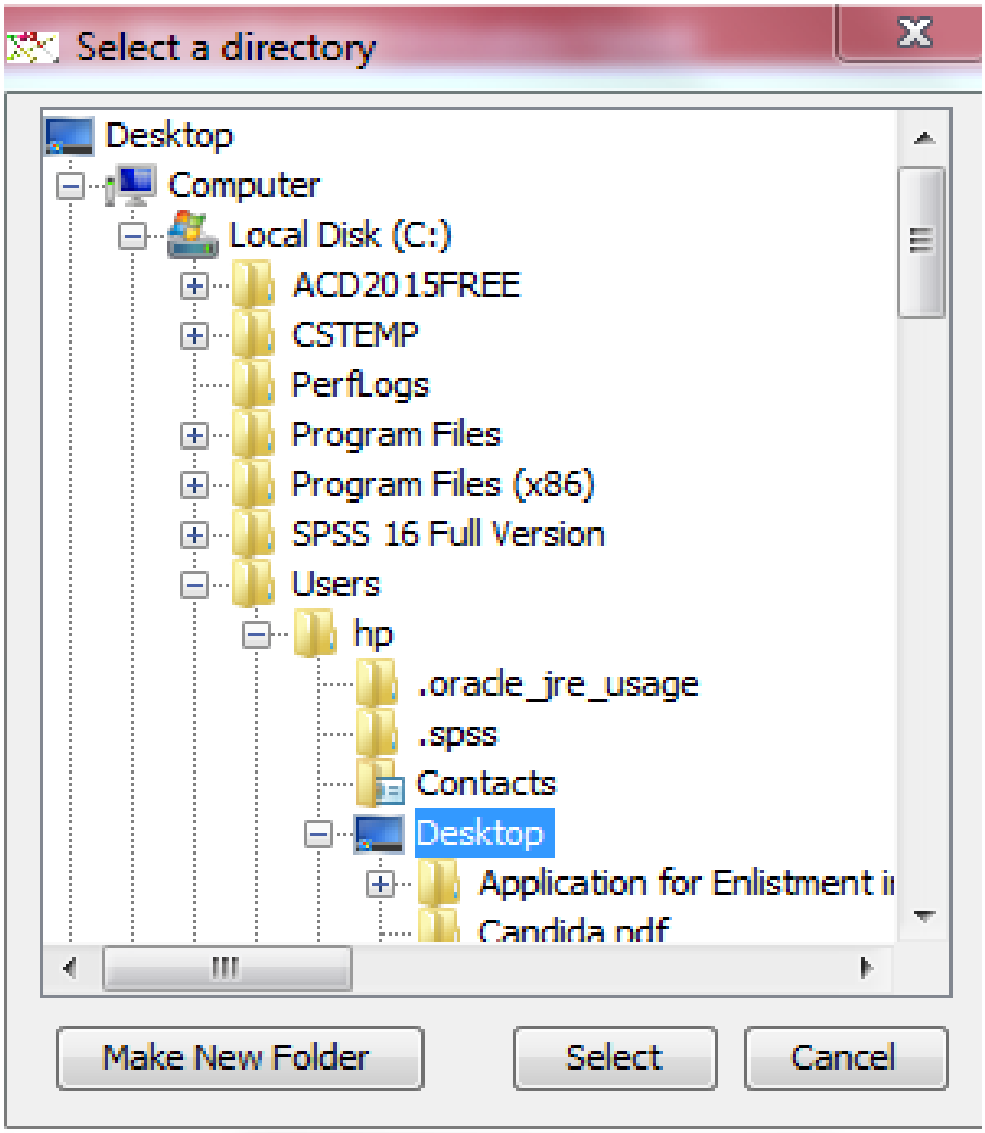


Plate 3.8: Padel Descriptor *molecules directory/files*

options- ***1D, 2D,***and***3D***on the Padel window (Plate 3.7) and marking appropriate boxes against each descriptors, the descriptors of choice were selected, the molecules were also standardized by checking the options such as ***retain 3D coordinates***, ***convert to 3D***,***remove salt***, ***detect aromaticity***, **standardize tautomers** and calculations commenced via the ***start***option at the bottom right of the interface (Plate 3.3). The results of the calculations were obtained from the output file specified above on Microsoft excel sheet.

(b) Computing molecular descriptors using Spartan software

By launching the Spartan 14 software, its interface (Plate 3.3) was created. The optimized molecule's structural file saved as ****sdf***file as described in section (b) was imported unto the interface by clicking the ***file menu → open*** option on the menu bar resulting to the propping of a file browser (Plate 3.4). On the file browser, the folder containing the optimized molecule was indicated via the ***file name*** checkbox at the bottom of the browser. The molecule was selected from its folder and ***all files* (*)** selected. By clicking on the ***open*** option also at the bottom of the browser, optimized molecule was imported unto the interface of Spartan 14 software (Plate 3.3). The properties of the molecule were then accessed by selecting ***display→ properties*** options on menu bar prompting the propping up a window shown in Plate 3.9. The properties of the molecule were then accessed via the ***molecule***, ***QSAR***, ***thermodynamic*** and ***acidity/basicity*** menus of the window by selecting appropriate menu. This same procedures apply to all molecules in the data set.

(c) Computing molecular descriptors using ChmBio3D Pro 12.0.1V

The ChmBio3D Pro 12.0.1V was launched creating its interface shown in Plate 3.10. Structure of the optimized molecule saved in ***sdf***file format as described in section (a) was imported into the model area window via the ***file→ import file menu*** on the menu

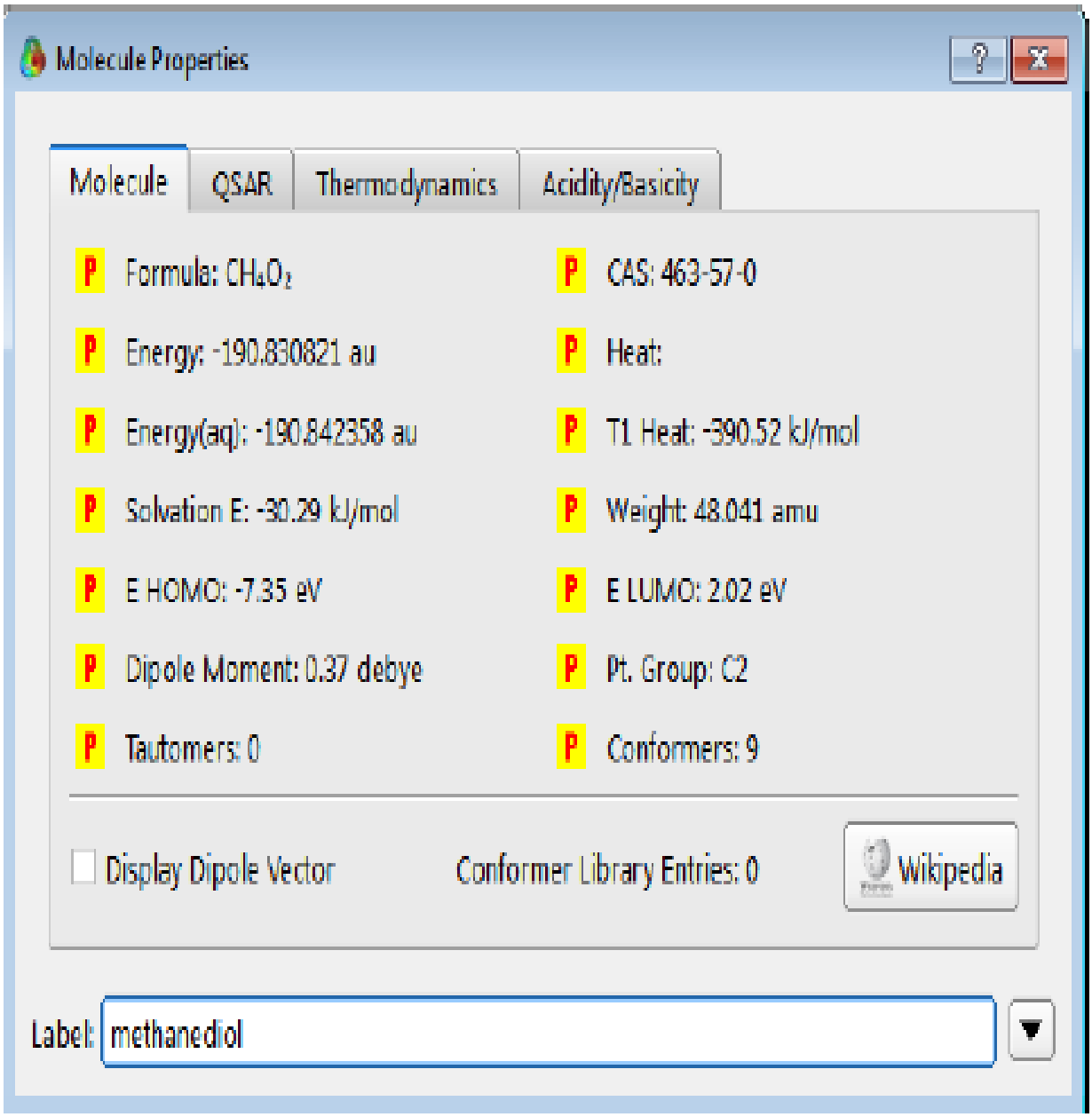


Plate 3.9: Spartan' 14 Molecular properties dialog

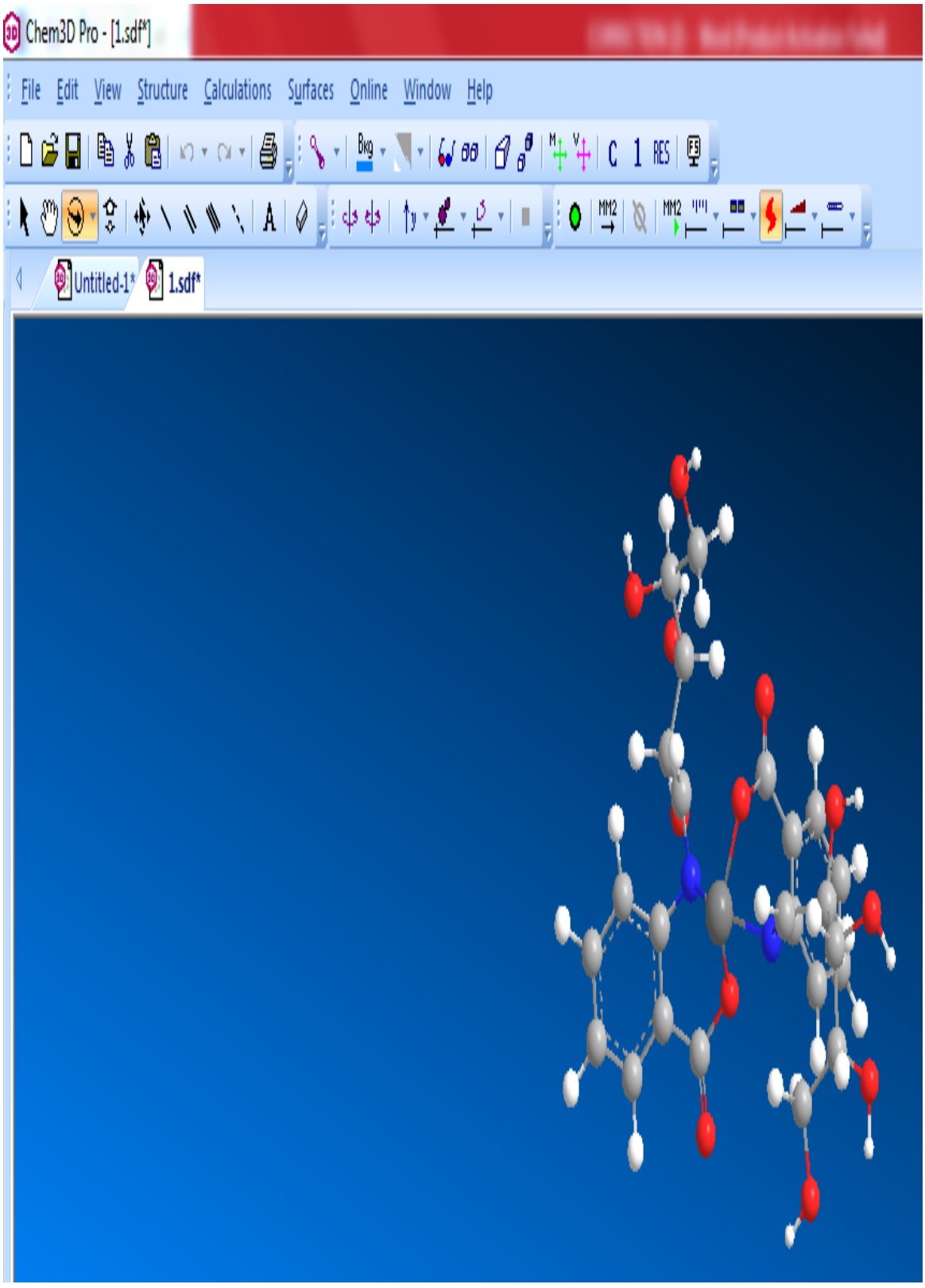


Plate 3.10: ChmBio3D Pro 12.0.1V interface

bar causing the popping up of a file browser (Plate 3.4). Through the ***file name*** and ***files of type*** check boxes of the browser, the name of the optimized molecule and its file extension (****sdf*** file extension in this case) were indicated respectively, and the molecule was imported unto the ChmBio3D Pro 12.0.1V interface by clicking ***open***on the browser's window. By clicking the ***calculation*** → ***Compute properties*** menus, a small window shown in Plate 3.11 popped up.On the window, the box corresponding to ***Chem prop Std*** was marked and by clicking the ***Ok*** option at the bottom of the window, the properties of the optimized molecule were computed and the results displayed at the bottom end of the ChmBio3D window. The same procedure applies to all other molecules in each of the data set.

3.2.4 Normalization of data

There exists a great deal of variability in the range and distribution of each variable in the data set. However, this may pose a problem for data mining algorithms such as Genetic function approximation. Such situation is handled by normalization of the data(Brameier and Banzhaf, 2001). Simple mathematical functions can be applied to achieve normality or symmetry in the data distribution. A commonly used approach is to apply logarithmic transformation on the variable of interest in order to achieve distribution approaching normality. This is typically performed on dependent variables such as the modeled biological/ chemical properties of interest whereby IC50 may be transformed to – logIC50. Practically, such mathematical operation is applied to each individual value of a given variable of interest(Guttman *et al*., 1944).

In this present study, normalization was performed on the dependent variables by transforming the experimentally determined MIC values to logarithmic scale (log MIC = pMIC).These operations were performed in order to reduce the dispersion of data set and to get linear response and well data fitting.

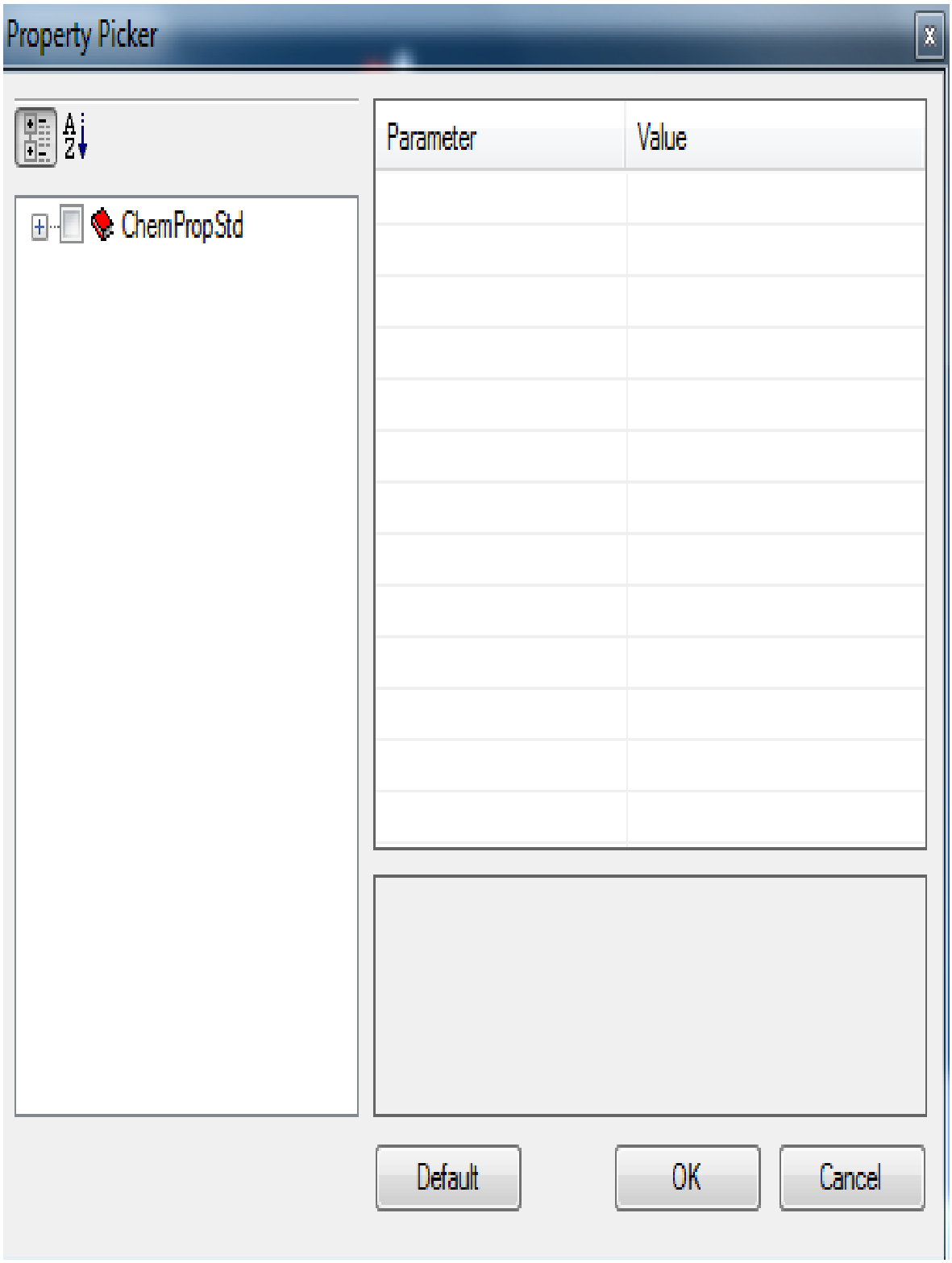


Plate 3.11: ChmBio3D Pro 12.0.1V molecular properties dialog box

Likewise, it is a common belief that the explanatory variables (independent variables) must be standardized in building regression models since different explanatory variables are almost always on different scales (i.e., measured in different units). However, the Material Studio software used for this GFA-regression analysis like other modern software provides standardized coefficients, i.e. the betas are estimated such that they convert the units of each explanatory variable into the units of the response variable appropriately.

3.2.5 Training and test set

In this stage, the data for each biological activity was split into training set and test set. The training set was used to adjust the parameters of the model while the test set was used to evaluate its prediction ability. This splitting was done manually and at random. At least 70% of the data set in each case was used as training set and the rest as test set. The data set for minimum inhibitory concentration (*MIC*) values of the molecules against *Candida albicans* and *Epidermophyton floccosum* contain one hundred and six (106) compounds whose activities have been determined experimentally. Eighty eight (88) compounds was used for modelling the *MIC* values of the compounds against *Candida albicans* (68 training sets and 20 test set) and Eighteen (18) compounds for modelling the *MIC* values of the compounds against *Epidermophyton floccosum* (13 training sets and 5 test set).

3.2.6 Learning process

During this process, the correlations between biological activities (MIC) of the complexes and the calculated descriptors were obtained via correlation analysis using the Microsoft excel package in Microsoft office 2013. Pearson's correlation matrix was

used to model, in order to select the suitable descriptors for each regression analysis. The selected descriptors were subjected to regression analysis with the experimentally determined activities as the dependent variable and the selected descriptors as the independent variables using Genetic function approximation (GFA) method in Material studio software. It is a distinctive characteristic of GFA that it could create a population of models rather than a single model. GFA algorithm, selecting the basic functions genetically, developed better models than those made using stepwise regression methods. And then, the models were estimated using the "lack of fit" (LOF), which was measured using a slight variation of the original Friedman formula (Khaled and El-Sherik, 2013), so that best model received the best fitness score (Hirashima *et al*., 1999; Khaled and El-Sherik, 2013; Khaled, 2011).

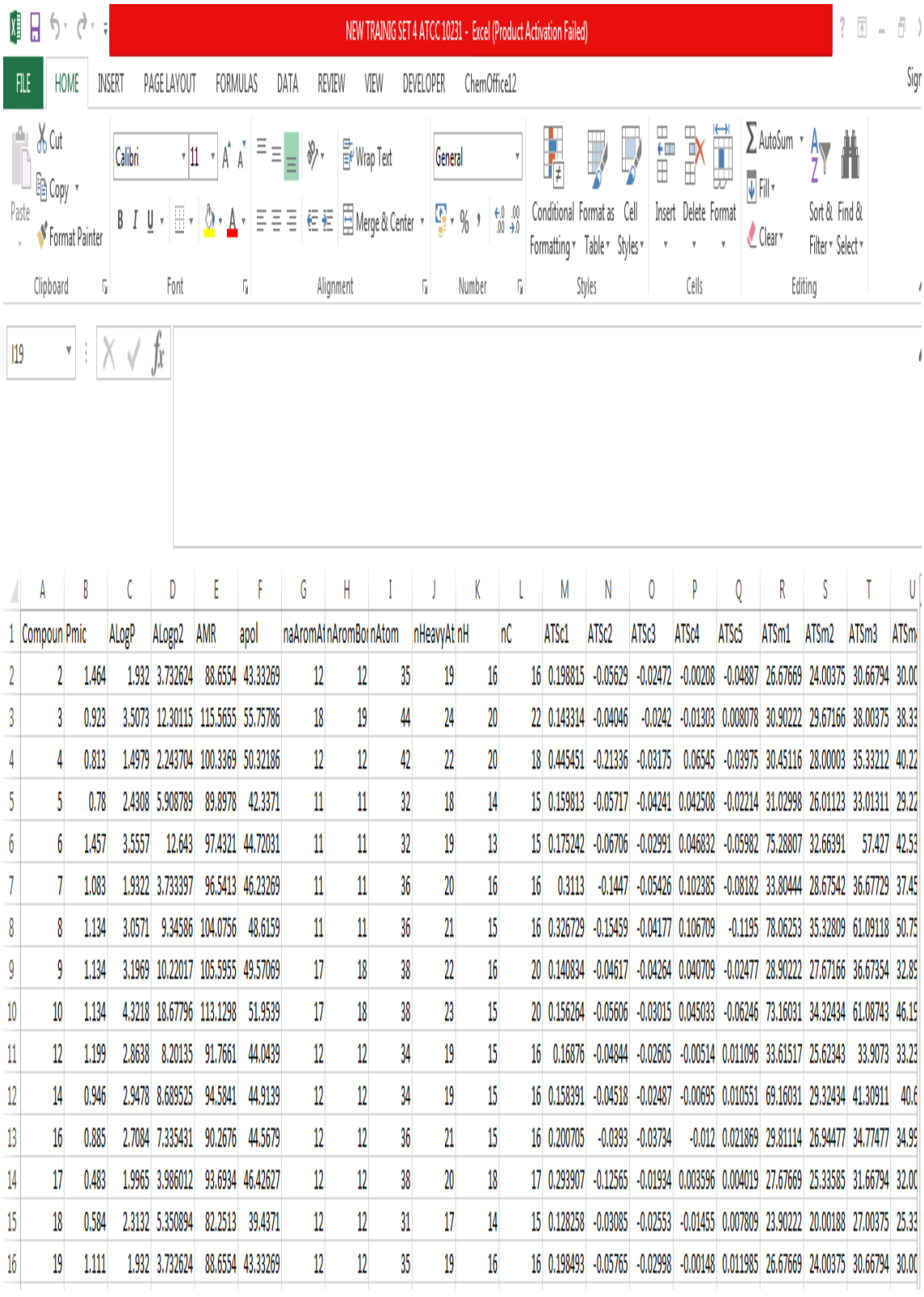
The genetic function approximation (GFA) algorithm offers a new approach to the problem of building quantitative structure-activity relationship (QSAR) and quantitative structure-property relationship (QSPR) models. Replacing regression analysis with the GFA algorithm enables the construction of models competitive with or superior to those produced by standard techniques and makes available additional information not provided by other techniques. Unlike most other analysis algorithms, GFA gives multiple models, where the populations of the models are created by evolving random initial models using a genetic algorithm. GFA can build models using not only linear polynomials but also higher-order polynomials, splines, and other nonlinear functions (Khaled and Abdel-Shafi, 2011). Friedman's multivariate adaptive regression splines (MARS) algorithm is a statistical technique for modeling data. It provides an error measure, called the lack-of-fit (LOF) score, eqn 3.1 that automatically penalizes models with too many features. It also inspired the use of splines as a powerful tool for nonlinear modeling (Khaled and Abdel-Shafi, 2011). The equation below is for LOF:

$$LOF = \frac{SSE}{\left(1-\frac{c+d_p}{M}\right)^2} \quad 3.1$$

where SSE is the sum of squares of errors, c is the number of terms in the model, other than the constant term, d is a user-defined smoothing parameter , p is the total number of descriptors contained in all model terms (again ignoring the constant term) and M is the number of samples in the training set. Unlike the commonly used least squares measure, the LOF measure cannot always be reduced by adding more terms to the regression model. While the new term may reduce the SSE, it also increases the values of c and p, which tends to increase the LOF score. Thus, adding a new term may reduce the SSE, but actually increases the LOF score. By limiting the tendency to simply add more terms, the LOF measure resists overfitting better than the SSE measure (Khaled and Abdel-Shafi, 2011) .

(a) Steps in correlation analysis in Excel

The Excel Microsoft package was launched by clicking on ***Microsoft excel***in the Microsoft office to create its interface as in Plate 3.12. Data comprising of the experimental pMIC of the molecules with their computed descriptors were entered into the cells with the bioactivity of the molecules occupying the first column (Plate 3.12). On the interface, ***Tools*** followed by ***data analysis*** options were selected. A window shown in Plate 3.13 containing list of statistical operations propped up. By scrolling down the list, ***correlation → OK***options were selected. A *correlation* window shown in Plate 3.14 with the options: ***output***and ***input***range propped up. In the ***input range*** box, the columns of the data comprising of the dependent variable and the computed descriptors were highlighted. The cell range automatically appeared in the box. The ***label***box was also ticked. Byclicking on the ***output range*** *box*, a region on the worksheet where the results



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Compoun	Pmic	ALogP	ALogp2	AMR	apol	naAromAt	nAromBor	nAtom	nHeavyAt	nH	nC	ATSc1	ATSc2	ATSc3	ATSc4	ATSc5	ATSm1	ATSm2	ATSm3	ATSm
2	2	1.464	1.932	3.732624	88.6554	43.33269	12	12	35	19	16	16	0.198815	-0.05629	-0.02472	-0.00208	-0.04887	26.67669	24.00375	30.66794	30.0
3	3	0.923	3.5073	12.30115	115.5655	55.75786	18	19	44	24	20	22	0.143314	-0.04046	-0.0242	-0.01303	0.008078	30.90222	29.67166	38.00375	38.3
4	4	0.813	1.4979	2.243704	100.3369	50.32186	12	12	42	22	20	18	0.445451	-0.21336	-0.03175	0.06545	-0.03975	30.45116	28.00003	35.33212	40.2
5	5	0.78	2.4308	5.908789	89.8978	42.3371	11	11	32	18	14	15	0.159813	-0.05717	-0.04241	0.042508	-0.02214	31.02998	26.01123	33.01311	29.2
6	6	1.457	3.5557	12.643	97.4321	44.72031	11	11	32	19	13	15	0.175242	-0.06706	-0.02991	0.046832	-0.05982	75.28807	32.66391	57.427	42.5
7	7	1.083	1.9322	3.733397	96.5413	46.23269	11	11	36	20	16	16	0.3113	-0.1447	-0.05426	0.102385	-0.08182	33.80444	28.67542	36.67729	37.4
8	8	1.134	3.0571	9.34586	104.0756	48.6159	11	11	36	21	15	16	0.326729	-0.15459	-0.04177	0.106709	-0.1195	78.06253	35.32809	61.09118	50.7
9	9	1.134	3.1969	10.22017	105.5955	49.57069	17	18	38	22	16	20	0.140834	-0.04617	-0.04264	0.040709	-0.02477	28.90222	27.67166	36.67354	32.8
10	10	1.134	4.3218	18.67796	113.1298	51.9539	17	18	38	23	15	20	0.156264	-0.05606	-0.03015	0.045033	-0.06246	73.16031	34.32434	61.08743	46.1
11	12	1.199	2.8638	8.20135	91.7661	44.0439	12	12	34	19	15	16	0.16876	-0.04844	-0.02605	-0.00514	0.011096	33.61517	25.62343	33.9073	33.2
12	14	0.946	2.9478	8.689525	94.5841	44.9139	12	12	34	19	15	16	0.158391	-0.04518	-0.02487	-0.00695	0.010551	69.16031	29.32434	41.30911	40.6
13	16	0.885	2.7084	7.335431	90.2676	44.5679	12	12	36	21	15	16	0.200705	-0.0393	-0.03734	-0.012	0.021869	29.81114	26.94477	34.77477	34.9
14	17	0.483	1.9965	3.986012	93.6934	46.42627	12	12	38	20	18	17	0.293907	-0.12565	-0.01934	0.003596	0.004019	27.67669	25.33585	31.66794	32.0
15	18	0.584	2.3132	5.350894	82.2513	39.4371	12	12	31	17	14	15	0.128258	-0.03085	-0.02553	-0.01455	0.007809	23.90222	20.00188	27.00375	25.3
16	19	1.111	1.932	3.732624	88.6554	43.33269	12	12	35	19	16	16	0.198493	-0.05765	-0.02998	-0.00148	0.011985	26.67669	24.00375	30.66794	30.0

Plate 3.12: Excel window of calculated descriptors

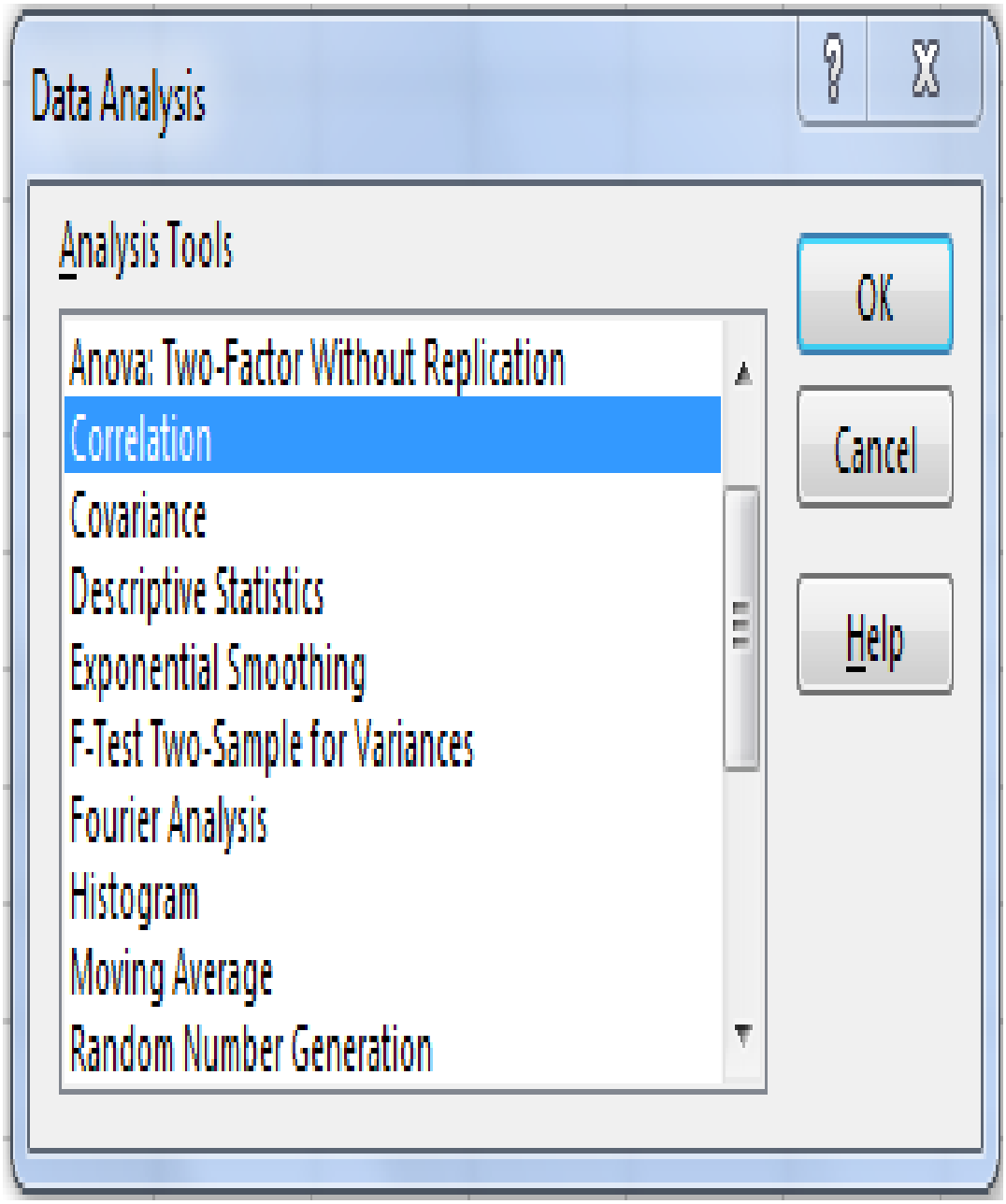


Plate 3.13: Microsoft Excel *Data Analysis* Dialog

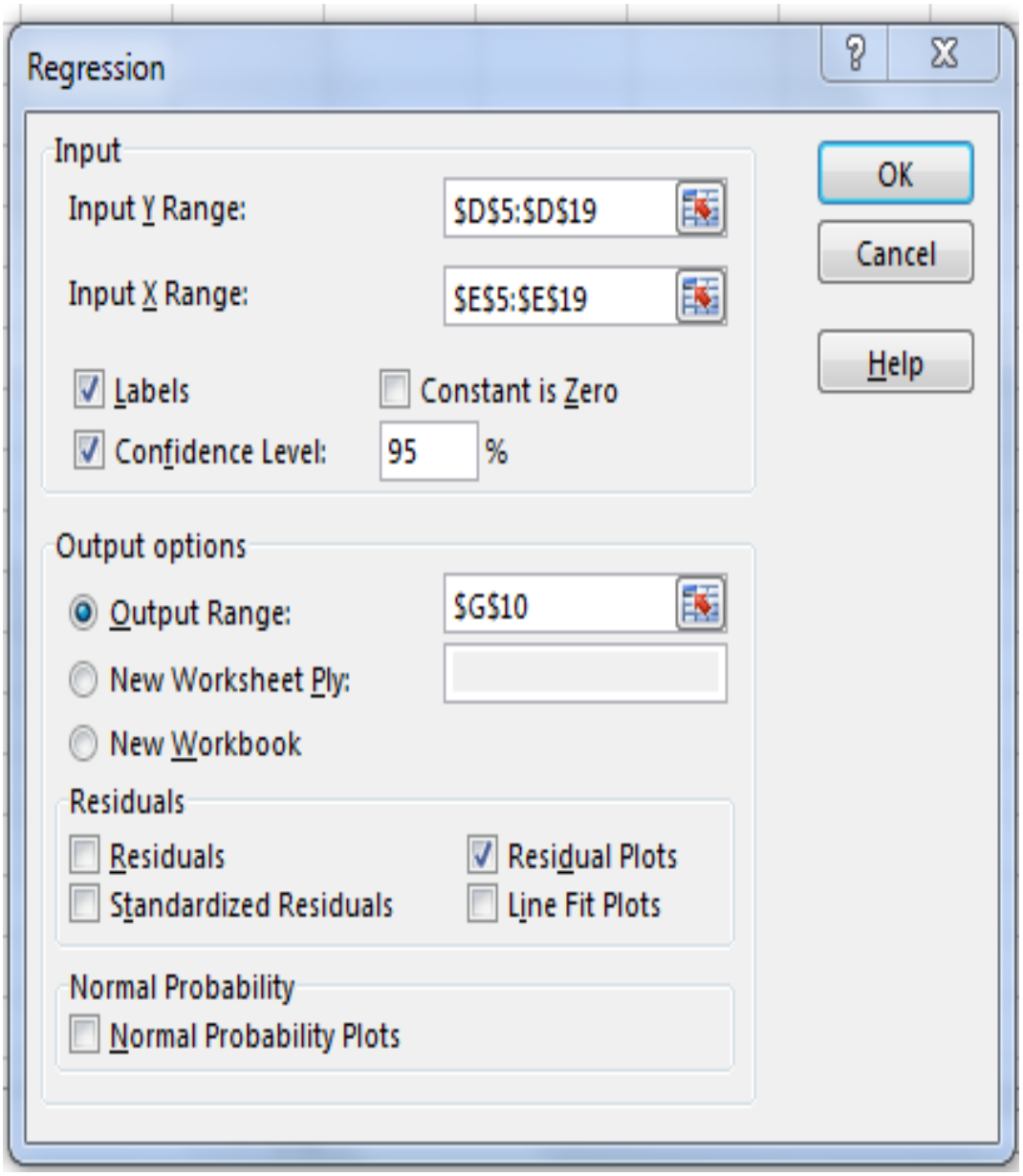


Plate 3.14: Microsoft Excel *Correlation* Dialog box

were to be displayed was selected by clicking on a single empty cell outside. The results were displayed in the formof a correlation matrix. From the correlation coefficient obtained from the second column of the matrix, the descriptors that were found to be strongly correlated to each bioactivity were copied to a new excel worksheet and saved as ***excel workbook*** file extension taking note of possible inter-correlation among the descriptors. Inter-correlation occur when two descriptors are strongly correlated to each other. In such case, only one of such descriptors were selected based on the one with the highest coefficient of correlation with the dependent variable (bioactivities).

(b) Steps involved in building GFA models

The *material studio software* was launched from the desk top prompting a prop up of the windows (work space window and project window) shown in Plate 3.15. By selecting theoptions, ***create a new project →Ok*** on the propped up window, a file browser (Plate 3.4) appeared. Through the file name option on the browser, the name of the file was indicated and saved as ***project files* (**stp*)** causing the appearance of full material studio work space shown in Plate 3.16. By clicking the ***file → new project***menus via the menu bar of the work space, a file browser (Plate 3.4) propped up. Through the ***file name*** option checkbox of the browser, the file name of the new project was indicated and saved as ***project files* (**stp*).** Through the file menu on the menu bar of the work space, study table was created by selecting ***new →study table →*** **Ok** options on a propped up document window shown in Plate 3.16 causing the appearance of the Material Studio study table or work space as shown on Plate 3.17.The computed descriptors with the corresponding bioactivity saved as ***excel work book*** extension described in section (a) above were highlighted, copied and pasted unto the spread sheet of Material studio's

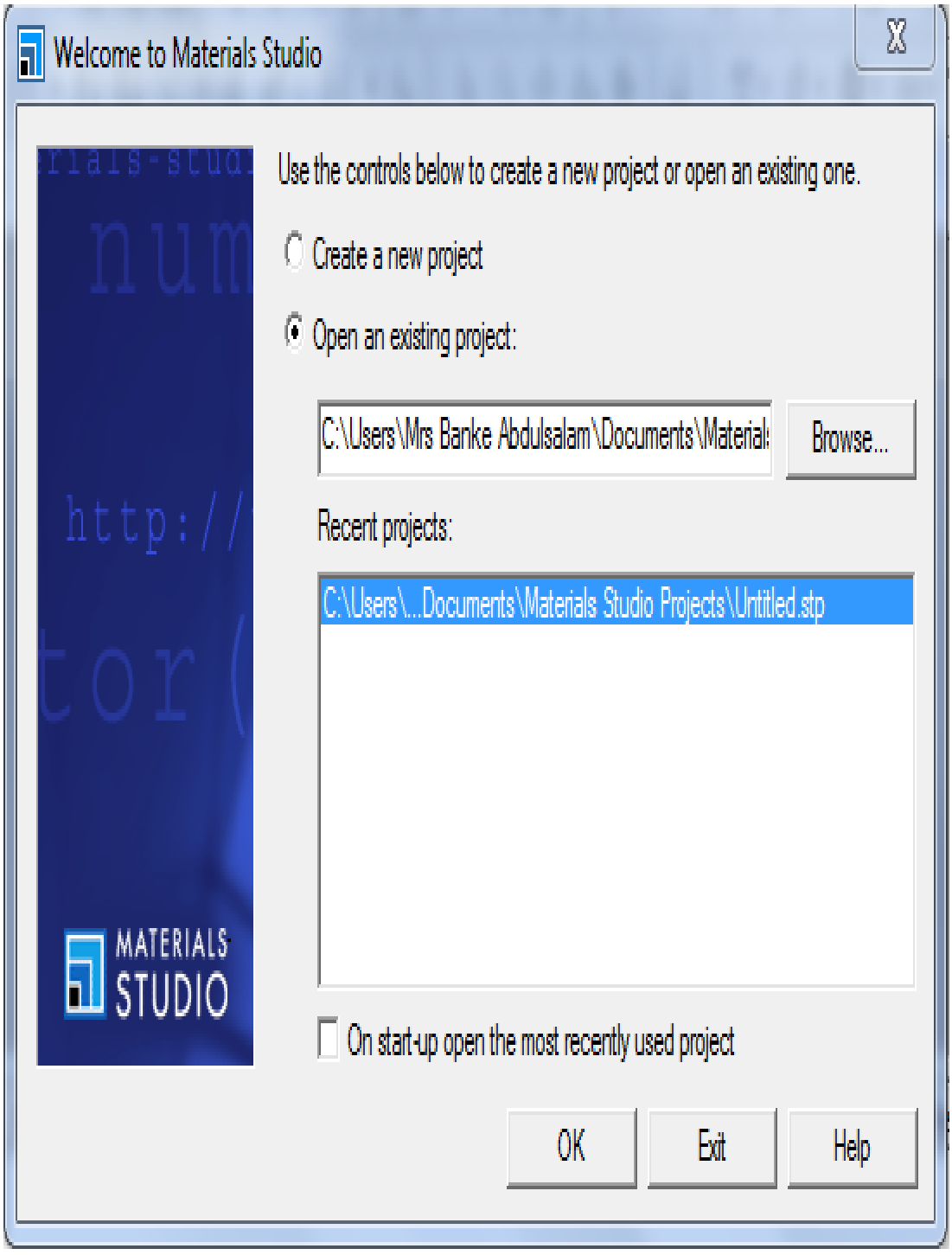


Plate 3.15: Material Studio's work space window and project window

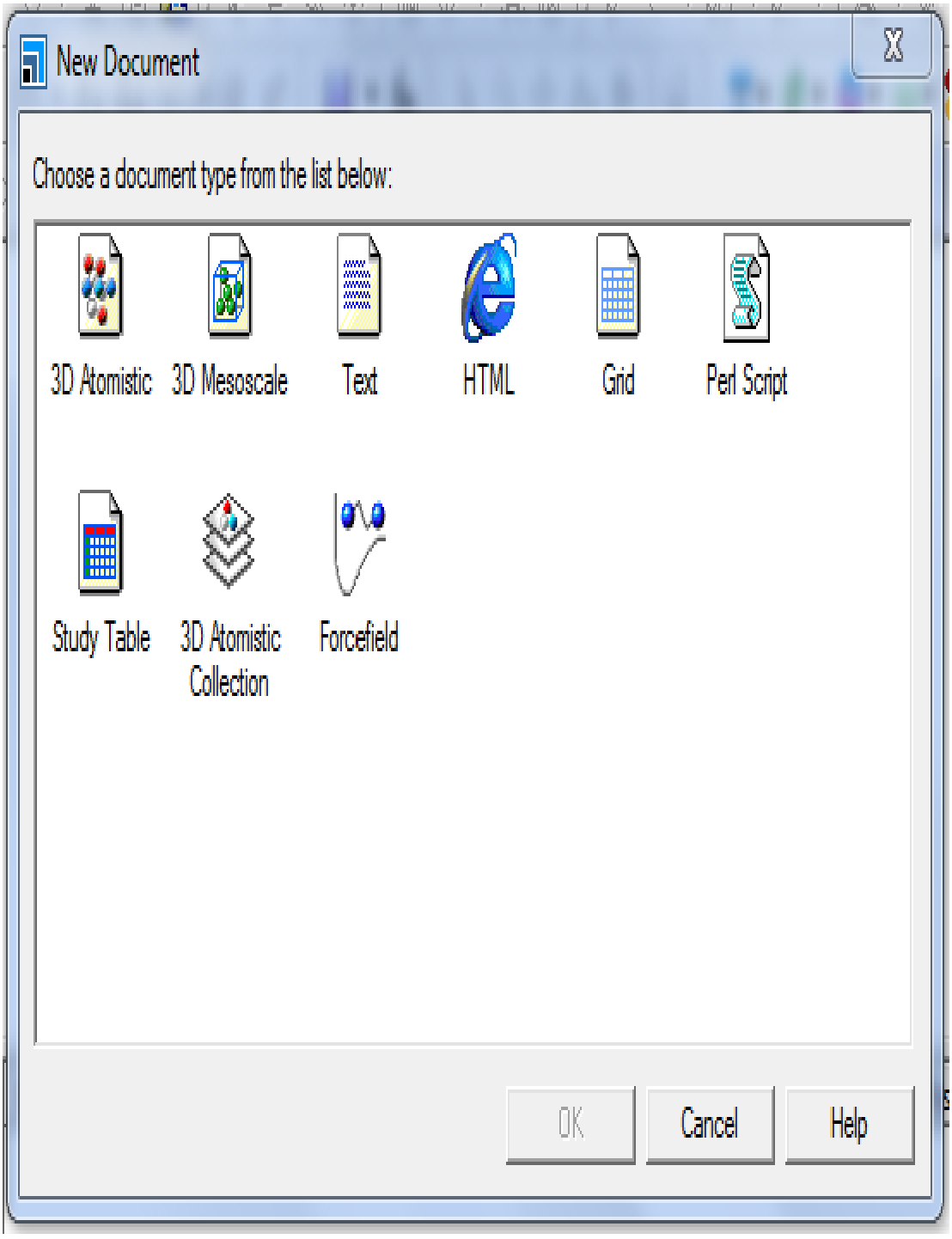


Plate 3.16: Material Studio's *document* window

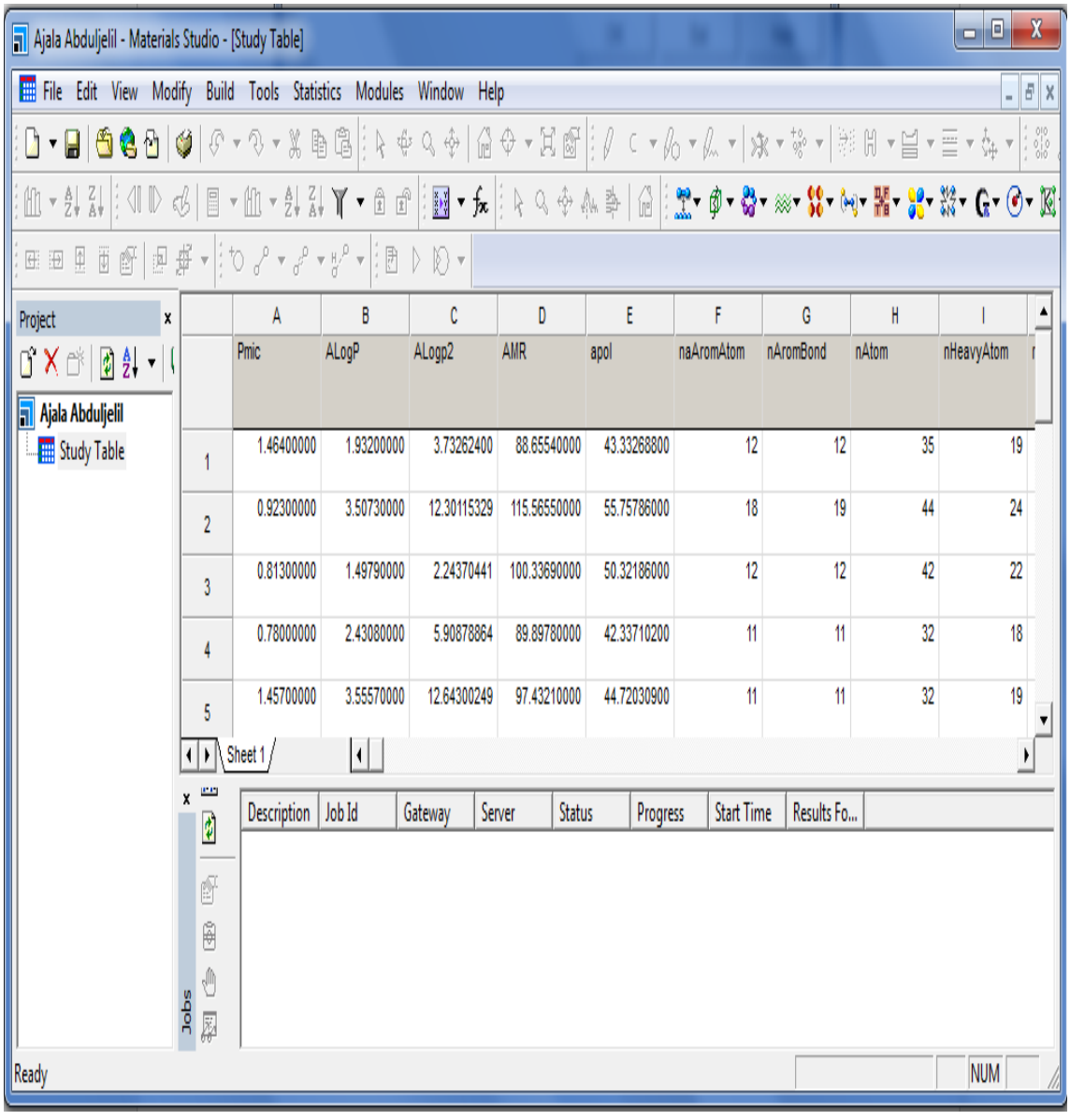


Plate 3.17**:** materials studio work space

study table with the first column comprising the dependent variable (pMIC) and subsequent columns comprising of the descriptors obtained from correlation study.

The column for dependent variable was highlighted and the options: **Statistics → *model building* → *Genetic function approximations***were selected. A Genetic function approximation window shown in Plate 3.18 appeared at the top left corner of the workspace from where the boxes against the predictor variables (descriptors) were marked via the *variable selection* menu of the window, the Population and Generation were set to 1,000 and 5,000 respectively via the *parameter* menu as shown in Plate 3.19.The number of top equations returned was set to 3, Mutation probability was set to 0.1, and the smoothing parameter was set to 0.5 in line with standard settings used in GFA – QSAR study. Equation length was adjusted by increasing the initial equation lenth to six and maximum equation lenth to 10 and checked the constant equation lenth via the *equation data* menu as shown in Plate 3.20. By clicking the ***Ok*** button, calculations commenced immediately. The results of the calculations were displayed in tabular form on the study table of the material studio as shown in Plate 3.21. The models were scored based on the one with the least Friedman's LOF(Khaled, 2011). Using this same procedure, models were generated for each data set.

3.2.7 Evaluation of the validity and predictability of the developed QSAR models

The fitting ability, stability, reliability and predictive ability of the best models were evaluated by internal and external validation parameters. The validation parameters were compared with the minimum recommended value (Table 3.7) for a generally acceptable QSAR model.

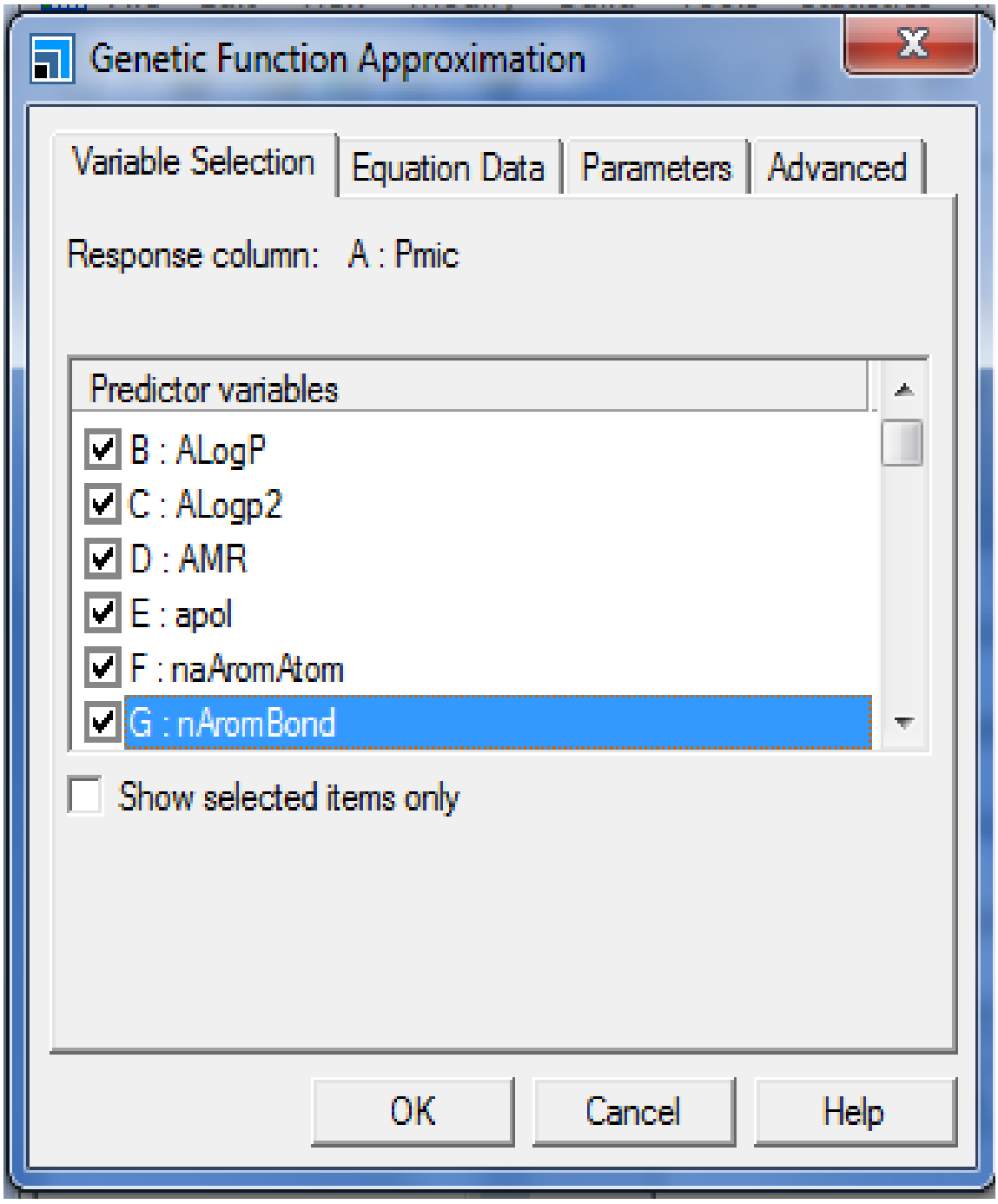


Plate 3.18: Material Studio's *Genetic function approximation* window

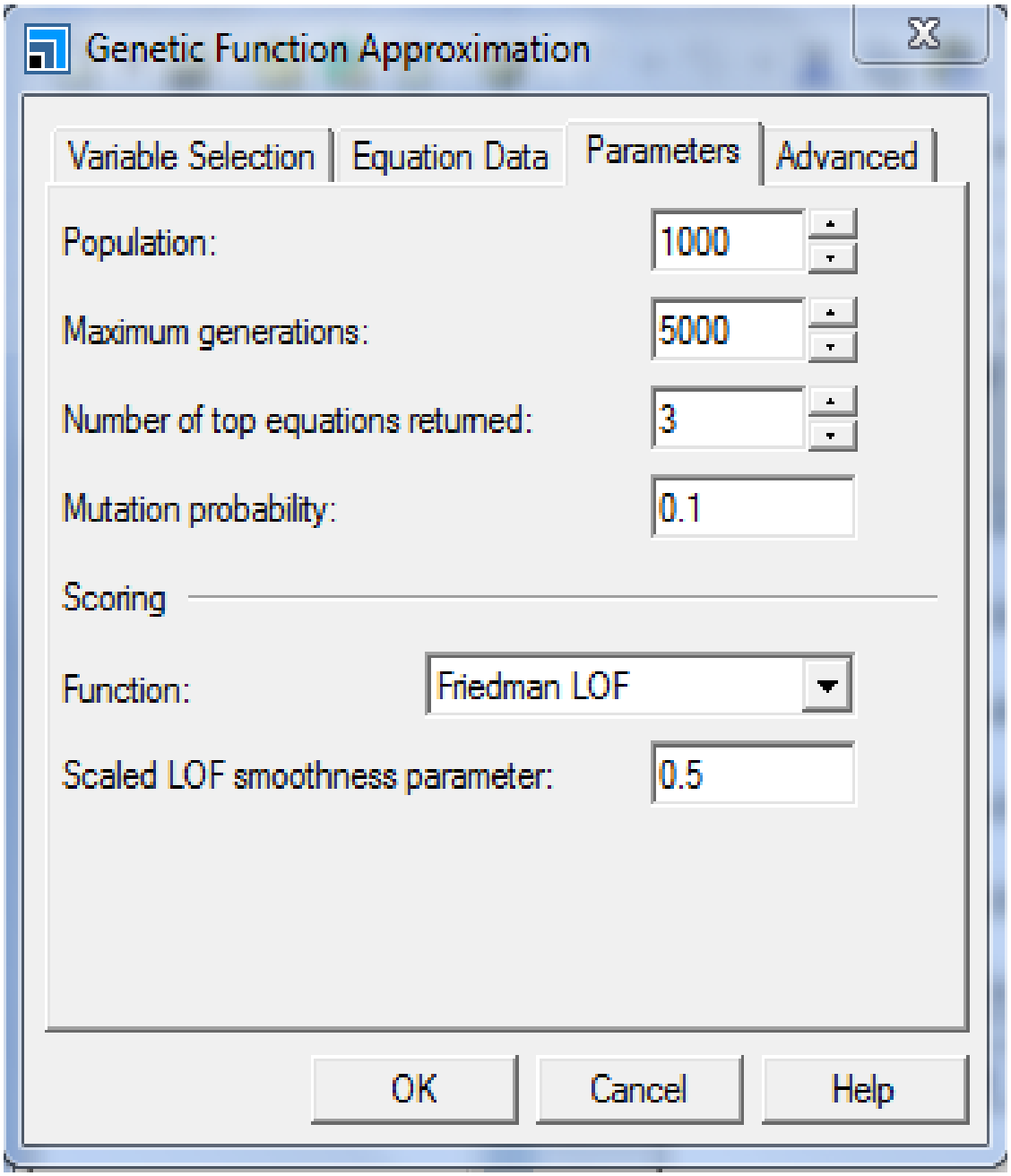


Plate 3.19: variable selection menu

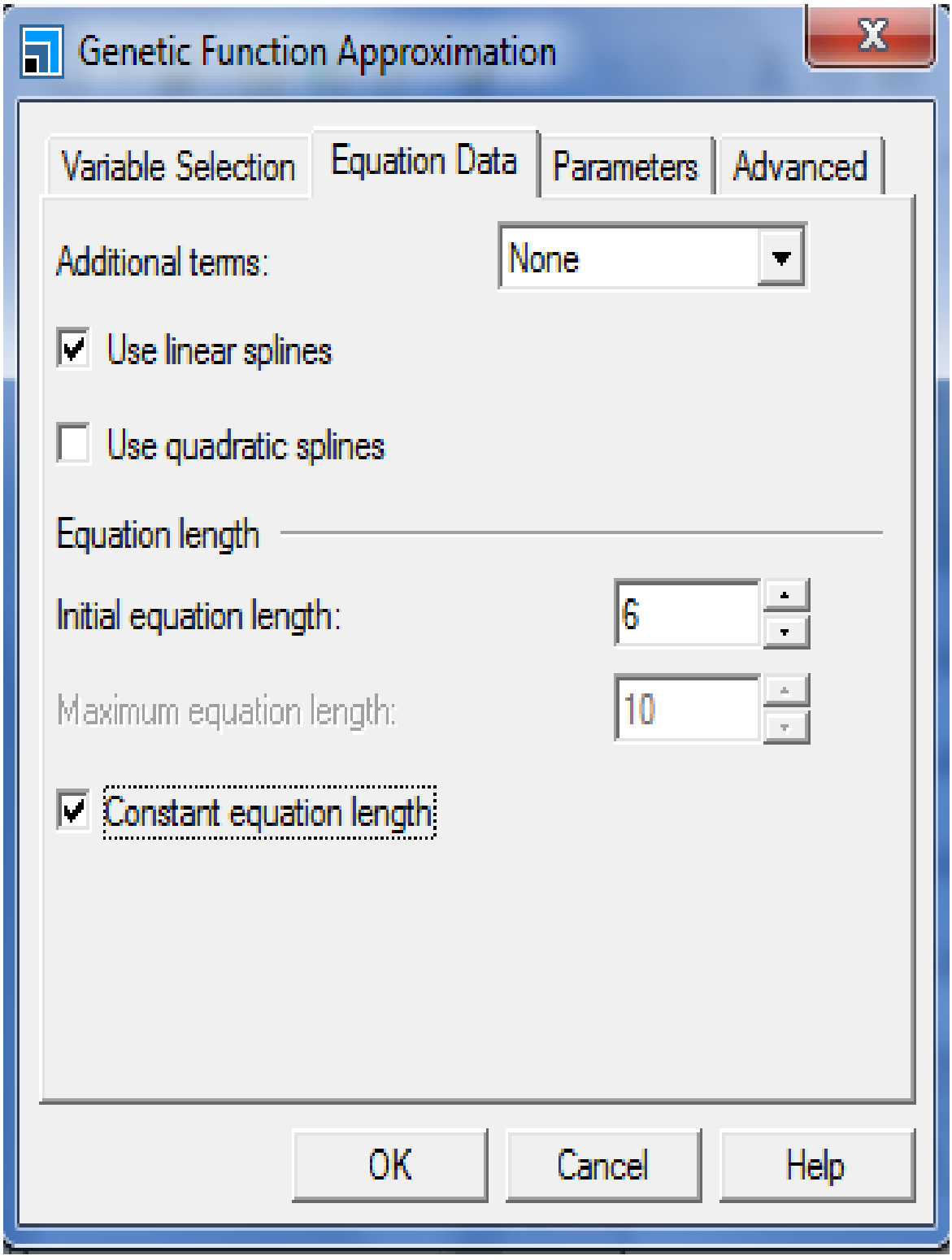
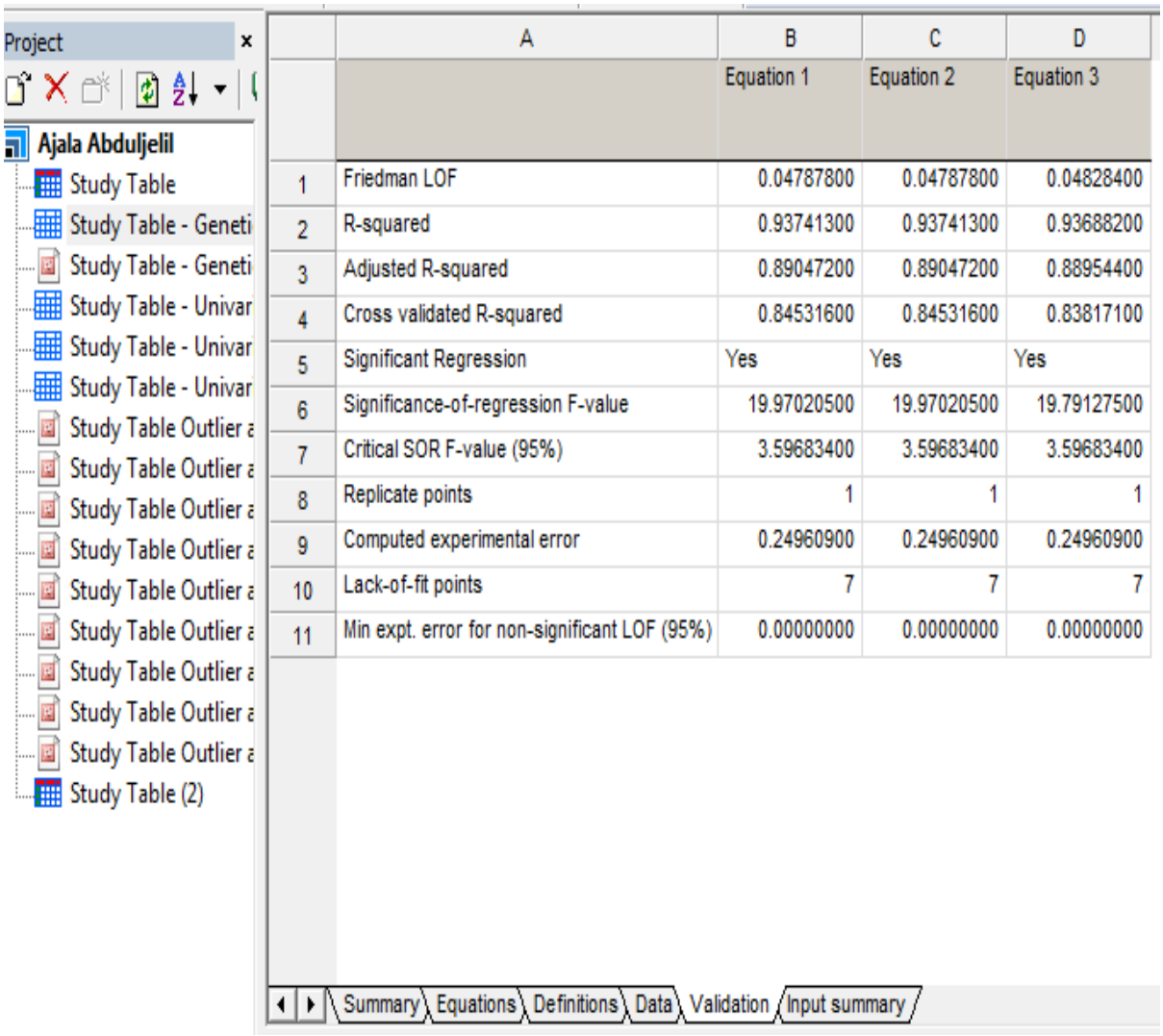


Plate 3.20: GFA equation data menu



	A	B	C	D
		Equation 1	Equation 2	Equation 3
1	Friedman LOF	0.04787800	0.04787800	0.04828400
2	R-squared	0.93741300	0.93741300	0.93688200
3	Adjusted R-squared	0.89047200	0.89047200	0.88954400
4	Cross validated R-squared	0.84531600	0.84531600	0.83817100
5	Significant Regression	Yes	Yes	Yes
6	Significance-of-regression F-value	19.97020500	19.97020500	19.79127500
7	Critical SOR F-value (95%)	3.59683400	3.59683400	3.59683400
8	Replicate points	1	1	1
9	Computed experimental error	0.24960900	0.24960900	0.24960900
10	Lack-of-fit points	7	7	7
11	Min expt. error for non-significant LOF (95%)	0.00000000	0.00000000	0.00000000

Plate 3.21: Result template on the GFA study table.

Table 3.7 : Validation metrics for a generally acceptable QSAR model

Name	Threshold
Coefficient of determination	≥ 0.6
LOO cross validation coefficient	> 0.5
External test set's coefficient of determination	≥ 0.6
Difference between R^2 and Q^2	≤ 0.3
Variation ratio	High
Golbraikh and Tropsha condition r^2	< 0.1
Golbraikh and Tropsha condition r'^2	< 0.1
Intercept	$0.85 \leq k$ or $k' \leq 1.15$
Variance inflation factor	$1 \leq VIF \leq 10$

(a) Internal validation parameters

The internal validation parameters was done using the data that created the model. The real usefulness of QSAR models is not just their ability to reproduce known data, verified by their fitting power (R^2), but mainly in their potential for predictive application. For this reason, the internal consistency of the training set was confirmed by using thesquare of the correlation coefficient (R^2), adjusted R-squared (R^2_{adj}), leave one out cross valudation coefficient (R^2_{cv}), variance ratio (F) and standard error of estimate (SE) method to ensure the robustness of the model, (Gramatica and Esterpapa, 2007)

(i) R^2 (square of the correlation coefficient) : estimates the proportion of the variation of biological activity explained by the regression (the explained variance of the model). The stronger the relationship between the dependent and the independent variables the more R^2 approaches unity (Gramatica, 2002).

$$R^2 = 1 - \frac{\sum(Y_{obs} - Y_{pred})^2}{\sum(Y_{obs} - \bar{Y}_{tra})^2} = 1 - \frac{RSS}{TSS} \qquad 3.2$$

In the above equation, Y_{obs} stands for the observed response value, while Y_{pred} is the model-derived predicted response and $\overline{Y}_{tra}$ is the mean experimental activity of the training set values, where RSS is the regression sum of squares and TSS is the total sum of squares.

(ii) Adjusted R^2 (R^2_{adj}): Adjusted R-Squared : (R^2_{adj}) is obtained by dividing the residual sum of squares and total sum of squares by their respective number of degrees of freedom. The value of R^2_{adj} decreases if an added variable to the equation does not reduce the unexplained variance (Gramaticaand Estherpapa, 2007)

$$R^2_{adj} = 1 - \frac{RSS/(n-p-1)}{TSS/(n-1)} = 1 - (1 - R^2)\left((n-1)/(n-p)\right) \quad 3.3$$

Where n = number of objects

P = number of predictor variables

RSS = residual sum of square

TSS = total sum of squares

(iii) R^2_{cv} (Leave one out cross validation coefficient):

The R^2_{cv} represents certain perturbations to the model and data size reduction, the corresponding statistics are always characterized by the relations $R^2 > Q^2$(Kiralj and Ferreira, 2009). The minimum acceptable statistics for regression models in QSAR and QSPR include conditions $Q^2 > 0.5$ and $R^2 > 0.6$ (Gramatica, 2002).

$$R^2_{cv} = 1 - \frac{\sum(y_i - \bar{y}_i)^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{PRESS_{CV}}{SSY_{CV}} \quad 3.4$$

Where $PRESS = \sum(y_{obs} - \bar{y}_{pred})^2$

y_i *and* $\bar{y}_i$ represent predicted and observed activity respectively of the training set and $\bar{y}$ is the mean activity value of the training set.

(iv) F-value or variance ratio (i.e Fisher statistic) :

The statistical significance of the regression model can be assessed by means of the Fisher statistic (F). The F-value or variance ratio is the ratio between explained and unexplained variance for a given number of degrees of freedom, respectively p and (n-p- 1), where n is the number of objects and p is the number of predictor variables.

$$F = \frac{MSS/(p)}{RSS/(n-p-1)} \quad 3.5$$

n = number of objects

p = number of predictor variable

mss = model sum of square

rss = residuals sum of square

(v) Standard error of estimate (s):

Low standard error of estimate is an indication of a good model. It is defined as follows:

$$S = \sqrt{\frac{(Yobs-Yp)^2}{N-P-1}} \qquad 3.6$$

Its degree of freedom is N-p-1 where p = number of independent variables in the model and N = number of objects. Y_{obs} and Y_p signify experimental and predicted property, respectively.

(b) External validation parameters

The external predictivity is to measure the model's predictive power for compounds. In order to estimate the predictive power of the models, one need to have more complex scores for the quantitative models. The predictability is quantified as: squared correlation coefficient (R^2_{pred}), and Golbraikh and Tropsha's criteria :

(i) Coefficient of multiple determination of prediction – external set (R^2_{pred}) : The real predictive ability of any *QSAR* model cannot be judged solely by using internal validation. It has to be validated on the basis of predictions of activities of molecules not used in the models (Gramatica, 2002). Prior to the development of the models, each data set was split into training and test set. QSAR models were built using the training set while the tests set were used for externally validating the models. The external validation test requires only one split of the complete data set into structurally similar training and external validation sets, (Kiralj and Ferreira, 2009). The predicted R^2 was computed in each case using the formulae below

$$R^2_{pred} = 1 - \frac{\sum(Y_{pred}-Y_{obs})^2}{\sum(Y_{obs}-\bar{Y}_{trai})^2} \qquad 3.7$$

Ypred.(test) and Y(obs) indicate predicted and observed activity values respectively of the test set compounds and $\bar{Y}$(trai) indicates mean experimental activity value of the training set.

(ii) Golbraikh and Tropsha's criteria**:** according to Golbraikh and Tropsha, models are considered satisfactory, if all the following conditions are met.

(a) R^2_{pred} □ 0.5

(b) $r^2 - r_0^2 / r^2$ □ 0.1

(c) $r^2 - r'^2_0 / r^2$ □ 0.1

(d) $0.85 \leq k \leq 1.15$

(e) $0.85 \leq k' \leq 1.15$

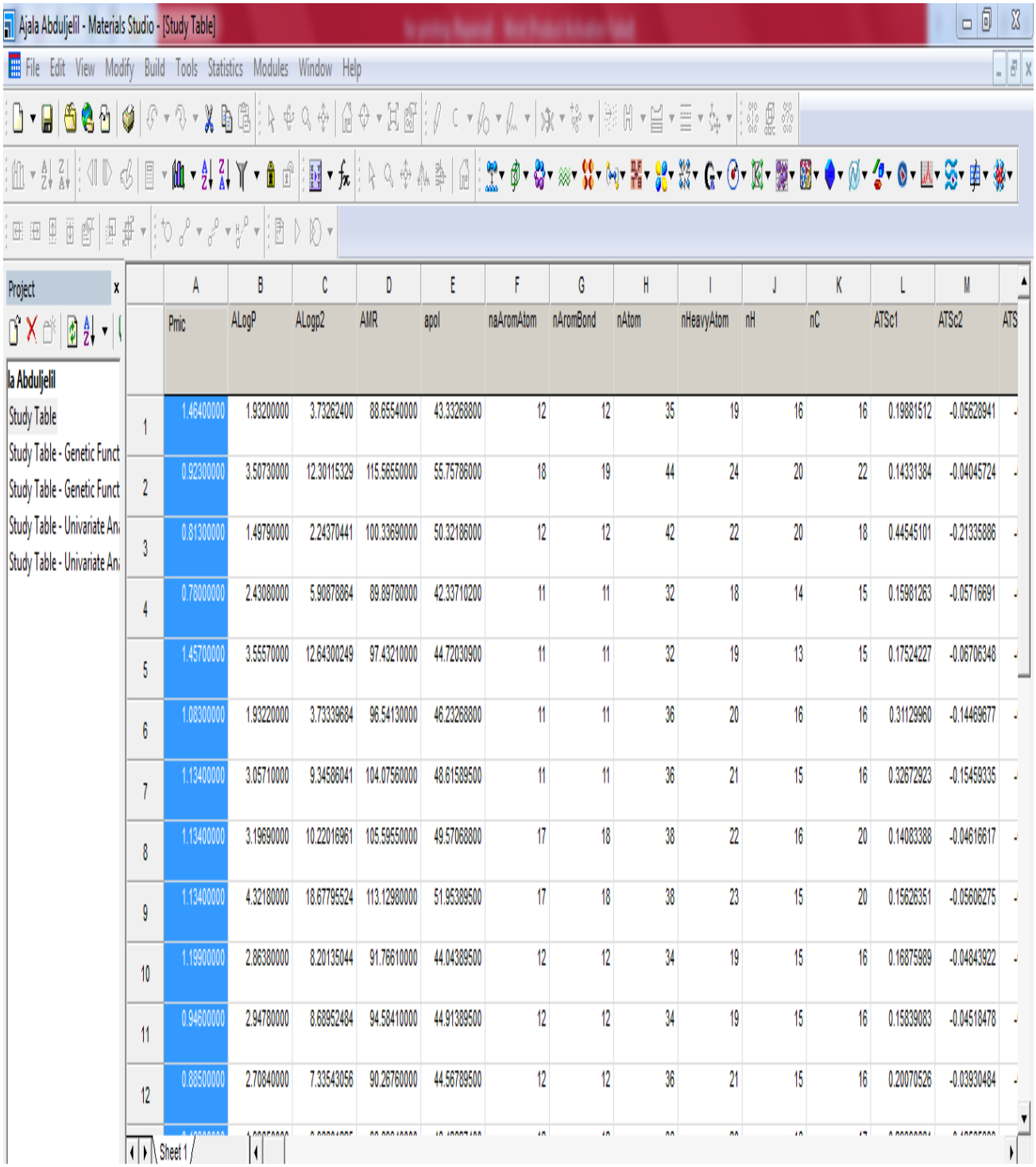
(iii) Procedure for univariate analysis of pMIC values

The work space of material studio was open (Plate 3.17), the pMIC was high-lighted as shown in Plate 3.22 and the following was clicked from the menu tools **statistic → initial analysis→ univariate analysis** and resulted to Plate 3.23.

(iv) Procedure for determining r_0^2, r^2, r'^2_0, k and k'

The Microsoft excel package was launched creating its interface (Plate3.12). On the work sheet, the data for predicted and observed activities were entered into two adjacent columns. Through the **Insert → Scatter** menus on the interface, a reduced version of the graph appears in a prop up window shown in Plate 3.24. By clicking **+** symbol → **Chart elements** at the top right of the window, *Trend line*, *axis title*, *Chart title* were added to the graph by marking the appropriate boxes. Through **Trend line** box → **more options→ Set intercept box**, the graph was either made to have intercept by leaving the intercept box unchecked or otherwise by setting the intercept to 0.0 (origin). The Parameters r^2 and r_0^2 are the squared correlation coefficients between the observed and

predicted values of the compounds with and without intercept, respectively. The parameter $r'_0{}^2$ bears the same meaning with $r_0{}^2$ but uses the reversed axes. k is the slope



	A	B	C	D	E	F	G	H	I	J	K	L	M
	Pmic	ALogP	ALogp2	AMR	apol	naAromAtom	nAromBond	nAtom	nHeavyAtom	nH	nC	ATSc1	ATSc2
1	1.46400000	1.93200000	3.73262400	88.65540000	43.33268800	12	12	35	19	16	16	0.19881512	-0.05628941
2	0.92300000	3.50730000	12.30115329	115.56550000	55.75786000	18	19	44	24	20	22	0.14331384	-0.04045724
3	0.81300000	1.49790000	2.24370441	100.33690000	50.32186000	12	12	42	22	20	18	0.44545101	-0.21335886
4	0.78000000	2.43080000	5.90878864	89.89780000	42.33710200	11	11	32	18	14	15	0.15981263	-0.05716691
5	1.45700000	3.55570000	12.64300249	97.43210000	44.72030900	11	11	32	19	13	15	0.17524227	-0.06706348
6	1.08300000	1.93220000	3.73339684	96.54130000	46.23268800	11	11	36	20	16	16	0.31129960	-0.14469677
7	1.13400000	3.05710000	9.34586041	104.07560000	48.61589500	11	11	36	21	15	16	0.32672923	-0.15459335
8	1.13400000	3.19690000	10.22016961	105.59550000	49.57068800	17	18	38	22	16	20	0.14083388	-0.04616617
9	1.13400000	4.32180000	18.67795524	113.12980000	51.95389500	17	18	38	23	15	20	0.15626351	-0.05606275
10	1.19900000	2.86380000	8.20135044	91.76610000	44.04389500	12	12	34	19	15	16	0.16875989	-0.04843922
11	0.94600000	2.94780000	8.68952484	94.58410000	44.91389500	12	12	34	19	15	16	0.15839083	-0.04518478
12	0.88500000	2.70840000	7.33543056	90.26760000	44.56789500	12	12	36	21	15	16	0.20070526	-0.03930484

Plate 3.22: High lighted pMIC work space of material studio

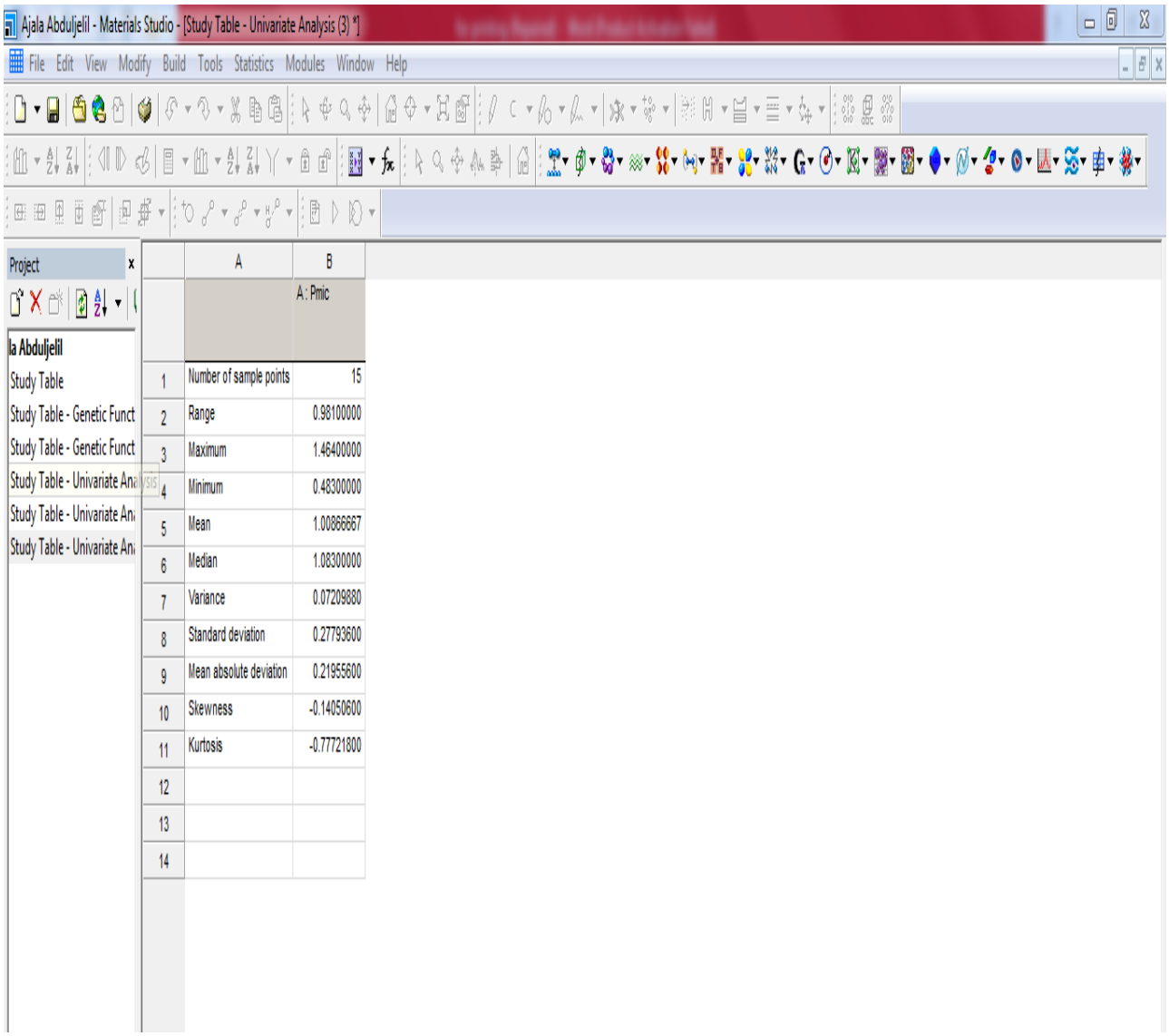


Plate 3.23: univariate analysis study table.

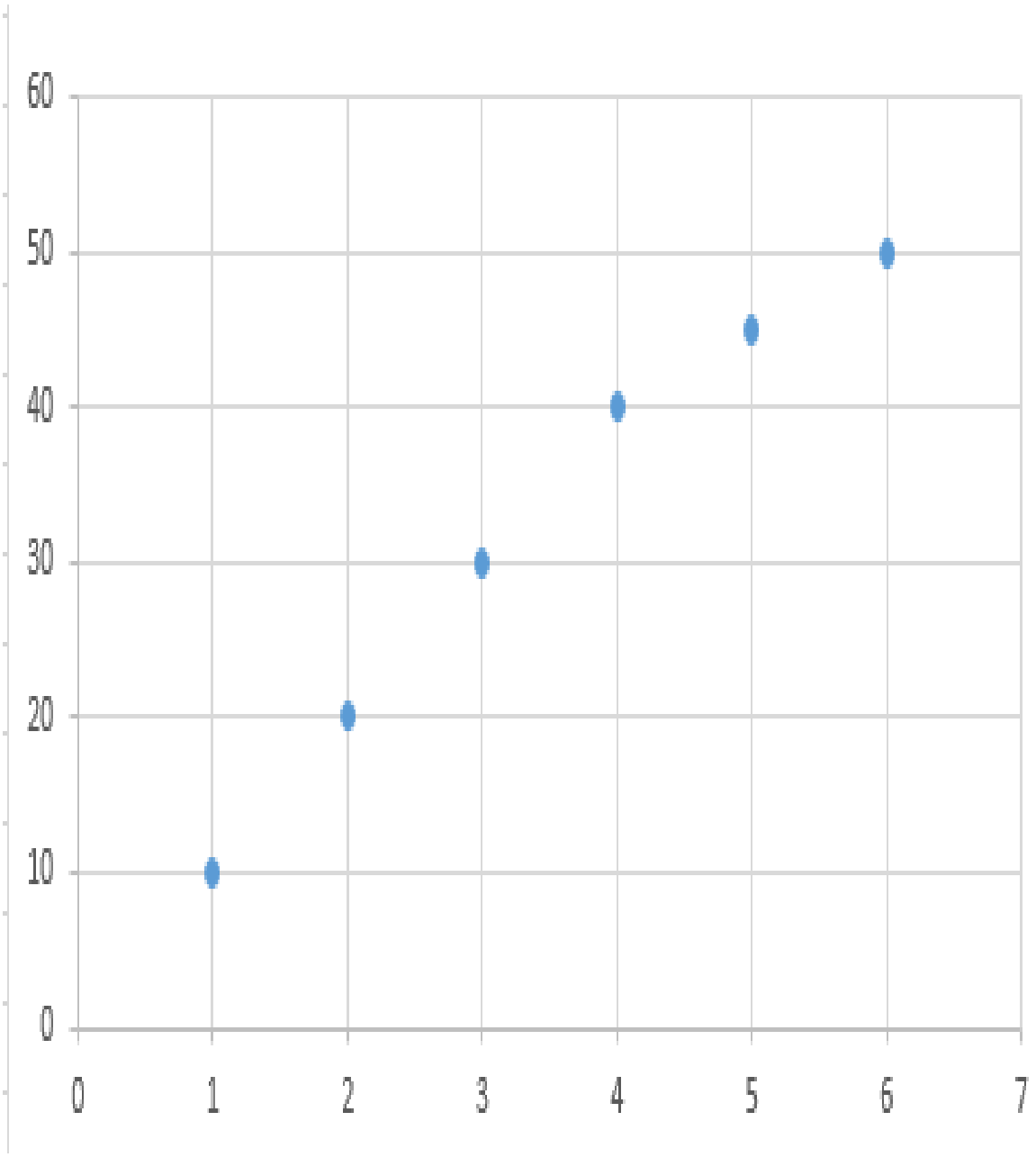


Plate 3.24: Microsoft excel Chart window

of the plot of the observed and predicted values of the compounds via the origin and k' the slope of the reversed axes (Alexander *et al*., 2015).

(v) Procedure for determining residual plots

Data was clicked from the down side of Plate 3.21 and lead to column showing actual, predicted and residual values for each of the generated model as shown in Plate 3.25, the residual and predicted columns were alighted, copied and paste on the excel interface (Plate 3.12). The following procedures were carried out, residual and predicted columns were alighted → **DATA** → **Data Analysis** → **Data analysis box** (Plate 3.13) → **Regression** → **ok** → **Regression box** (Plate 3.14), Residual column was alighted for input Y range while predicted column was alighted for input X range, residual plots box was checked and this resulted to residual plots of Figure 4,5, 4.6, 4.7 and 4.8.

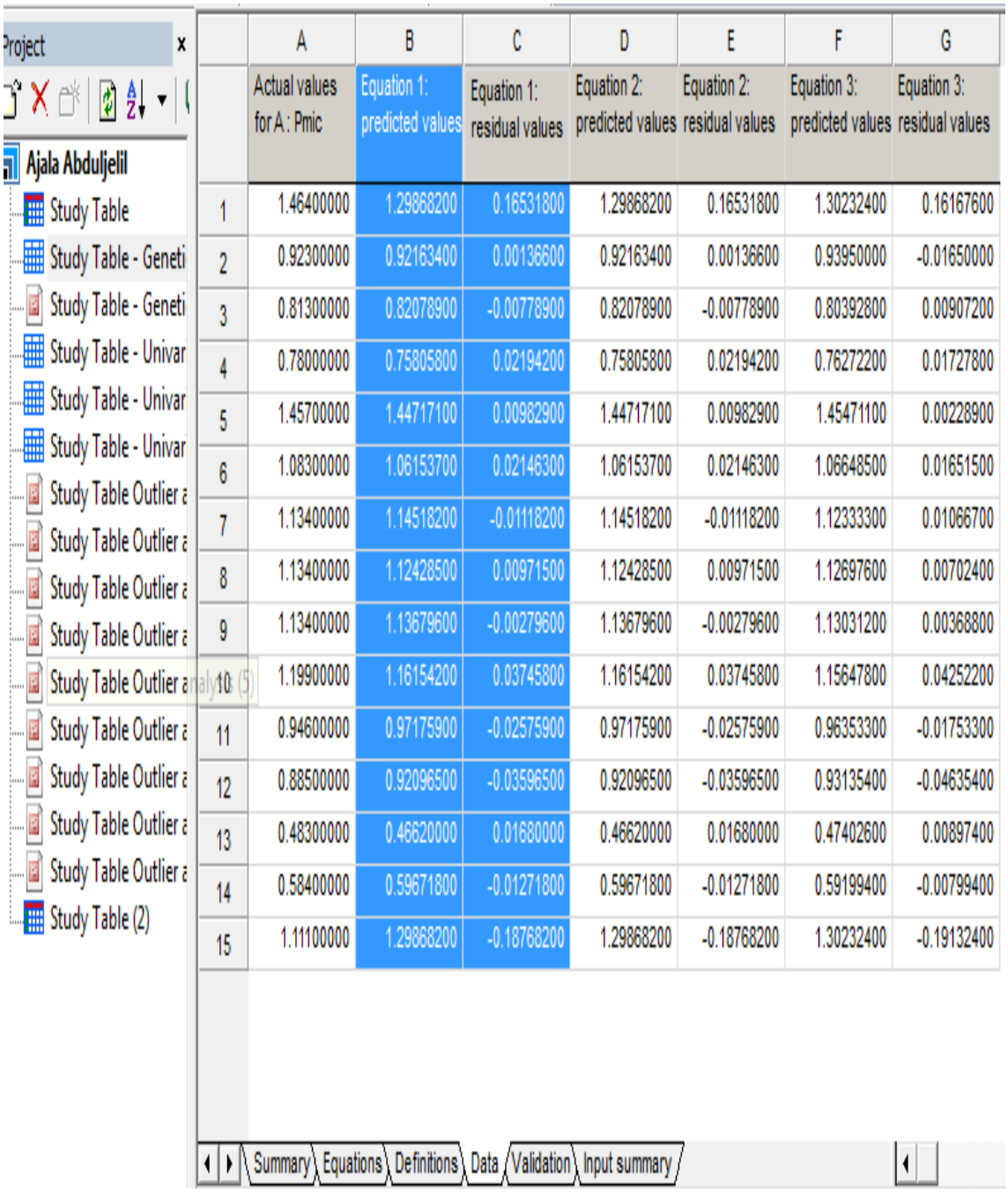
(vi) Procedure for determining outlier analysis of models

The work space of material studio was open and pMIC was high-lighted (Plate 3.17 and Plate 3.22) and the following was clicked from the menu tools **statistic** → **model building** → **outlier analysis** and resulted to Plate 3.26, the option box was clicked in Plate 3.26 and lead to Plate 3.27, in the cutoff box, four (4) was inputed as value and checked, also scale residual by standard deviation and plot square root of residual values was also checked resulted to Plate 3.28.

3.3 Multicollinearity Among Descriptors

The multi-collinearity between the descriptors used in the model can be detected by calculating their variance inflation factors (VIF), which can be calculated as follows:

$$VIF = {}^{1}/_{(1-R^2)} \quad 3.8$$



	A	B	C	D	E	F	G
	Actual values for A: Pmic	Equation 1: predicted values	Equation 1: residual values	Equation 2: predicted values	Equation 2: residual values	Equation 3: predicted values	Equation 3: residual values
1	1.46400000	1.29868200	0.16531800	1.29868200	0.16531800	1.30232400	0.16167600
2	0.92300000	0.92163400	0.00136600	0.92163400	0.00136600	0.93950000	-0.01650000
3	0.81300000	0.82078900	-0.00778900	0.82078900	-0.00778900	0.80392800	0.00907200
4	0.78000000	0.75805800	0.02194200	0.75805800	0.02194200	0.76272200	0.01727800
5	1.45700000	1.44717100	0.00982900	1.44717100	0.00982900	1.45471100	0.00228900
6	1.08300000	1.06153700	0.02146300	1.06153700	0.02146300	1.06648500	0.01651500
7	1.13400000	1.14518200	-0.01118200	1.14518200	-0.01118200	1.12333300	0.01066700
8	1.13400000	1.12428500	0.00971500	1.12428500	0.00971500	1.12697600	0.00702400
9	1.13400000	1.13679600	-0.00279600	1.13679600	-0.00279600	1.13031200	0.00368800
10	1.19900000	1.16154200	0.03745800	1.16154200	0.03745800	1.15647800	0.04252200
11	0.94600000	0.97175900	-0.02575900	0.97175900	-0.02575900	0.96353300	-0.01753300
12	0.88500000	0.92096500	-0.03596500	0.92096500	-0.03596500	0.93135400	-0.04635400
13	0.48300000	0.46620000	0.01680000	0.46620000	0.01680000	0.47402600	0.00897400
14	0.58400000	0.59671800	-0.01271800	0.59671800	-0.01271800	0.59199400	-0.00799400
15	1.11100000	1.29868200	-0.18768200	1.29868200	-0.18768200	1.30232400	-0.19132400

Plate 3.25 : Column of actual, predicted and residual values for the developed model.

Plate 3.26: Outlier analysis checkbox

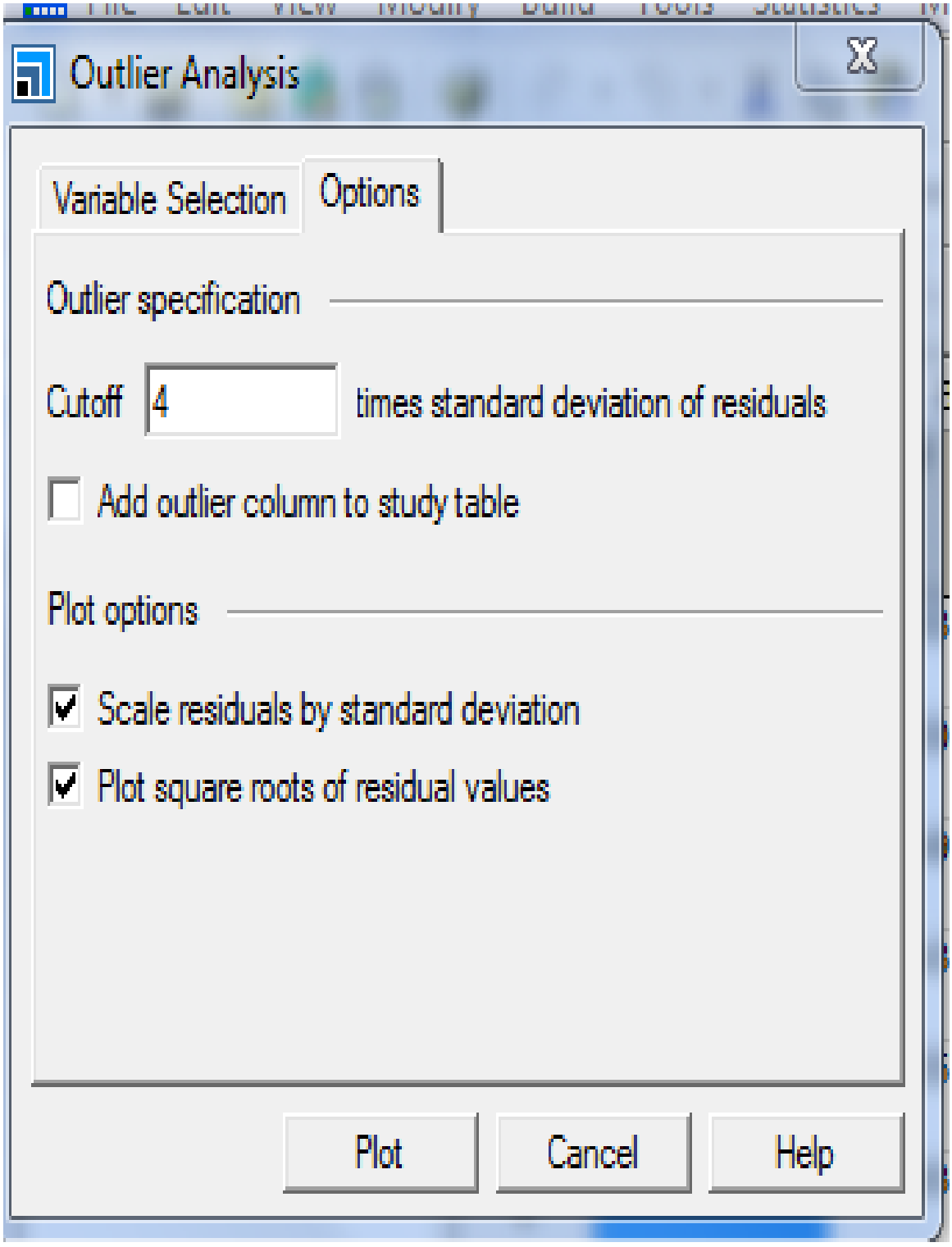


Plate 3.27: Outlier analysis option box

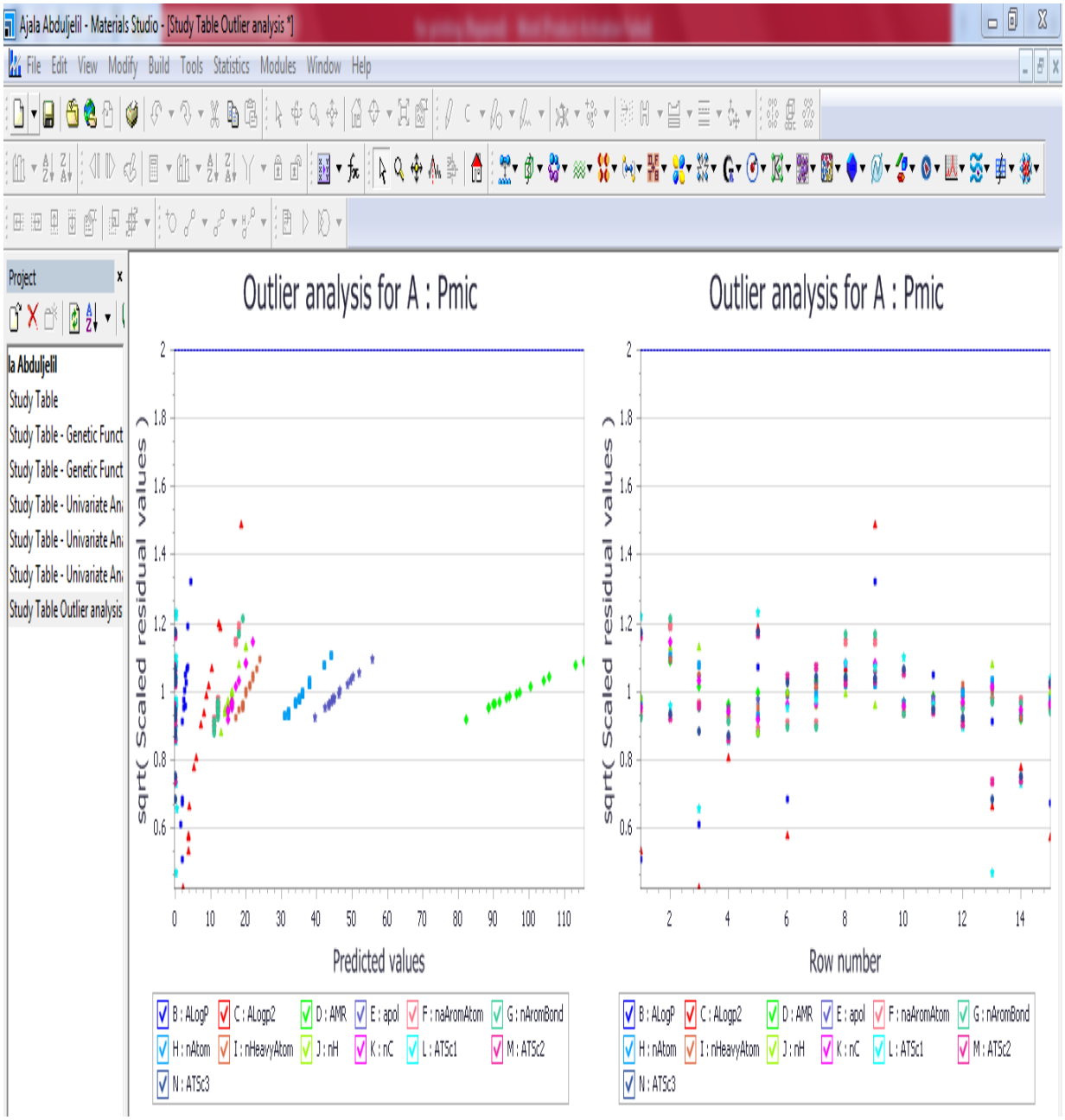


Plate 3.28: Outlier analysis graph

Where R^2 is the correlation coefficient of the multiple regression between the variables within the model. If VIF equals to 1, then no inter-correlation exists for each variable; if VIF falls into the range of 1–5, the related model is acceptable; and if VIF is larger than 10, the related model is unstable and a recheck is necessary (Jaiswal *et al*., 2004; Shapiro *et al*., 1998).

3.4 Applicability Domain (AD)

Not even a robust, significant, and validated QSAR model can be expected to reliably predict the modelled property for the entire universe of chemicals. In fact, only

predictions for new molecules falling within the model domain can be considered reliable like those for training chemicals and not extrapolations of the model. The applicability domain is a theoretical region in physicochemical space (the response and chemical structure space) for which a QSAR model should make predictions with a given reliability. This region was defined by the nature of the chemicals in the training set, and can be characterized in various ways (Netzeva *et al*., 2005). While the range of the independent variable X (descriptor) is useful to define the chemical domain of a univariate QSAR (based on one descriptor), in Multiple linear regression, where the data and residual distribution is generally normal, the predicted data Y are obtained from the experimental data Y through the Hat matrix of influence.

$$\hat{Y} = \boldsymbol{X(X^T X)^{-1} X^T y = H_y} \qquad \textbf{3.9}$$

where X is the descriptor matrix.

A simple measure of a molecule being too far from the applicability domain of the model is its leverage in the original variable space, h_{ii}, which is defined as:

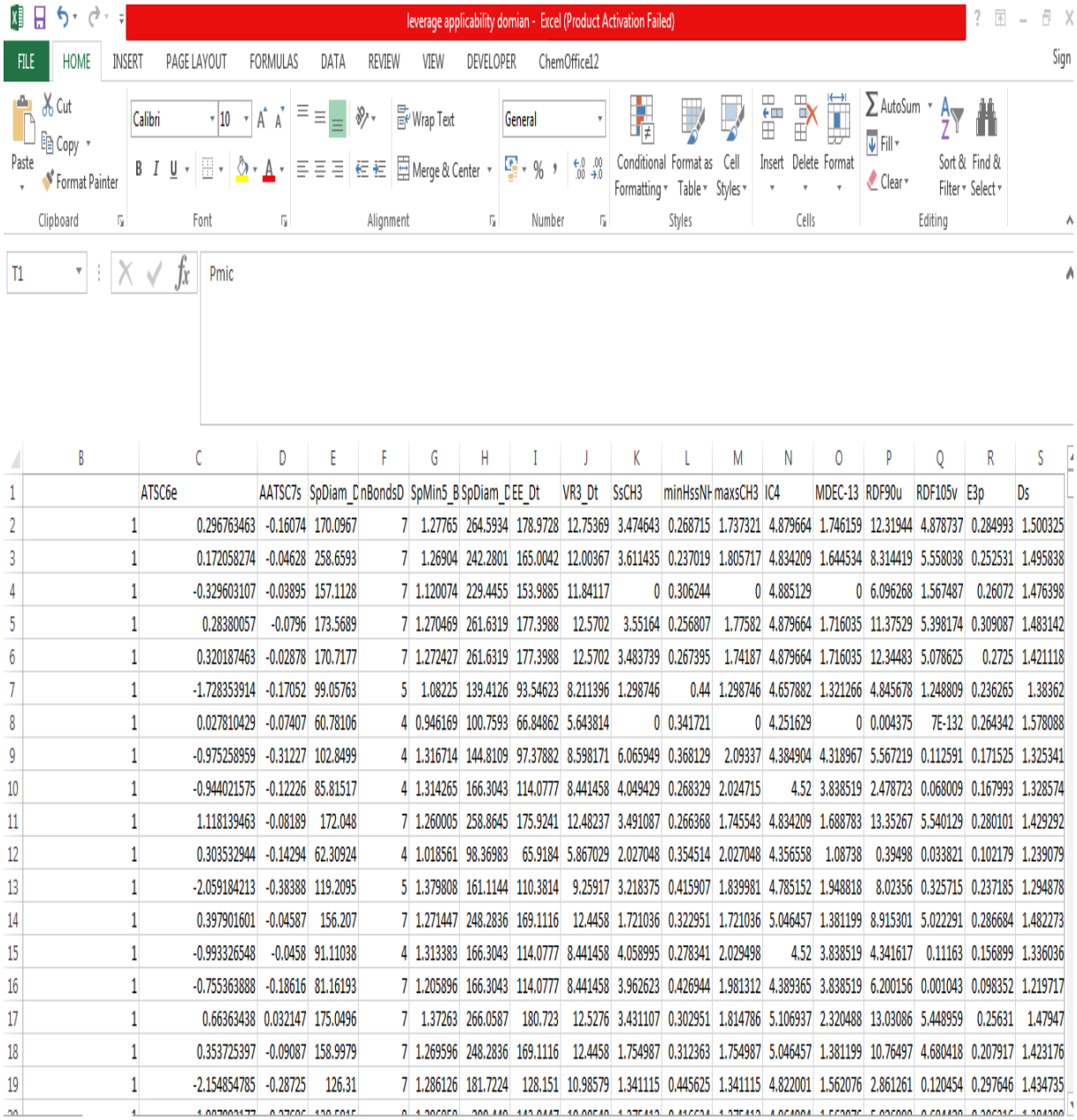
$$\boldsymbol{\mathfrak{h}_{ii} = x_I^T (X^T X)^{-1} \qquad x_i (i = 1 \ldots\ldots n)} \qquad \textbf{3.10}$$

where x_i is the descriptor row-vector of the query compound, and X is the n by p matrix of p model descriptor values for n training set compounds. The superscript T refers to the transpose of the matrix/vector. The ii main diagonal entry of the Hat matrix (H, (ℏii)) provides a measure of how far observation i is from the center of the X data (leverage). The warning leverage ℏ* is generally fixed at 3k/n, where n is the number of training compounds, and k the number of model parameters plus one (p + 1). A molecule with high leverage in the training greatly influences the regression: the fitted regression line is forced near the observed value and the residuals are small, thus the molecule in the training set is not an outlier for the response fitting. To visualize the complete AD of a QSAR model (both for structure and response), the plot of standardized crossvalidated residuals against leverage (Hat diagonal) values (ℏ) (the Williams plot) was used for an immediate and simple graphical detection of both the response outliers (i.e., compounds with cross-validated standardized residuals greater than 2.5-3 standard deviation units, >2.5–3s) and structurally influential molecules in a model (ℏ>ℏ*).

3.4.1 Procedure for computing applicability domain of William's plot

This was done using the Microsft Office Excel 2007. The descriptors matrix of the ones that appeared in the model was copied and pasted on an already opened excel worksheet as shown in Plate 3.29. This was done by clicking on the right button on the key pad of the laptop → *clicking* **Paste special** to bring up a drop down menu as shown in Plate 3.30. Transpose option was checked and **ok** was selected to give the transpose of the original matrix. The transposed version (X^T_{tr} = m x n) of the original descriptor matrix (X_{tr}) generated was multiplied by the original matrix 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



	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1		ATSC6e	AATSC7s	SpDiam_D	nBondsD	SpMin5_B	SpDiam_D	EE_Dt	VR3_Dt	SsCH3	minHssNH	maxsCH3	IC4	MDEC-13	RDF90u	RDF105v	E3p	Ds
2	1	0.296763463	-0.16074	170.0967	7	1.27765	264.5934	178.9728	12.75369	3.474643	0.268715	1.737321	4.879664	1.746159	12.31944	4.878737	0.284993	1.500325
3	1	0.172058274	-0.04628	258.6593	7	1.26904	242.2801	165.0042	12.00367	3.611435	0.237019	1.805717	4.834209	1.644534	8.314419	5.558038	0.252531	1.495838
4	1	-0.329603107	-0.03895	157.1128	7	1.120074	229.4455	153.9885	11.84117	0	0.306244	0	4.885129	0	6.096268	1.567487	0.26072	1.476398
5	1	0.28380057	-0.0796	173.5689	7	1.270469	261.6319	177.3988	12.5702	3.55164	0.256807	1.77582	4.879664	1.716035	11.37529	5.398174	0.309087	1.483142
6	1	0.320187463	-0.02878	170.7177	7	1.272427	261.6319	177.3988	12.5702	3.483739	0.267395	1.74187	4.879664	1.716035	12.34483	5.078625	0.2725	1.421118
7	1	-1.728353914	-0.17052	99.05763	5	1.08225	139.4126	93.54623	8.211396	1.298746	0.44	1.298746	4.657882	1.321266	4.845678	1.248809	0.236265	1.38362
8	1	0.027810429	-0.07407	60.78106	4	0.946169	100.7593	66.84862	5.643814	0	0.341721	0	4.251629	0	0.004375	7E-132	0.264342	1.578088
9	1	-0.975258959	-0.31227	102.8499	4	1.316714	144.8109	97.37882	8.598171	6.065949	0.368129	2.09337	4.384904	4.318967	5.567219	0.112591	0.171525	1.325341
10	1	-0.944021575	-0.12226	85.81517	4	1.314265	166.3043	114.0777	8.441458	4.049429	0.268329	2.024715	4.52	3.838519	2.478723	0.068009	0.167993	1.328574
11	1	1.118139463	-0.08189	172.048	7	1.260005	258.8645	175.9241	12.48237	3.491087	0.266368	1.745543	4.834209	1.688783	13.35267	5.540129	0.280101	1.429292
12	1	0.303532944	-0.14294	62.30924	4	1.018561	98.36983	65.9184	5.867029	2.027048	0.354514	2.027048	4.356558	1.08738	0.39498	0.033821	0.102179	1.239079
13	1	-2.059184213	-0.38388	119.2095	5	1.379808	161.1144	110.3814	9.25917	3.218375	0.415907	1.839981	4.785152	1.948818	8.02356	0.325715	0.237185	1.294878
14	1	0.397901601	-0.04587	156.207	7	1.271447	248.2836	169.1116	12.4458	1.721036	0.322951	1.721036	5.046457	1.381199	8.915301	5.022291	0.286684	1.482273
15	1	-0.993326548	-0.0458	91.11038	4	1.313383	166.3043	114.0777	8.441458	4.058995	0.278341	2.029498	4.52	3.838519	4.341617	0.11163	0.156899	1.336036
16	1	-0.755363888	-0.18616	81.16193	7	1.205896	166.3043	114.0777	8.441458	3.962623	0.426944	1.981312	4.389365	3.838519	6.200156	0.001043	0.098352	1.219717
17	1	0.66363438	0.032147	175.0496	7	1.37263	266.0587	180.723	12.5276	3.431107	0.302951	1.814786	5.106937	2.320488	13.03086	5.448959	0.25631	1.47947
18	1	0.353725397	-0.09087	158.9979	7	1.269596	248.2836	169.1116	12.4458	1.754987	0.312363	1.754987	5.046457	1.381199	10.76497	4.680418	0.207917	1.423176
19	1	-2.154854785	-0.28725	126.31	7	1.286126	181.7224	128.151	10.98579	1.341115	0.445625	1.341115	4.822001	1.562076	2.861261	0.120454	0.297646	1.434735

Plate 3.29: Excel Worksheet for computing applicability domain

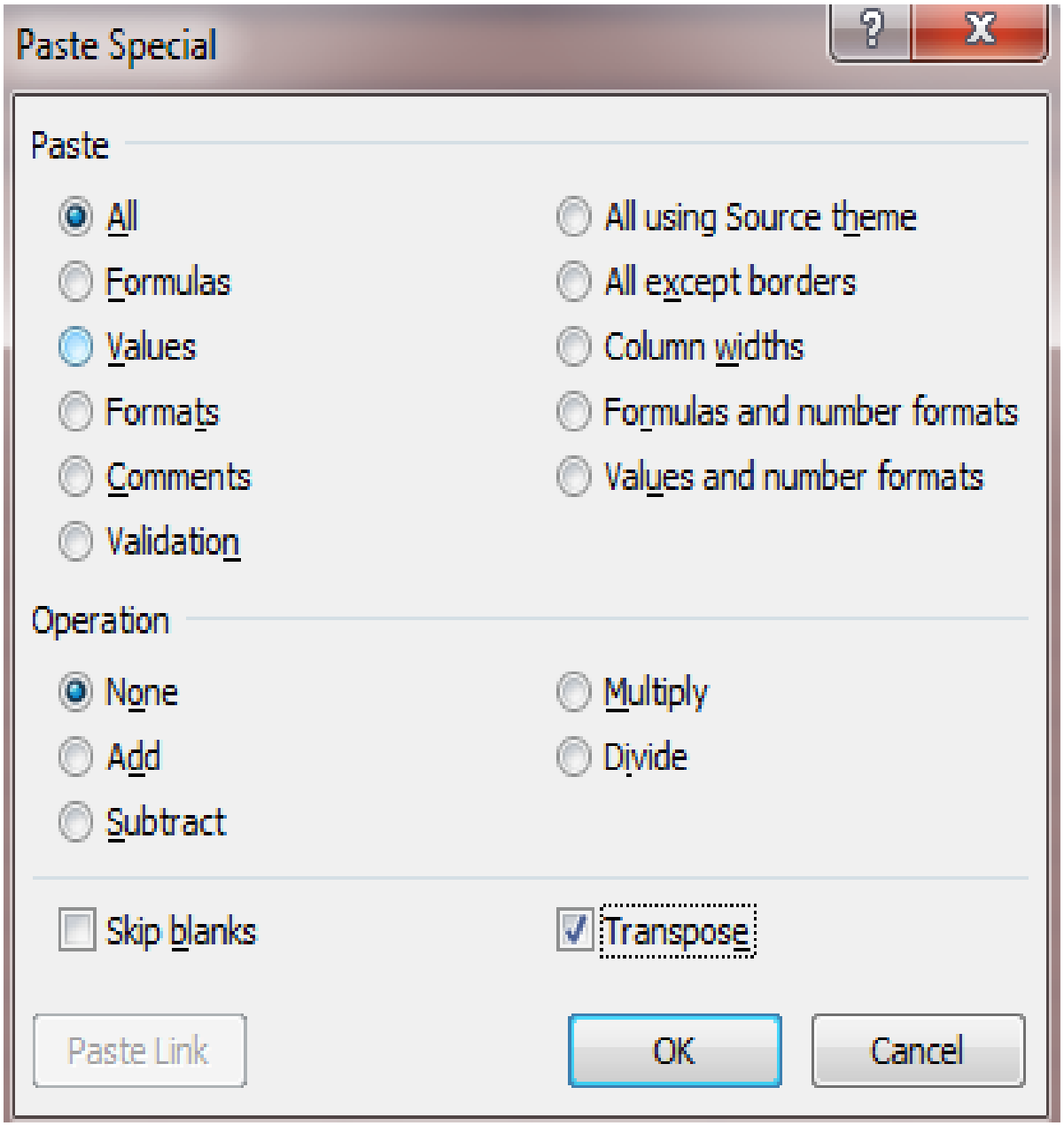


Plate 3.30: Paste special Dialogue Box

number of rows equal to that of the number of rows of the transposed matrix. In the first cell of the highlighted section MMULT (array1, array2) was typed. Array 1 is the range of cells covered by the original matrix on the worksheet while array2 is the range of cells covered by the transpose. These ranges were selected by highlighting the sections covered by the respective matrices on the work sheet matrix as shown in Plate 3.31. The CTRL+ SHIFT + ENTER was pressed to give the result of the multiplication. The inverse of the result of the multiplication was found by first highlighting a section on the worksheet where the result of the operation was displayed, the number of rows and columns on this section was the same as that of the matrix whose inverse was being sought. In the first cell of the highlighted section (Plate 3.32) =Transpose (Range) was typed, 'range' is the range of cells covered by the matrix to be transposed e.g. A1:M7. CTRL+ SHIFT + ENTER was pressed to give the result of the operation. The original X_{tr}, the inverse $(X^{T}_{tr}*X_{tr})^{-1}$ and the transposed matrix X^{T} were multiplied using the same multiplication steps described above to obtain an $n \times n$ symmetric matrix (hat matrix) whose diagonal elements (leverages) describe the distance of each training set compound to the structural centroid of the model as shown in Plate 3.33, the leverages were extracted from the hat matrix. The calculation of the hat matrix as well as the extraction of the diagonal elements (leverages) for the training set compound entering the model, were performed. As for the test/external validation set, the steps described above were also carried out 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 was also generated for the test / validation set as shown in Plate 3.34.

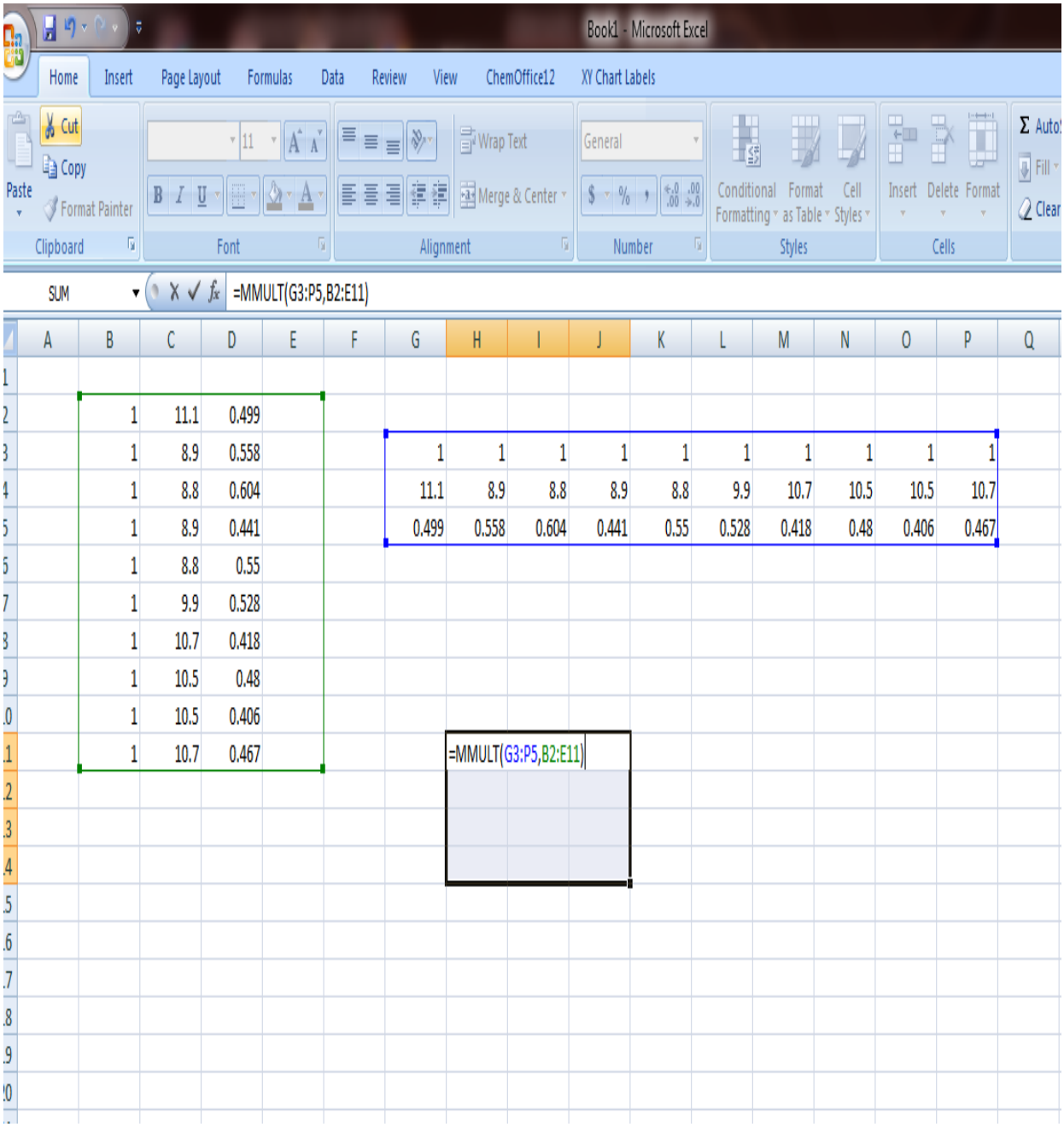


Plate 3.31: Excel sheet for computing leverages

Plate 3.32: Excel Worksheet for leverage procedures

leverage applicability domain - Excel (Product Activation Failed)

INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW DEVELOPER ChemOffice2

Calibri 11 Wrap Text General Merge & Center Conditional Formatting Format as Table Cell Styles Insert Delete Format AutoSum Fill Clear Sort & Filter Find & Select

Font Alignment Number Styles Cells Editing

THE HAT MATRIX OF TRAINING SET

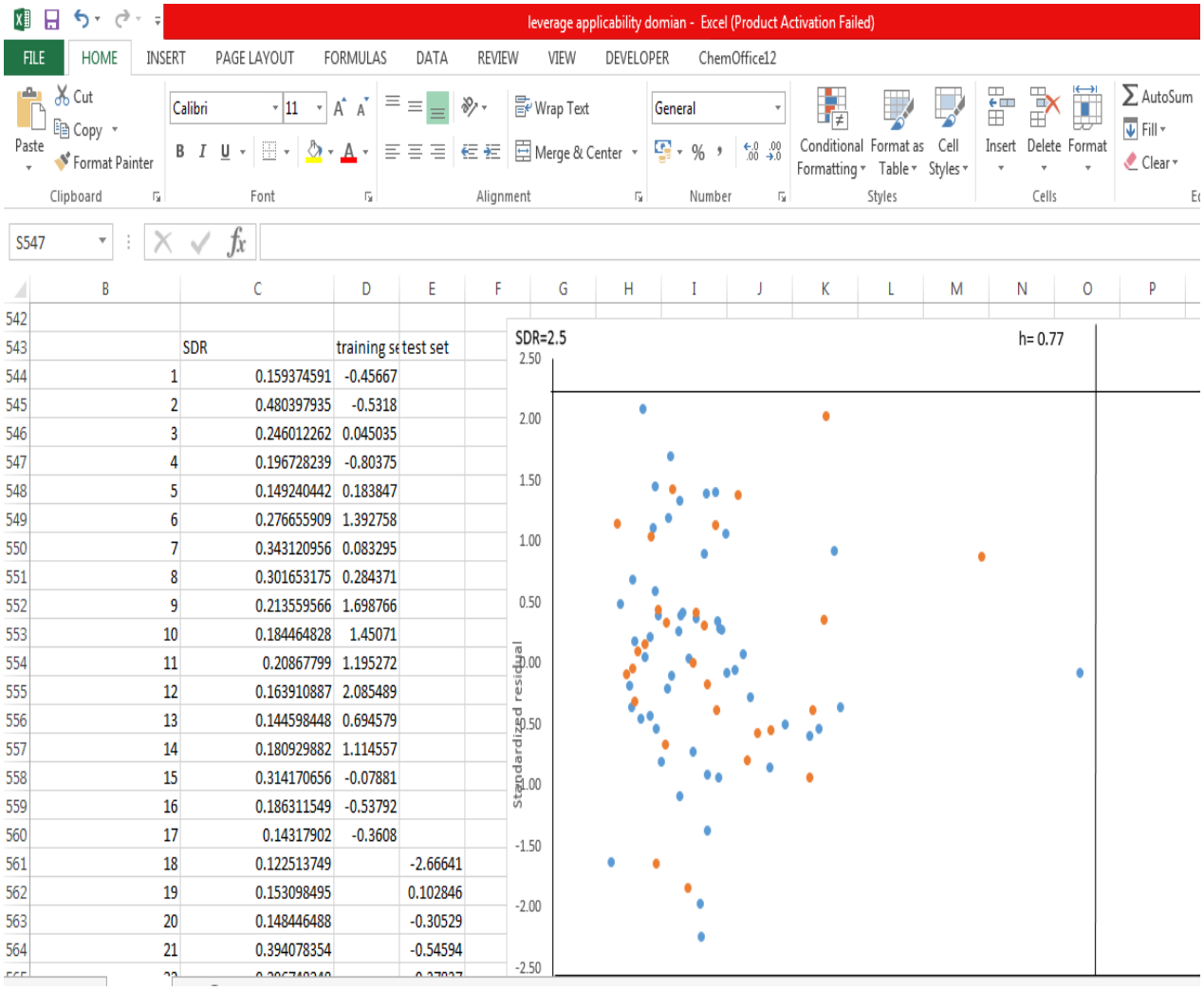
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	1
1	0.159374591	0.069113291	0.087043	0.163194	0.130679	0.017085	0.076699	0.030409	0.034265	0.115809	-0.00166	0.018728	0.061433	0.009818	0.02357
2		0.480397935	0.037856	0.09226	0.066174	-0.02317	0.007835	0.114072	-0.00502	0.062946	0.046519	0.036081	-0.01276	-0.00756	-0.077
3			0.246012	0.073652	0.058839	0.058968	0.151283	0.026721	0.058656	0.009786	0.022445	0.015162	0.005183	0.037617	0.0373
4				0.196728	0.160486	0.014085	0.057084	-0.00345	0.018375	0.147054	-0.01822	0.005921	0.076568	-0.01082	-0.0104
5					0.14924	0.004203	0.008611	9.73E-05	0.007648	0.13988	-0.0023	0.00084	0.060812	-0.00502	0.02871
6						0.276656	0.073963	0.017411	0.002506	-0.05352	0.061112	0.079105	0.04947	0.013191	0.01223
7							0.343121	0.030051	0.018582	0.004127	0.01066	-0.02186	0.030666	0.001619	-0.051
8								0.301653	0.074111	-0.01426	0.035745	0.115815	-0.07622	0.078967	0.09846
9									0.21356	-0.04462	0.041005	0.015769	0.004886	0.185801	0.07575
10										0.184465	-0.02392	-0.03107	0.076694	-0.05164	0.00870
11											0.208678	0.045754	-0.00516	0.038344	0.07911
12												0.169311	-0.03194	0.006236	-0.0187
13													0.144598	-0.00369	-0.0354
14														0.18093	0.09418
1															0.31417

Plate 3.33 Hat matrix of training set

Plate 3.34 Hat matrix of test set

the original $X_{te/ext}$, the inverse $(X^{T}_{te/ext}*X_{te/est})^{-1}$ and the transposed matrix $X^{T}_{te/ext}$ were multiplied. The leverages from both training (blue colour) and test sets hat matrices (Plate 3.33 and 3.34 respectively) were plotted as a function of the standardized cross validated residuals (SDR) to obtain the Williams plot.

A plot of the standardized residual (SDR) against the extracted leverages (Hat diagonal) values (h) were made as shown in Plate 3.35. On that plot along the abscissa the cut-off leverage value was made the boundary. And along the ordinate ± 2.5 or at most ± 3 standard deviation unit was used as boundary because point that lie within ± 3 standardised residual from the mean covers ninety nine percent of the normally distributed data . Any compounds with cross-validated standard residual larger than 2.5 standard deviation unit were considered as outlier.



		SDR	training se	test set
544	1	0.159374591	-0.45667	
545	2	0.480397935	-0.5318	
546	3	0.246012262	0.045035	
547	4	0.196728239	-0.80375	
548	5	0.149240442	0.183847	
549	6	0.276655909	1.392758	
550	7	0.343120956	0.083295	
551	8	0.301653175	0.284371	
552	9	0.213559566	1.698766	
553	10	0.184464828	1.45071	
554	11	0.20867799	1.195272	
555	12	0.163910887	2.085489	
556	13	0.144598448	0.694579	
557	14	0.180929882	1.114557	
558	15	0.314170656	-0.07881	
559	16	0.186311549	-0.53792	
560	17	0.14317902	-0.3608	
561	18	0.122513749		-2.66641
562	19	0.153098495		0.102846
563	20	0.148446488		-0.30529
564	21	0.394078354		-0.54594

Plate 3.35: The William's Plot of Applicability Domain

CHAPTER FOUR

4.0 RESULTS

4.1 Results of Molecular Optimization

The following properties were obtained from the optimized structures: Molecular properties, QSAR descriptors, thermodynamic properties as well as acidity and basicity properties.

4.2 Descriptor Calculation

Padel descriptor tool kit, Spartan 14 and ChmBio3D Pro 12.0.1V softwares were used to calculate 1,537 descriptors ranging from 0D, 1D, 2D, and 3D.

4.3 GFA Derived Models for MIC of Anti-*Candida albicans* and *Epidermophyton floccosum* Compounds

Models 1, 2, 3 (set 1: Chalcone derivatives), Models 4, 5, 6 (set 2: Coumarin and Chalcone derivatives), Models 7, 8, 9 (set 3: Chalcones derivatives) and Models 10, 11, and 12 (set 4: Neolignans derivatives) gaves the linear models for pMIC of anti-*Candida albicans* and *Epidermophyton floccosum* organism respectively, using multiple linear regression and genetic function approximation.

The Genetic Function Approximation (GFA) led to the derivation of twelve models. Models 1, 4, 7 and 10 were chosen as the best models for predicting the pMIC of anti-*Candida albicans* and *Epidermophyton floccosum* molecules based on the statistical criterion. The internal and external validation parameters of the models conform to the minimum standard for a robust QSAR model (see section 3.2.7), confirming the stability and robustness of the models.

Model 1:

$$pMIC = -0.1015\, mw - 0.007566\, caa - 0.98076\, cLogp - 55.9669\, mr + 0.02751\, mv + 141.3109\, p + 8.76211$$

n = 20, R^2 = 0.9864, R^2_{adj} = 0.9762, F-value = 96.62, Friedman LOF = 0.0231, R^2_{pred} = 0.986 and R^2_{cv} = 0.928

Model 2

$$pMIC = -3.4824\, sc5 - 2.1052\, mdec11 + 1.3502\, weta.volume - 0.0158\, psa - 5.2359H + 3.1779$$

n = 20, R^2 = 0.9134,R^2_{adj} = 0.8825, F-value = 29.53, Friedman LOF = 0.076, R^2_{pred} = 0.669 and R^2_{cv} = 0.7838.

Model 3

$$pMIC = -3.999\, sc5 - 0.2194\, mdecI3 + 1{,}3789\, weta.volume - 0.0171\, psa - 5.2759\, Gibbs + 3.1149$$

n = 15, R^2 = 0.8859,R^2_{adj} = 0.8452, F-value = 21.74, Friedman LOF = 0.101, R^2_{pred} = 0.338 and R^2_{cv} = 0.7576

Model 4

$$pMIC = 0.04703\, SHBint5 + 3.017924\, WD.unity + 0.05263\, wlambda2.eneg + 0.01986 wlambda3.polar - 1.1356$$

n= 28, R^2 = 0.9064, R^2_{adj} = 0.8830, F-value = 38.749, Friedman LOF = 0.0502, R^2_{pred} = 0.8722 and R^2_{cv} = 0.7321

Model 5

$$pMIC = 0.6703\ MLFER_A + 2.7097 WD.unity - 0.0519 wlambda2.eneg + 0.01729 wlambda3.polar - 0.7781$$

n= 28, R^2 = 0.9022, R^2_{adj} = 0.8778, F-value = 36.92, Friedman LOF = 0.0522, R^2_{pred} = 0.5271 and R^2_{cv} = 0.5170

Model 6

$$pMIC = 1.2252\ minaaCH - 2.4607 MDEO12 + 0.0017 DPSA1 - 0.00477\ solv\ energy + 1.7573$$

n= 28, R^2 = 0.8983, R^2_{adj} = 0.8728, F-value = 35.32, Friedman LOF = 0.0546, R^2_{pred} = -0.557and R^2_{cv} = 0.7871

Model 7

$$pMIC = -0.1052\ BCUTw11 + 0.006425 PNSA3 + 0.0743\ MOMI_{XZ} - 0.07419\ psa + 0.00978 \max elpot - 0.604521 cLogp + 3.5836$$

n= 32, R^2 = 0.8085, R^2_{adj} = 0.7625, F-value = 17.5862, Friedman LOF = 0.7321, R^2_{pred} = 0.7941 and R^2_{cv} = 0.6578

Model 8

$$pMIC = -0.123 BCUTw_{11} + 0.006374 PNSA_3 - 1.5864 Logp - 0.003887\ mol\ wt - 0.201508\ psa - 4.2385 G$$

n= 32, R^2 = 0.8048, R^2_{adj} = 0.7579, F-value = 17.176, Friedman LOF = 0.1903, R^2_{pred} = 0.4912 and R^2_{cv} = 0.6643

Model 9

$$pMIC = -0.1259BCUTw_{11} + 0.006429PNSA_3 - 1.6496Logp - 0.00348\ mol\ wt$$
$$= 0.01970\ psa - 4.1929\ H + 6.5124$$

n= 32, R^2 = 0.8036, R^2_{adj} = 0.7565, F-value = 17.049, Friedman LOF = 0.1915, R^2_{pred} = 0.4002 and R^2_{cv} = 0.666

Model 10

pMIC = -0.05424se – 0.003097mep + 23.45799H – 23.455G^o + 1.139116

n= 13, R^2 = 0.942815, R^2_{adj} = 0.8954, F-value = 20.5407, Friedman LOF = 0.01523, R^2_{pred} = 0.7202 and R^2_{cv} = 0.8196

Model 11

pMIC = -0.0451sol energy–0.002197max elpot+21.4542H – 21.4521G+ 0.001568energy change +0.2052

n= 13, R^2 = 0.9215, R^2_{adj} = 0.8766, F-value = 19.02, Friedman LOF = 0.0161, R^2_{pred} = 0.6251 and R^2_{cv} = 0.7779

Model 12

pMIC =-0.0466sol energy+0.002408maxelpot+21.2891H–21.2869G + 0.00087energy (lumo)+1.0585

n= 13, R^2 = 0.9345, R^2_{adj} = 0.8805, F-value = 17.2127, Friedman LOF = 0.6174, R^2_{pred} = 0.5905 and R^2_{cv} = 0.7933

4.3.1: Univariate analysis of the pMIC value of models 1, 4, 7 and 10

A univariate analysis was performed on the data in the training set of Tables 3.2, 3.3, 3.4 (a and b) and 3.5 (a, b and c). The results were presented in Tables 4.1, 4.2, 4 3 and 4.4 as a tool to assess the quality of the data and its suitability for next statistical analysis.

4.3.2: Plot of Experimental against Predicted pMIC of models 1, 4, 7 and 10

Fig 4.1, 4.2, 4.3 and 4.4 showed the plot of the experimental against predicted activity of the models.

4.3.3: Residual plot of models 1, 4, 7 and 10

Fig 4.5, 4.6, 4.7 and 4.8 showed the plot of the residual against predicted activity of models.

4.3.4: Comparison of observed and predicted pMIC of models 1, 4, 7 and 10

The comparison of the predicted pMIC of the models with their experimental values were shown in Table : 4.5, 4.6, 4.7 and 4.8.

4.3.5: External validation of models 1, 4, 7 and 10

Table 4.9, 4.10, 4.11 and 4.12 with low residual values, show the predictive power of the models.

4.3.6: Variance Inflation Factor for the Descriptors in models 1, 4, 7 and 10

The multi-collinearity between the descriptors used in the models 1, 4 and 7 were checked by calculating their variance inflation factors (VIF), shown on Table 4.13, 4.14, 4.15 and 4.16 using equation 3.8

Table 4.1: Univariate analysis of the pMIC value of model 1

		pMIC
1	Number of sample points	15
2	Range	1.55
3	Maximum	4.72
4	Minimum	3.17
5	Mean	3.99
6	Median	4.09
7	Variance	0.16
8	Standard deviation	0.41
9	Mean absolute deviation	0.33
10	Skewness	-0.12
11	Kurtosis	-0.77

Table: 4.2: Univariate analysis of the pMIC value of model 4

		pMIC
1	Number of sample points	21
2	Range	1.03
3	Maximum	3.60
4	Minimum	2.57
	Mean	2.99
6	Median	2.89
7	Variance	0.09
8	Standard deviation	0.31
9	Mean absolute deviation	0.26
10	Skewness	0.23
11	Kurtosis	-1.28

Table 4.3: Univariate analysis of the pMIC value of model 7

		pMIC
1	Number of sample points	32
2	Range	1.65
3	Maximum	1.74
4	Minimum	0.09
5	Mean	0.89
6	Median	0.84
7	Variance	0.17
8	Standard deviation	0.41
9	Mean absolute deviation	0.33
10	Skewness	0.02
11	Kurtosis	-0.67

Table 4.4: Univariate analysis of the pMIC value of model 10

		pMIC
1	Number of sample points	13
2	Range	1.23
3	Maximum	1.38
4	Minimum	0.15
5	Mean	0.79
6	Median	0.71
7	Variance	0.09
8	Standard deviation	0.31
9	Mean absolute deviation	0.23
10	Skewness	-0.02
11	Kurtosis	-0.26

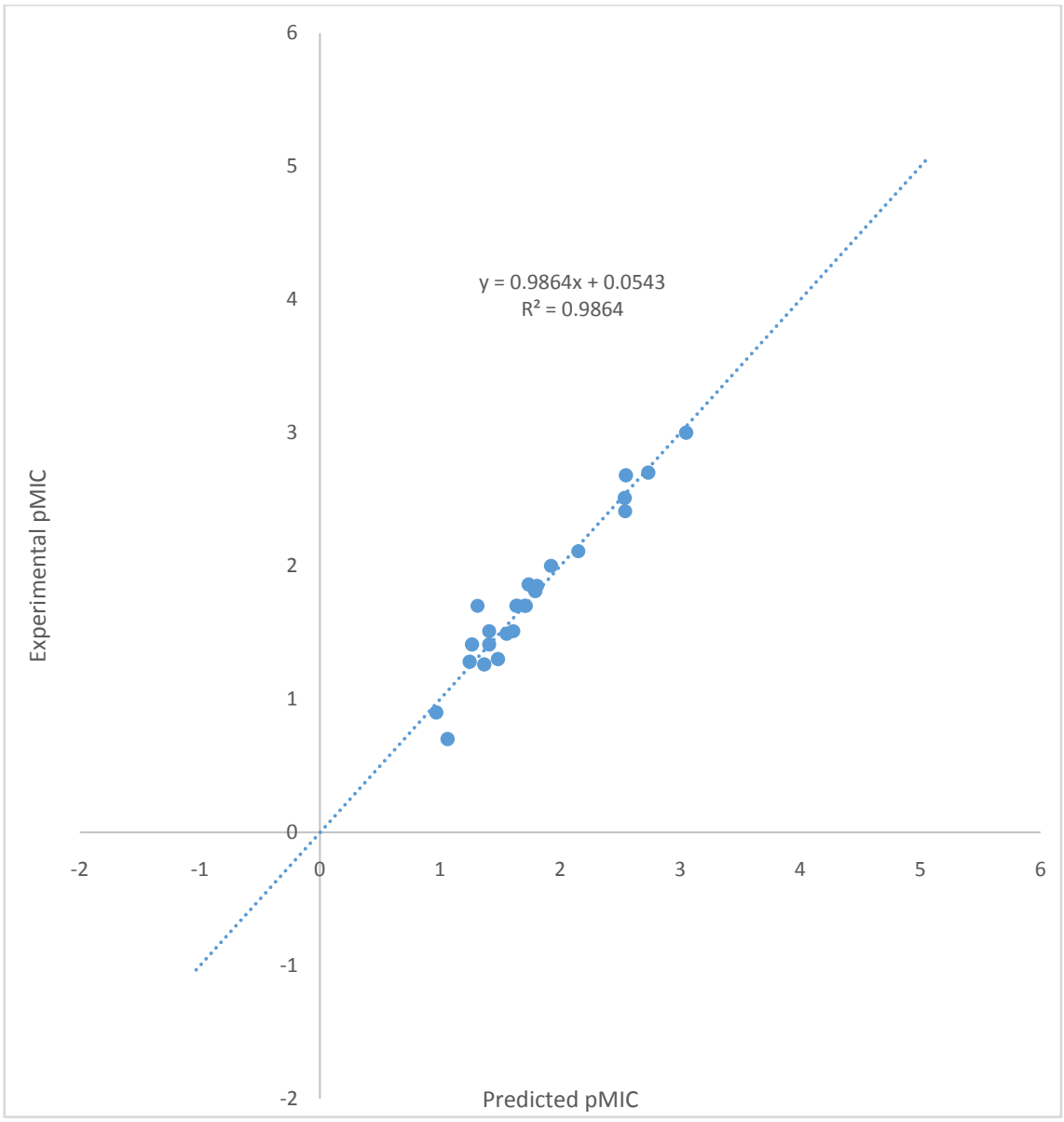


Fig. 4.1: Plot of Experimental Versus Predicted pMIC of model 1

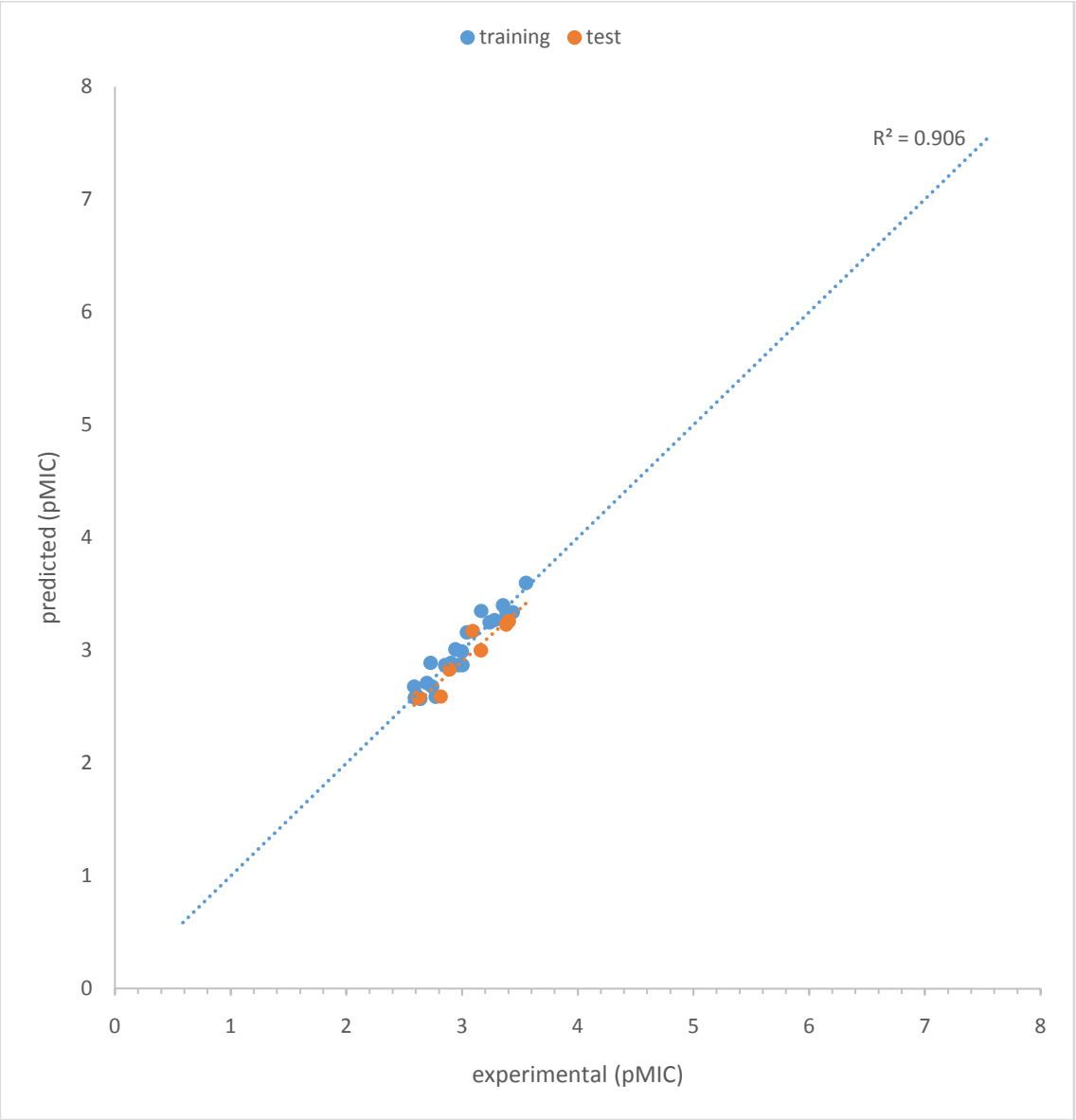


Fig 4.2: Plot of experimental versus predicted pMIC of model 4

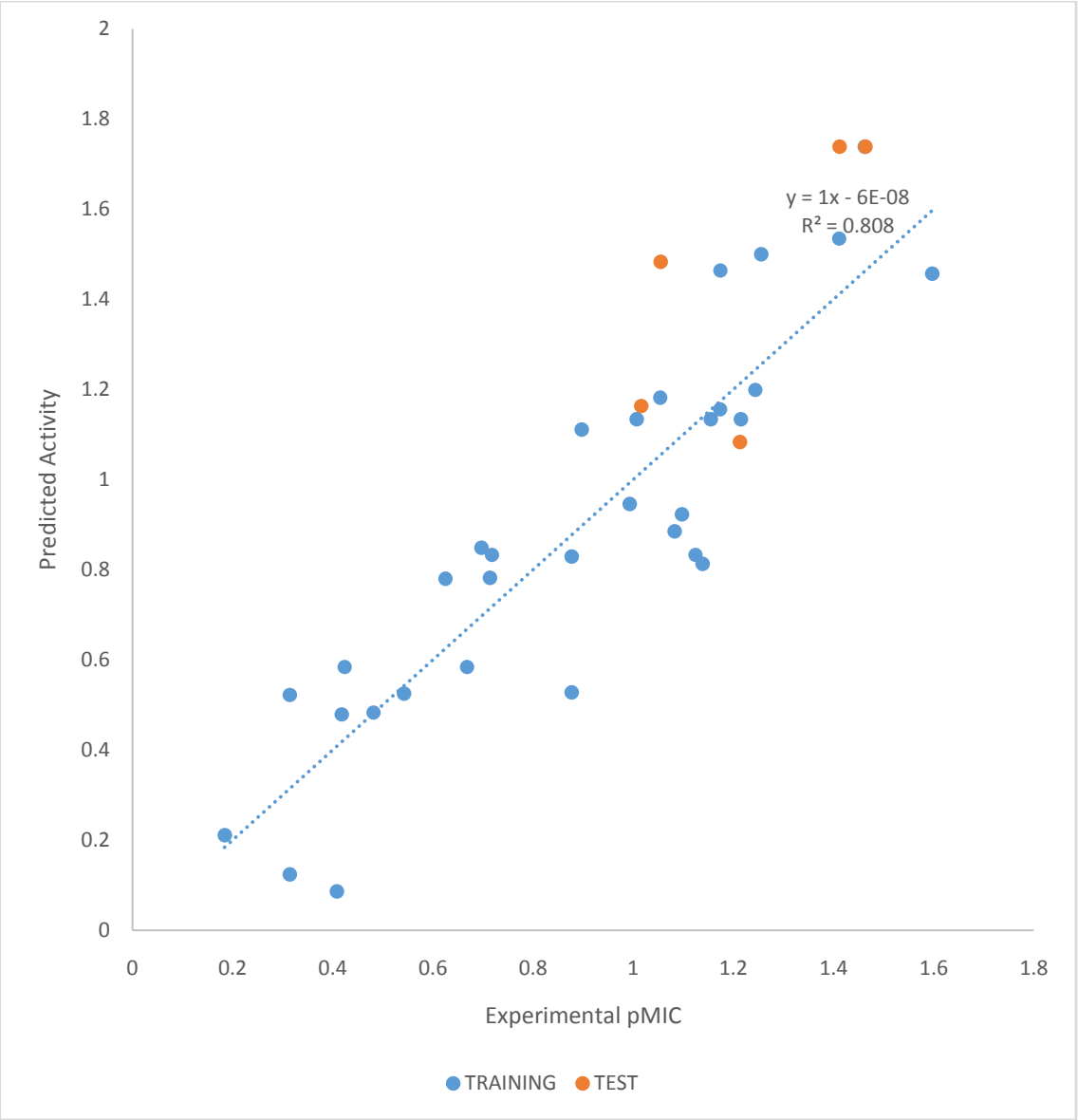


Fig 4.3: Plot of Experimental versus predicted pMIC of model 7

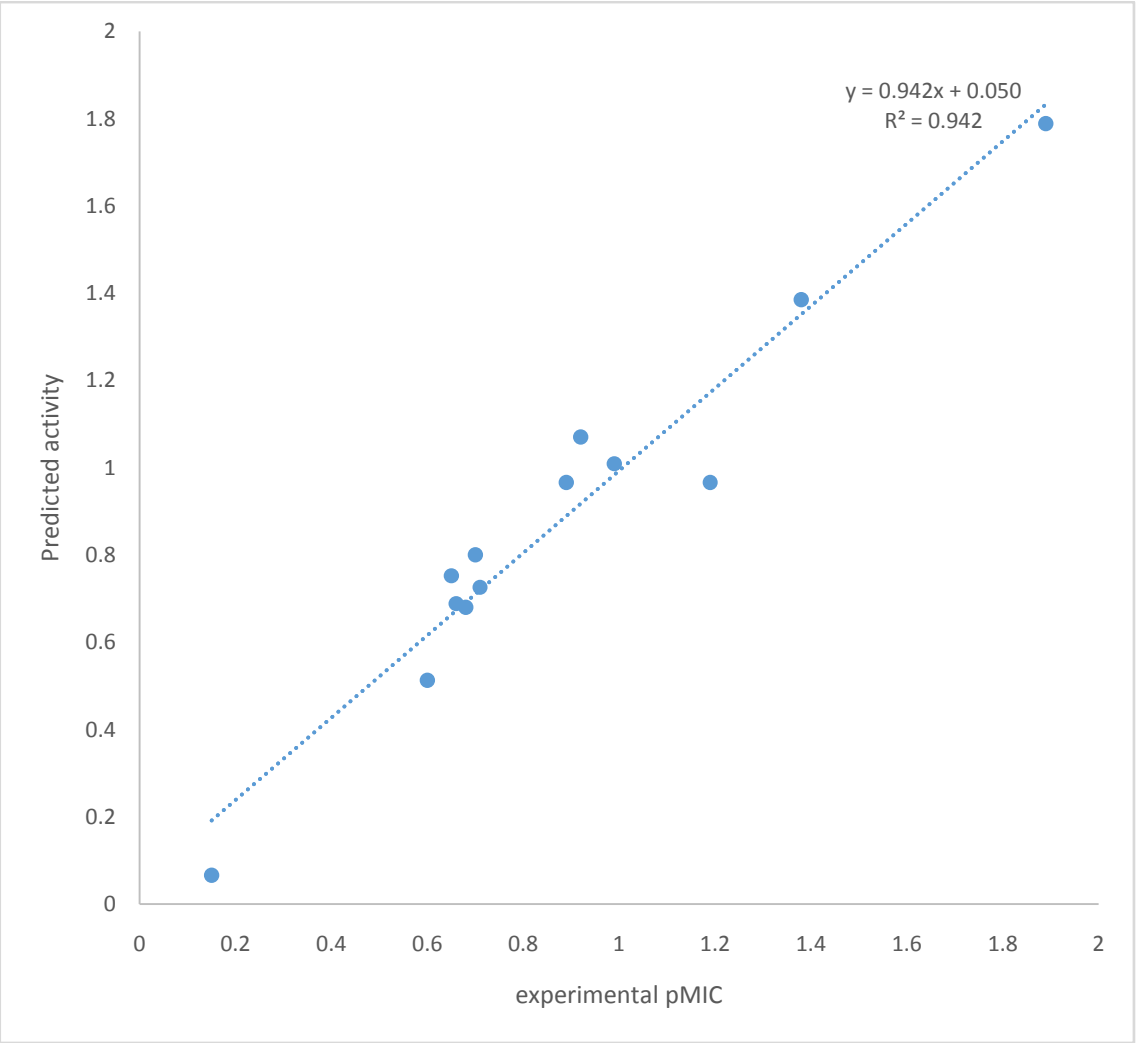


Fig 4.4: Plot of experimental versus predicted pMIC of model 10

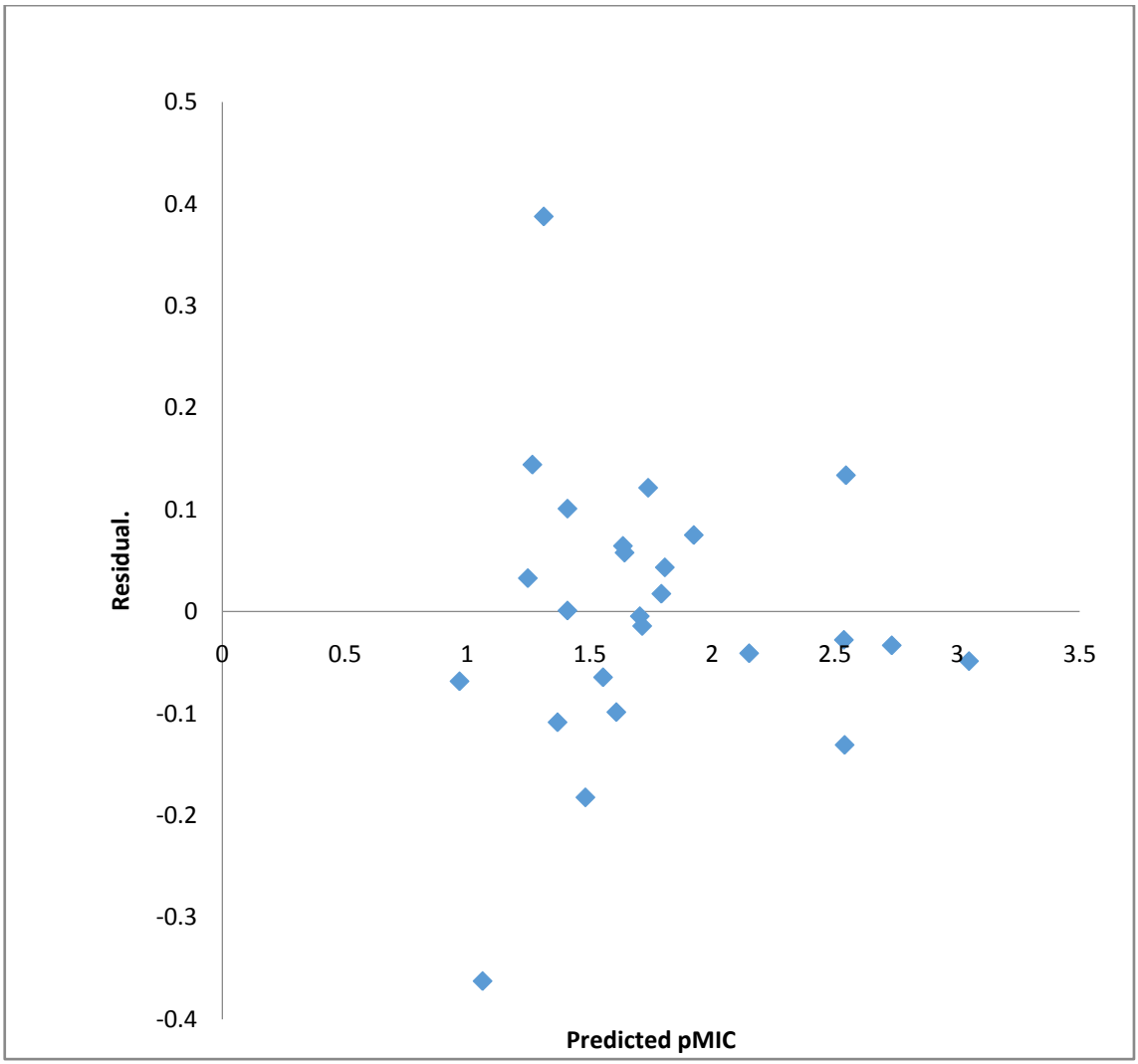


Fig. 4.5: Residual Plot of Model 1

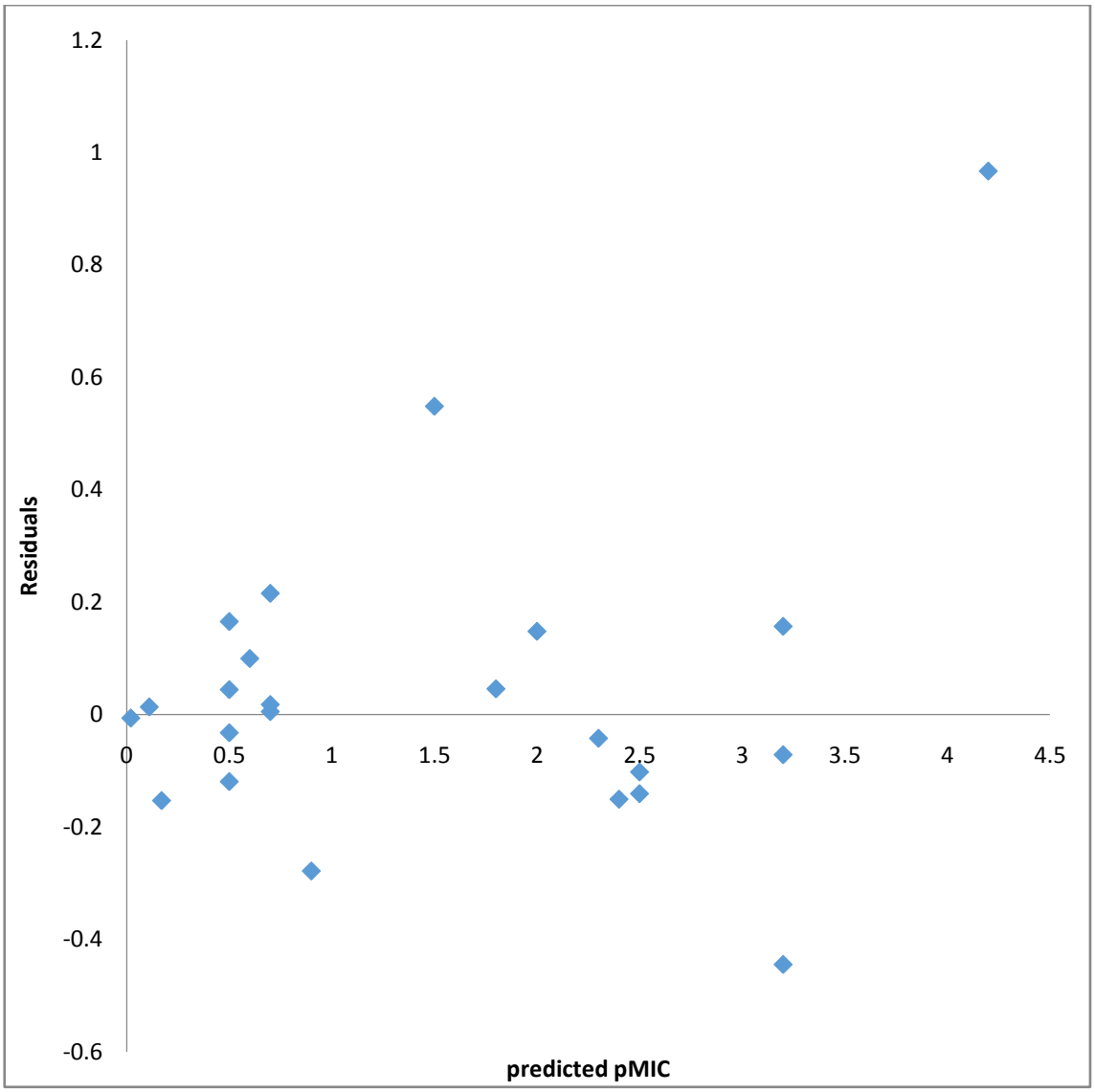


Fig. 4.6: Residual Plot of Model 4

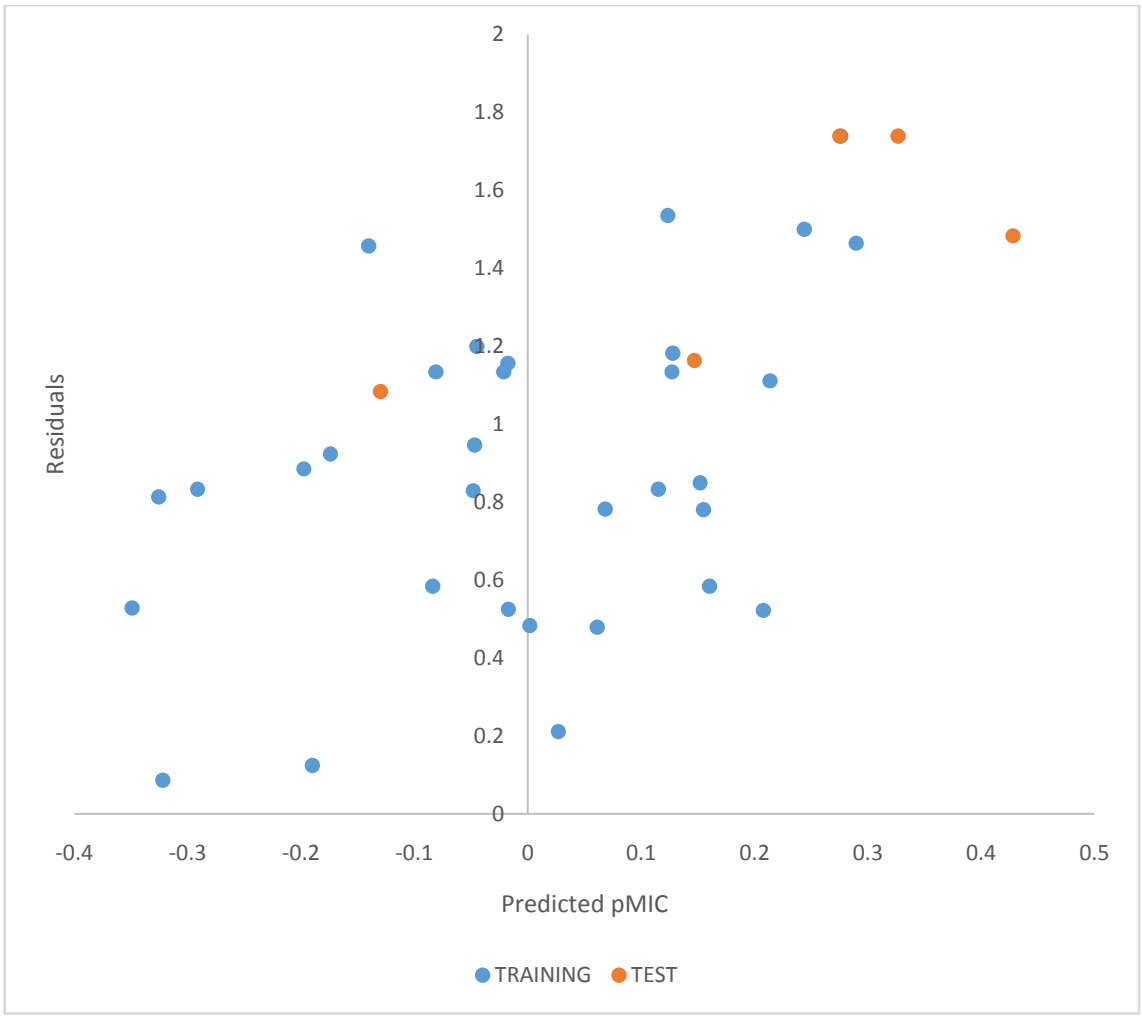


Fig 4.7: Residual Plot of Model 7

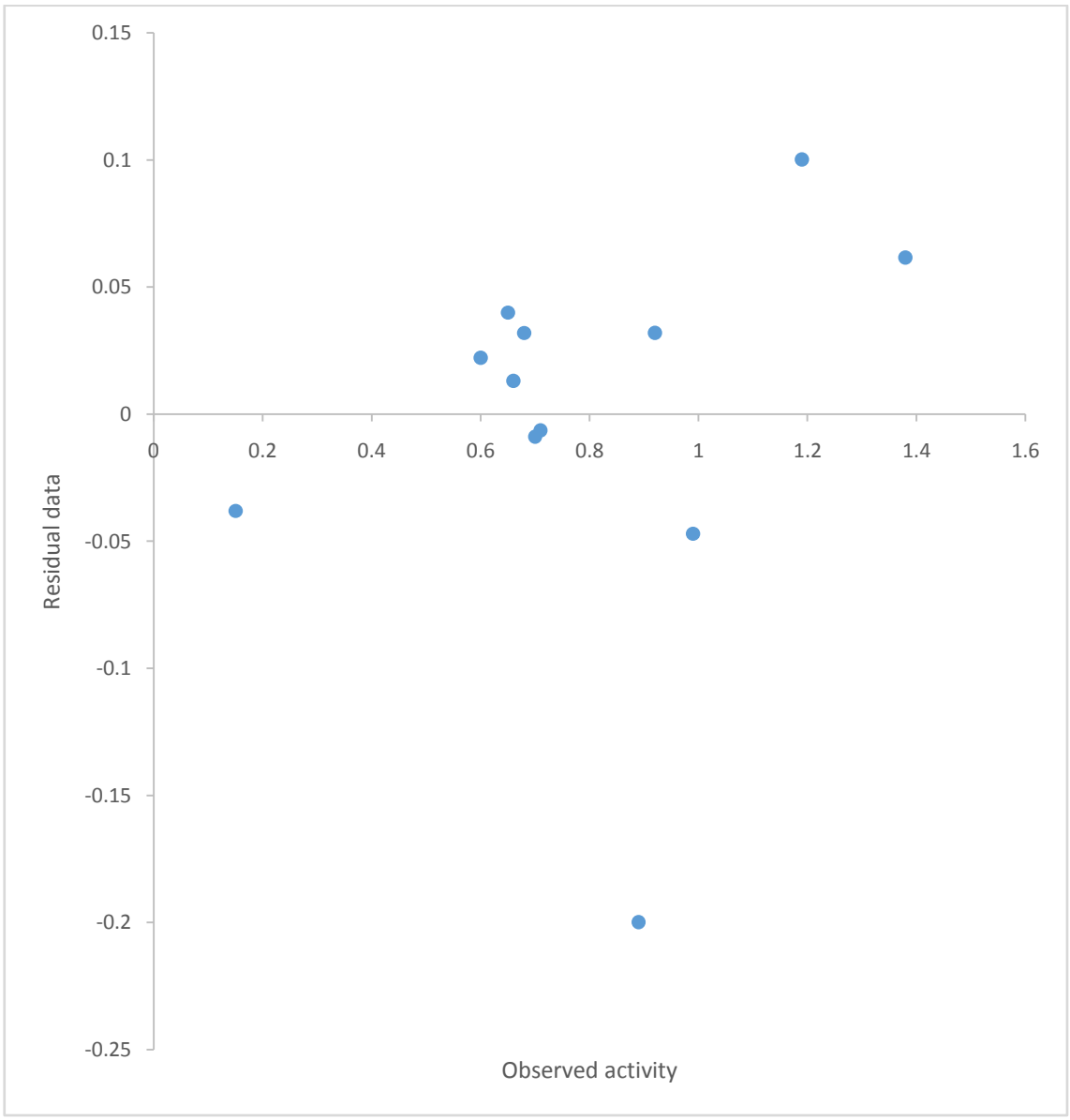


Fig 4.8: Residual Plot of Model 10

Table 4.5: Comparison of observed and predicted pMIC of model 1

Compounds	Actual values	predicted values	residuals values
1	4.53	4.560	-0.03
2	4.16	4.097	0.06
3	3.52	3.608	0.09
6	4.15	4.139	0.01
7	3.80	3.776	0.02
8	3.83	3.850	0.02
9	4.13	4.195	-0.07
10	3.83	3.855	-0.03
12	4.47	4.485	-0.02
13	4.72	4.720	0.00
14	3.83	3.789	0.04
15	4.14	4.0995	0.04
17	3.51	3.534	-0.02
19	4.09	3.998	0.09
20	3.51	3.535	-0.03

Table 4.6: Comparison of observed and predicted pMIC of model 4

Compound	Actual values	predicted values	residual values
1	2.68000000	2.74047800	-0.06
2	2.89000000	2.90442700	-0.01
3	3.60000000	3.55309600	0.05
4	2.87000000	2.85161400	0.02
5	3.25000000	3.23910800	0.01
6	2.87000000	2.96696900	-0.10
7	2.99000000	2.99727500	-0.00
8	3.01000000	2.94309900	0.07
9	2.71000000	2.69563800	0.015
10	2.68000000	2.58562300	0.09
11	3.35000000	3.16478300	0.19
12	3.30000000	3.38182300	-0.08
13	3.34000000	3.43747900	-0.10
14	3.27000000	3.28108900	-0.01
15	3.40000000	3.35432400	0.05
16	2.58000000	2.59206700	-0.01
17	2.59000000	2.77082600	-0.18
18	2.89000000	2.72726800	0.16
19	3.16000000	3.04156800	0.12
20	2.87000000	3.00249100	-0.13
21	2.57000000	2.63895400	-0.07

Table 4.7: Comparison of observed and predicted pMIC of model 7

Compound	Actual values for A : Pmic	predicted values	residual values
1	1.46400000	1.17415500	0.28984500
2	0.92300000	1.09714100	-0.17414100
3	0.81300000	1.13876000	-0.32576000
4	0.78000000	0.62494400	0.15505600
5	1.45700000	1.59743200	-0.14043200
6	1.13400000	1.15515100	-0.02115100
7	1.13400000	1.21516000	-0.08116000
8	0.52200000	0.31406600	0.20793400
9	1.19900000	1.24385700	-0.04485700
10	0.94600000	0.99292000	-0.04692000
11	0.88500000	1.08261800	-0.19761800
12	0.48300000	0.48121300	0.00178700
13	0.58400000	0.42352200	0.16047800
14	1.11100000	0.89727600	0.21372400
15	0.82900000	0.87721400	-0.04821400
16	1.53500000	1.41141100	0.12358900
17	0.52500000	0.54215300	-0.01715300
18	0.21100000	0.18402400	0.02697600
19	0.47900000	0.41776600	0.06123400
20	1.15600000	1.17350700	-0.01750700
21	0.78200000	0.71369000	0.06831000
22	0.83300000	0.71784500	0.11515500

Table 4.7: Comparison of observed and predicted pMIC of model 7 (continuation)

23	1.13400000	1.00680700	0.12719300
24	0.83300000	1.12437100	-0.29137100
25	0.12400000	0.31406600	-0.19006600
26	1.50000000	1.25590100	0.24409900
27	1.73900000	1.46277400	0.27622600
28	0.84900000	0.69688500	0.15211500
29	1.18200000	1.05402700	0.12797300
30	0.58400000	0.66801500	-0.08401500
31	0.08600000	0.40811300	-0.32211300
32	0.52800000	0.87721400	-0.34921400

Table 4.8: Comparison of observed and predicted pMIC of model 10

Observed activity	Predicted activity	residual
0.99000000	1.00890000	-0.019
0.92000000	1.07022400	-0.150
1.89000000	1.78829400	0.102
1.38000000	1.38493100	-0.005
0.71000000	0.72571600	-0.016
1.19000000	0.96620300	0.224
0.89000000	0.96620300	-0.076
0.65000000	0.75221400	-0.102
0.60000000	0.51273700	0.087
0.68000000	0.68004800	-4.800
0.15000000	0.06590000	0.084
0.70000000	0.80034900	-0.100
0.66000000	0.68828000	-0.028

Table 4.9: Golbraikh and Tropsha external validation parameters for model 1

s/n	parameter	value
1	r^2	0.9399
2	r'^2_0	0.9399
3	r_0^2	0.9364
4	k	0.9948
5	K'	1
6	$r^2 - r_0^2 / r^2$	0.0037
7	$r^2 - r'^2_0 / r^2$	0

Table 4.10: Golbraikh and Tropsha external validation parameters for model 4

S/N	Parameter	Value
1	r^2	0.9467
2	$r'_0{}^2$	0.9467
3	$r_0{}^2$	0.9448
4	k	0.9795
5	K'	1
6	$r^2 - r_0{}^2 / r^2$	0.0020
7	$r^2 - r'_0{}^2 / r^2$	0

Table: 4.11: Golbraikh and Tropsha external validation parameters for model 7

S/n	Parameter	Value
1	r^2	0.9341
2	$r'_0{}^2$	0.9341
3	$r_0{}^2$	0.9298
4	K	0.9952
5	K'	1
6	$r^2 - r_0{}^2 / r^2$	0.005
7	$r^2 - r'_0{}^2 / r^2$	0

Table 4.12: Golbraikh and Tropsha external validation parameters for model 10

S/n	Parameter	Value
1	r^2	0.9465
2	$r'_0{}^2$	0.9465
3	$r_0{}^2$	0.9436
4	K	0.9970
5	K'	1
6	$r^2 - r_0{}^2 / r^2$	0.0030
7	$r^2 - r'_0{}^2 / r^2$	0

Table 4.13: Variance Inflation Factor (VIF) for the descriptors in model 1

S/no	Dependent variable	R^2	VIF
1	MW	0.584	2.38
2	CAA	0.406	1.68
3	cLogp	0.521	2.09
4	MR	0.65	2.86
5	MV	0.75	4.00
6	P	0.77	4.35

Table 4.14: Variance Inflation Factor (VIF) for the descriptors in model 4

S/NO	Dependent variable	R^2	VIF
1	SHBint5	0.07	1.08
2	WD.Unity	0.04	1.04
3	Wlambda2.energy	0.08	1.09
4	Wlambda3.polar	0.05	1.05

Table 4.15: Variance Inflation Factor (VIF) for the descriptors in model 7

S/NO	Dependent variable	R^2	VIF
1	BCUTw_11	0.25	1.34
2	PNSA_3	0.15	1.18
3	MOMI-XZ	0.48	1.93
4	PSA	0.22	1.27
5	MEP	0.22	1.28
6	cLogp	0.24	1.32

Table 4.16: Variance Inflation Factor (VIF) for the descriptors in model 10

s/no	Dependent variable	R^2	VIF
1	SE	0.081	1.09
2	MEP	0.039	1.04
3	H^o	0.065	1.07
4	G^o	0.055	1.06

4.3.7: Outlier analysis of models 1, 4, 7 and 10

Fig 4.9, 4.10, 4.11 and 4.12 indicated the outliers analysis.

4.3.8: William's plot applicability domain of models 1, 4, 7 and 10

The physicochemical, structural and information on the training set of the developed model was given in Fig 4.13, 4.14, 4.15 and 4.16.

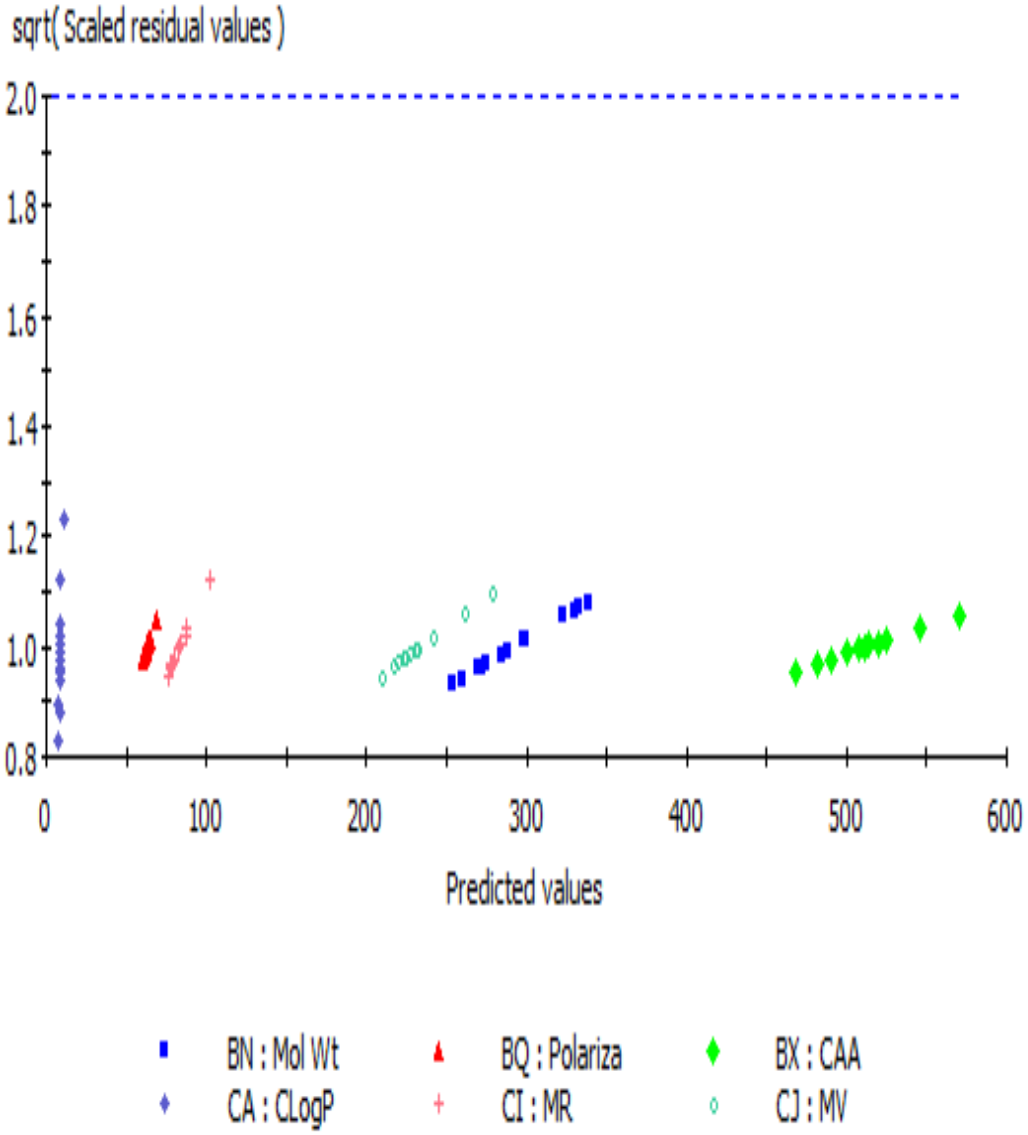


Fig 4.9i: Outlier analysis of model 1

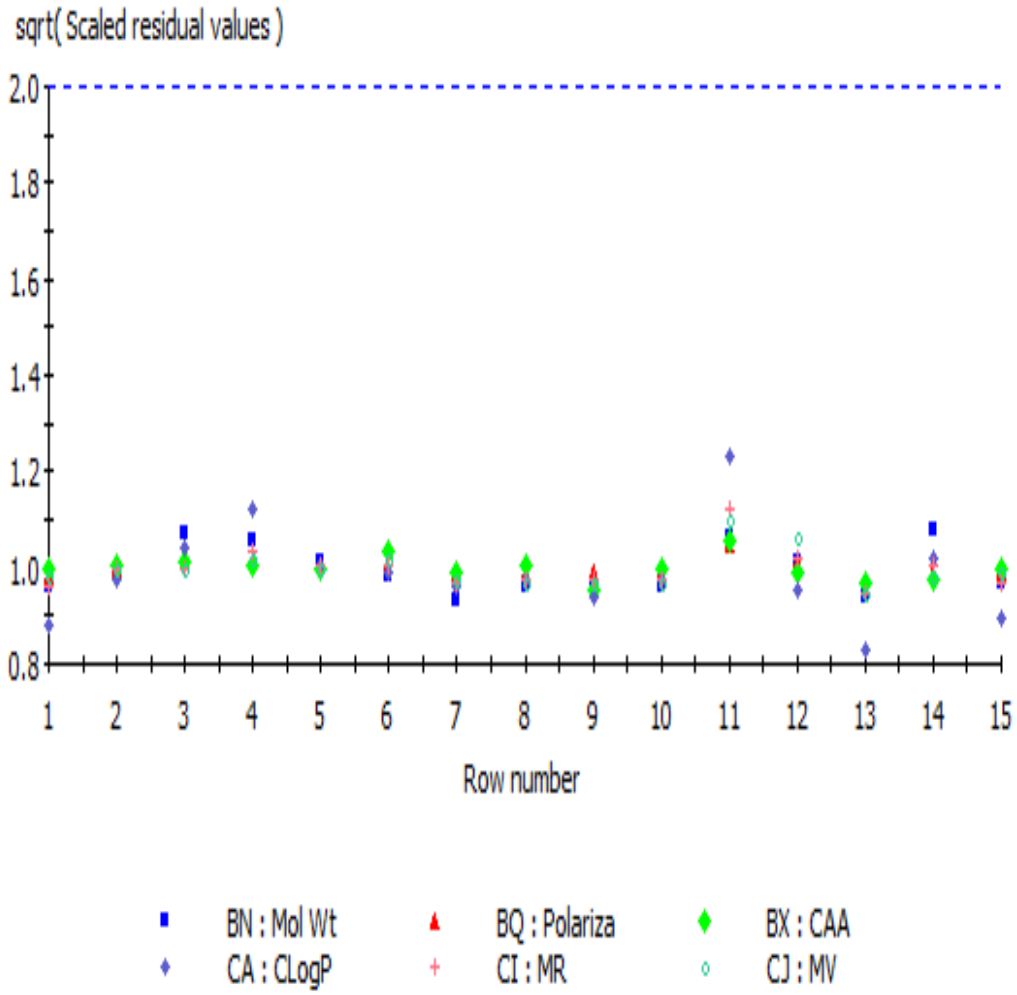


Fig 4.9ii : Outlier Analysis of model 1

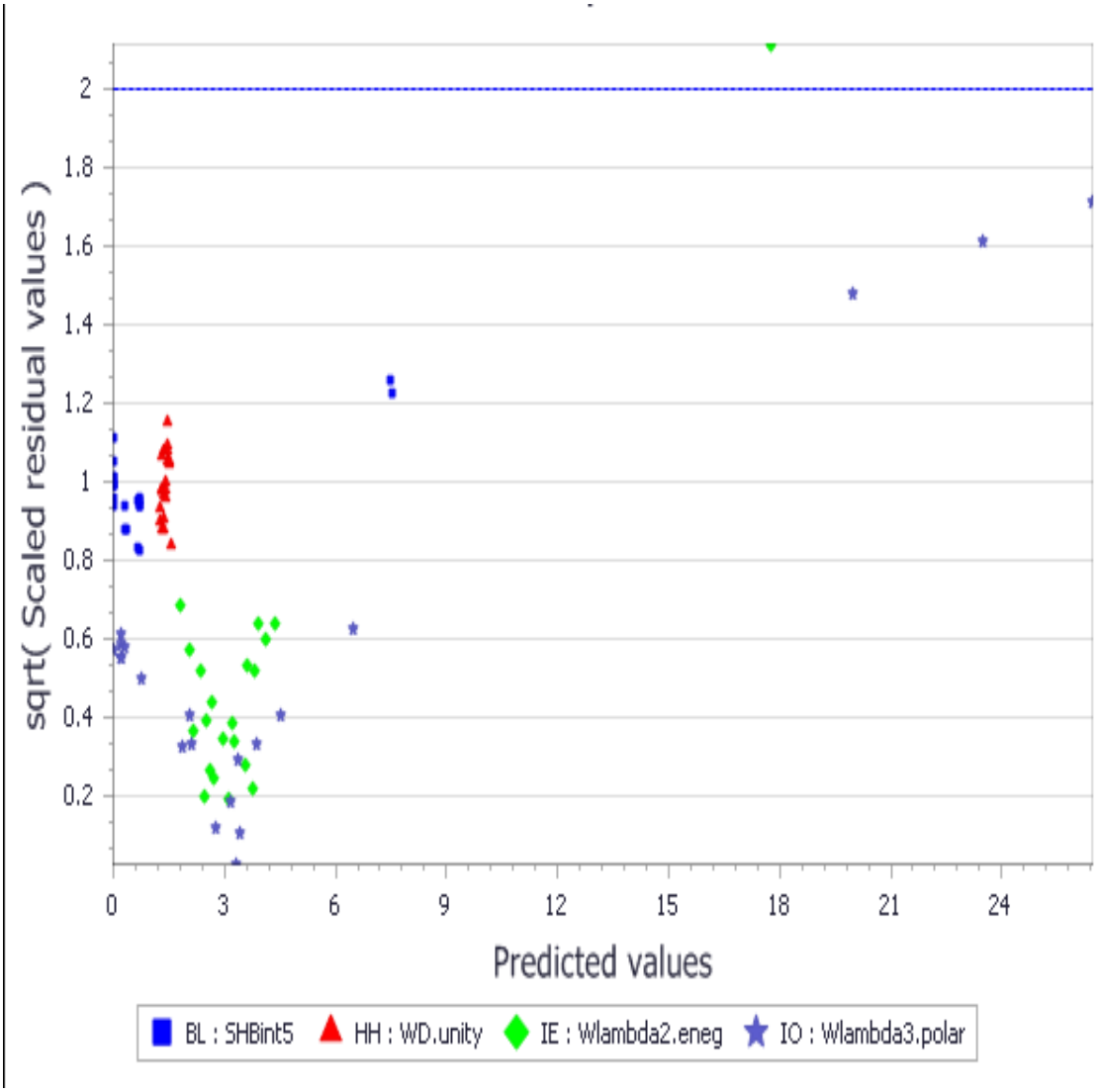


Fig 4.10i : Outlier analysis of model 4

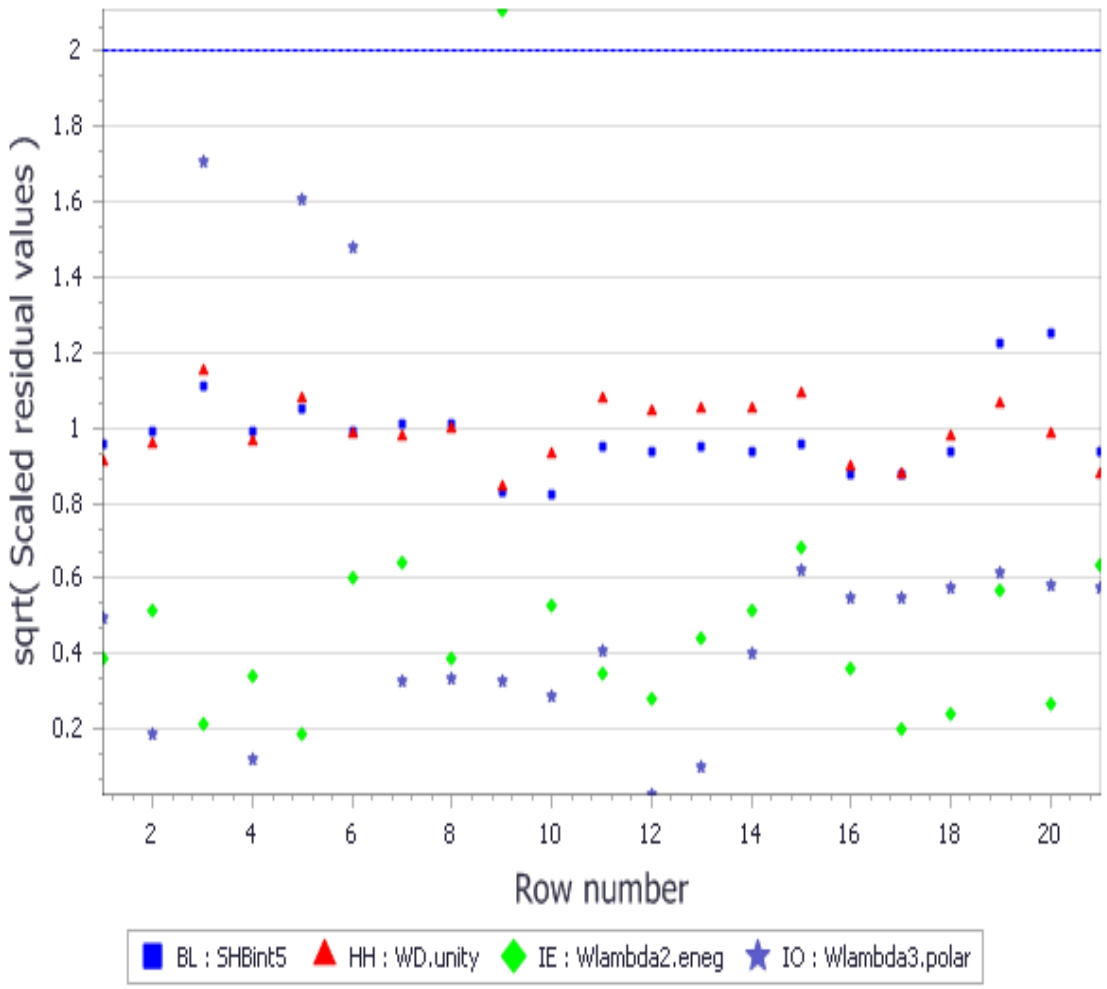


Fig 4.10ii: Outlier Analysis of model 4

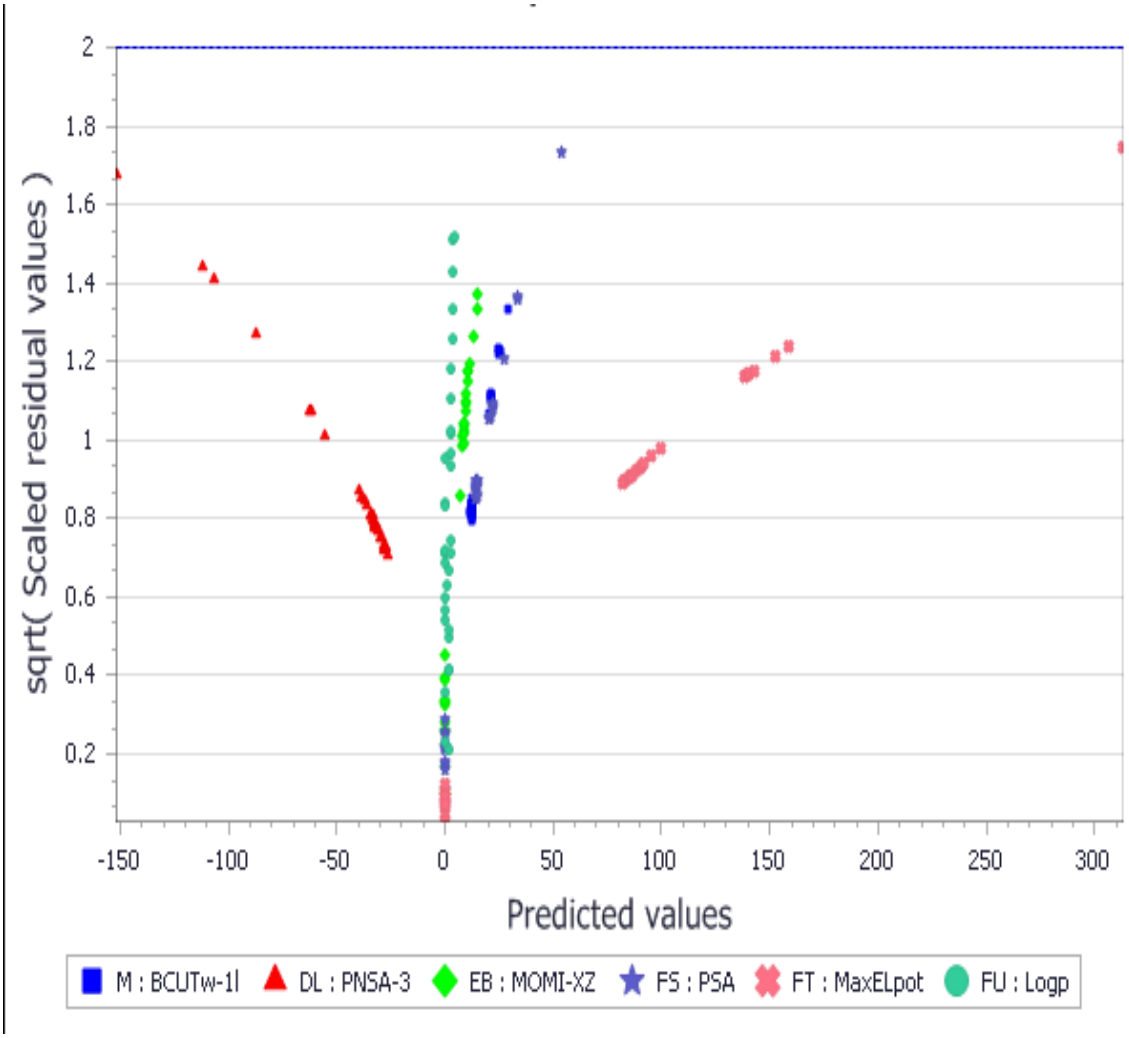


Fig 4.11i: Outlier analysis of model 7

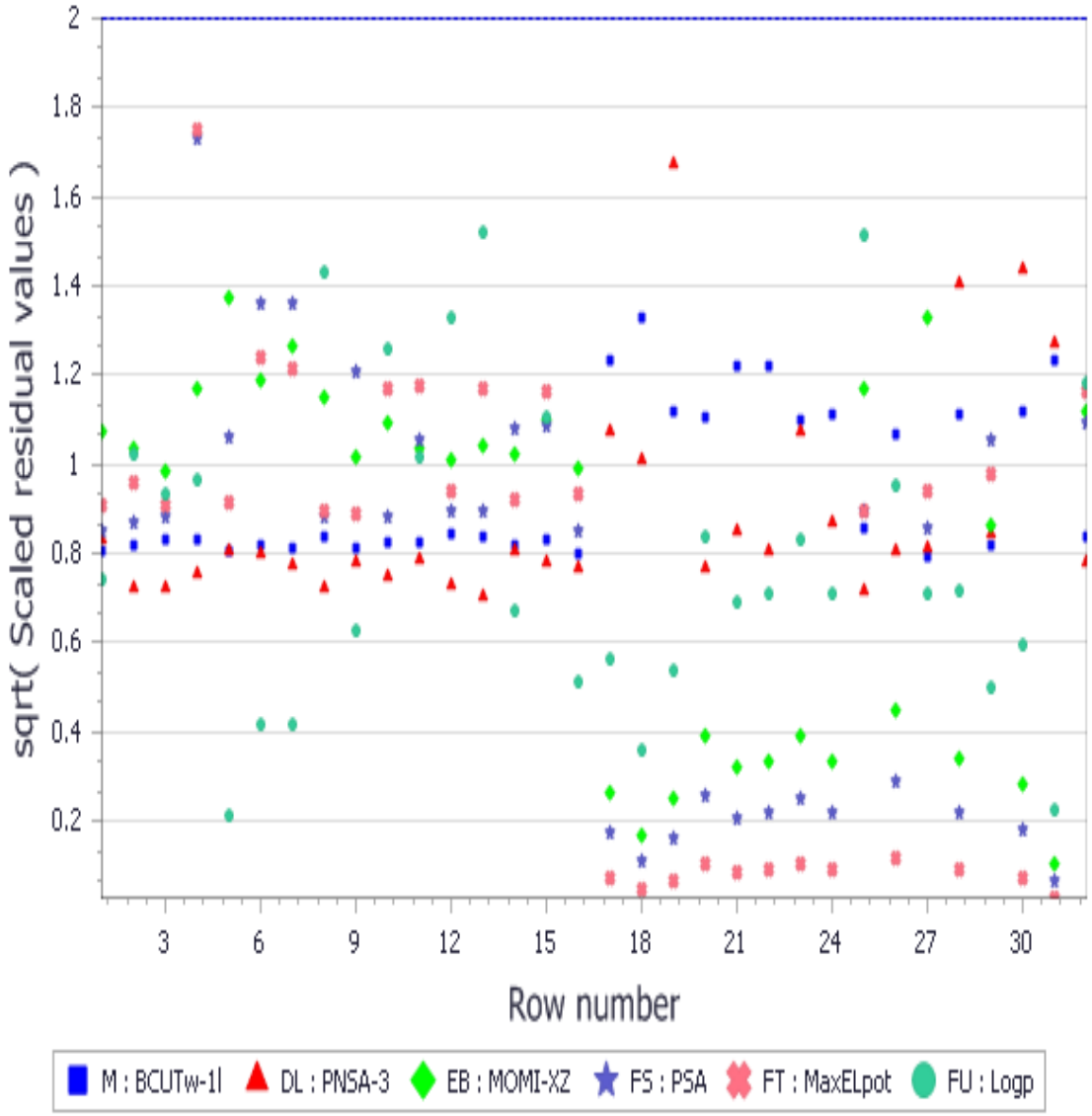


Fig 4.11ii :Outlier analysis of model 7

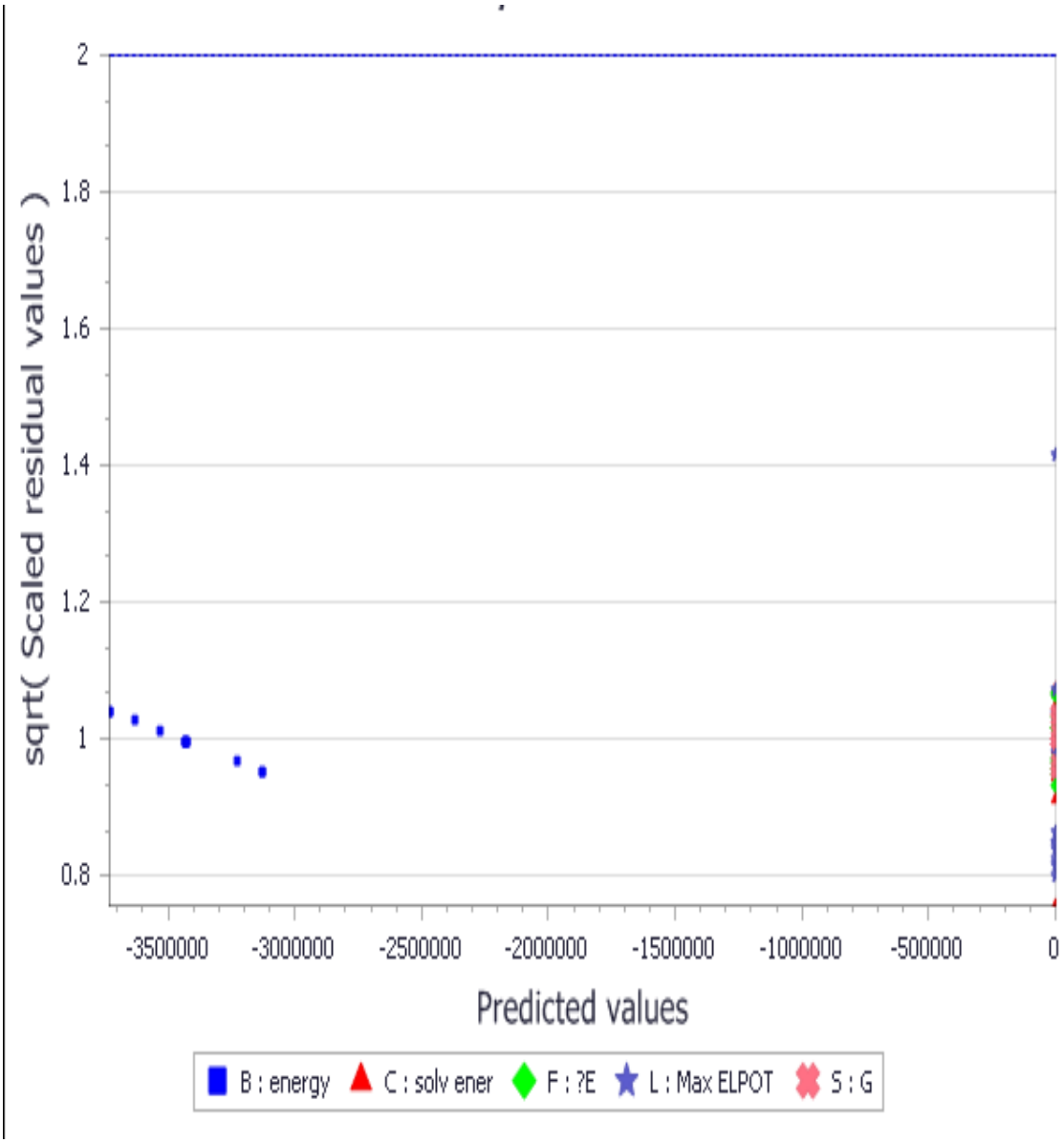


Fig 4.12i : Outlier analysis of model 10

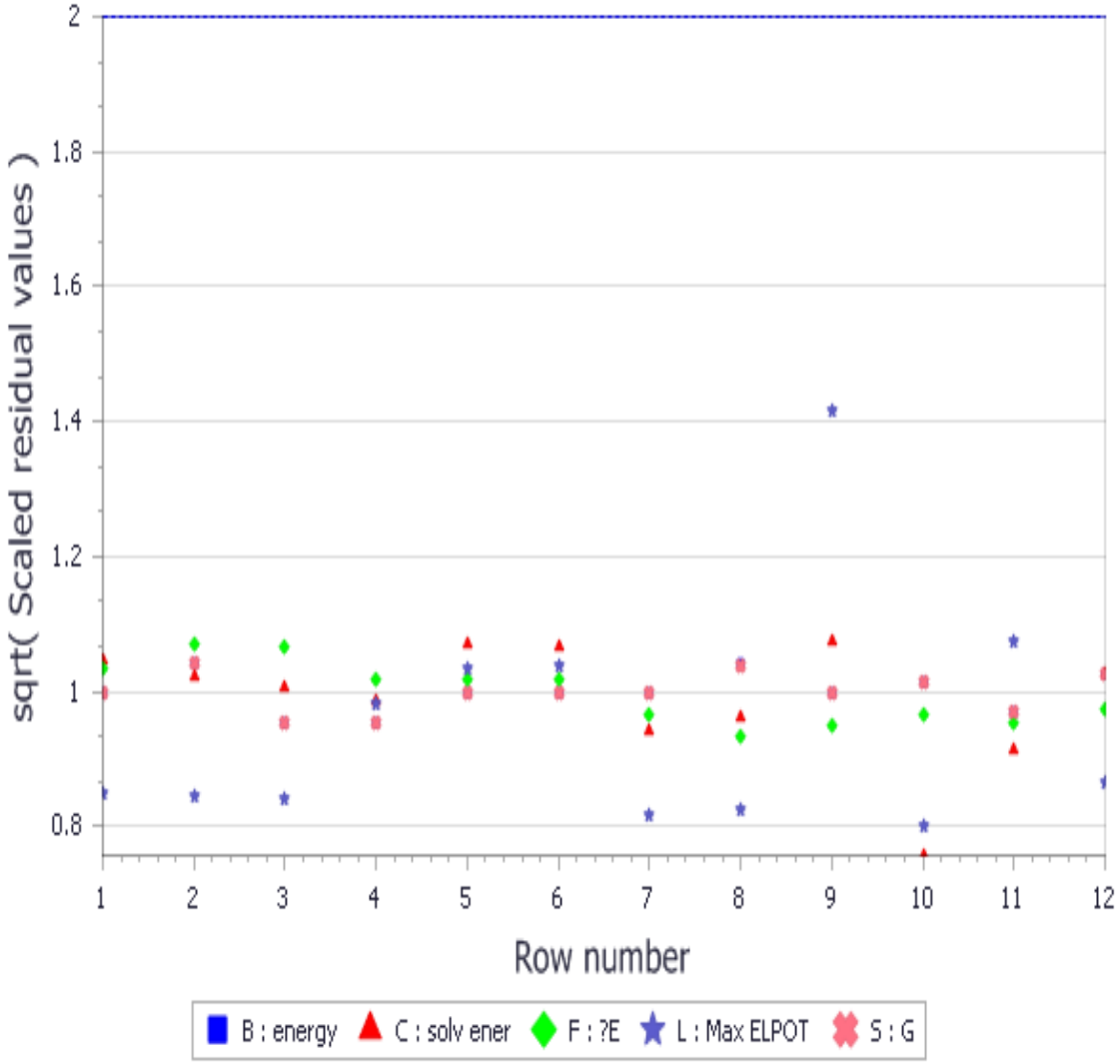


Fig 4.12ii: Outlier Analysis of model 10

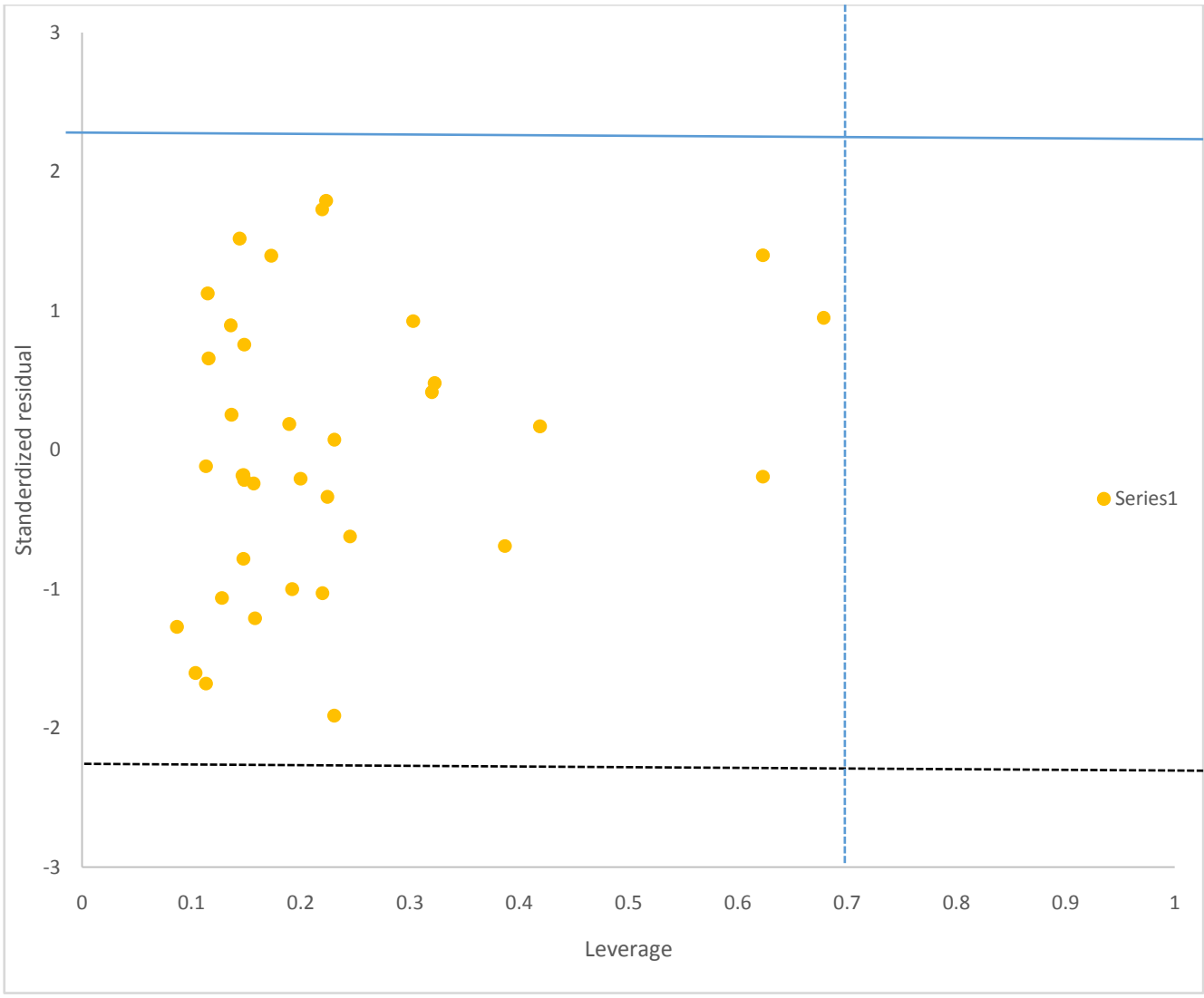


Fig 4.13: William's plot applicability domain of model 1

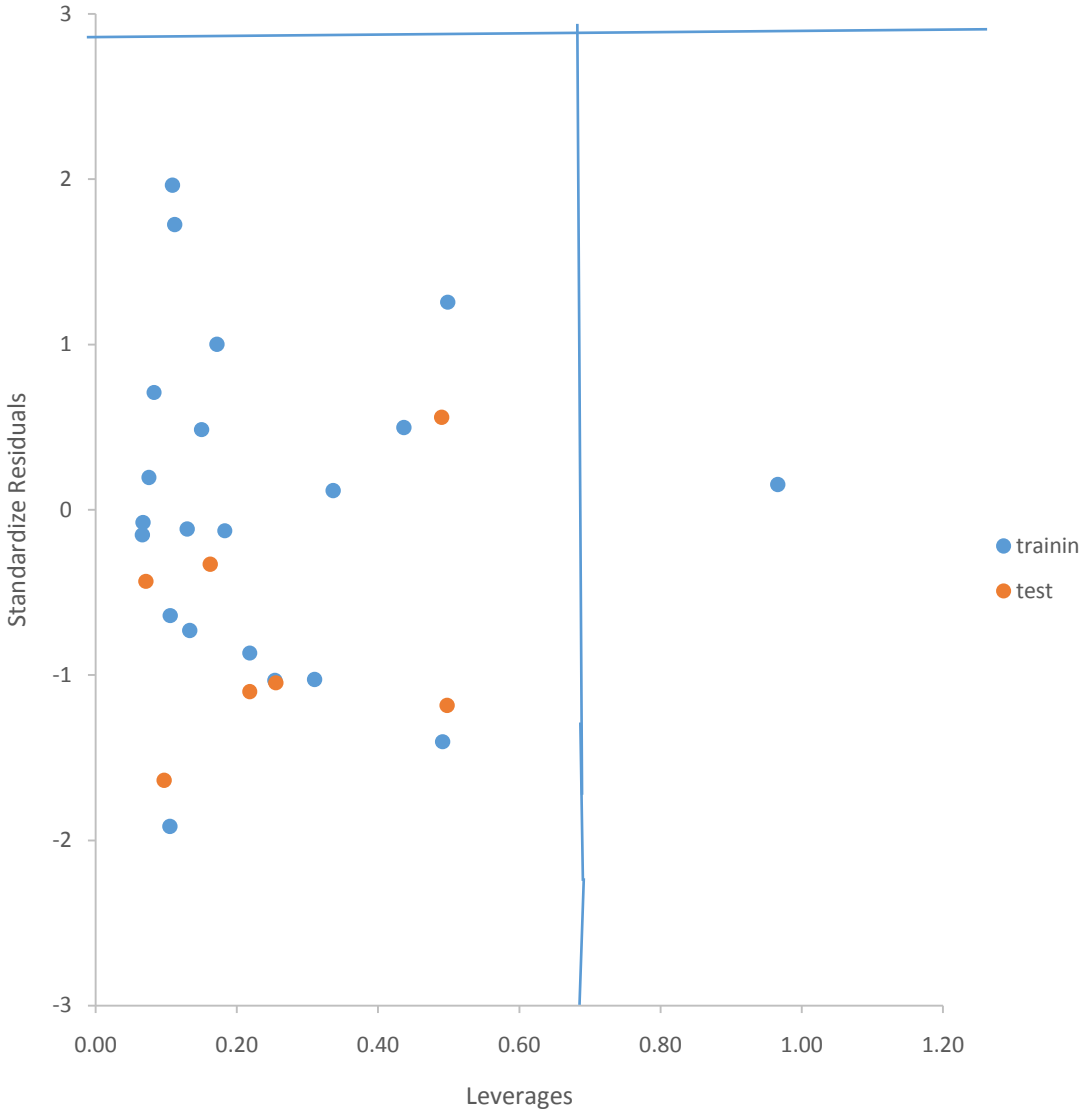


Fig 4.14: The William's plot Applicability Domain of model 4

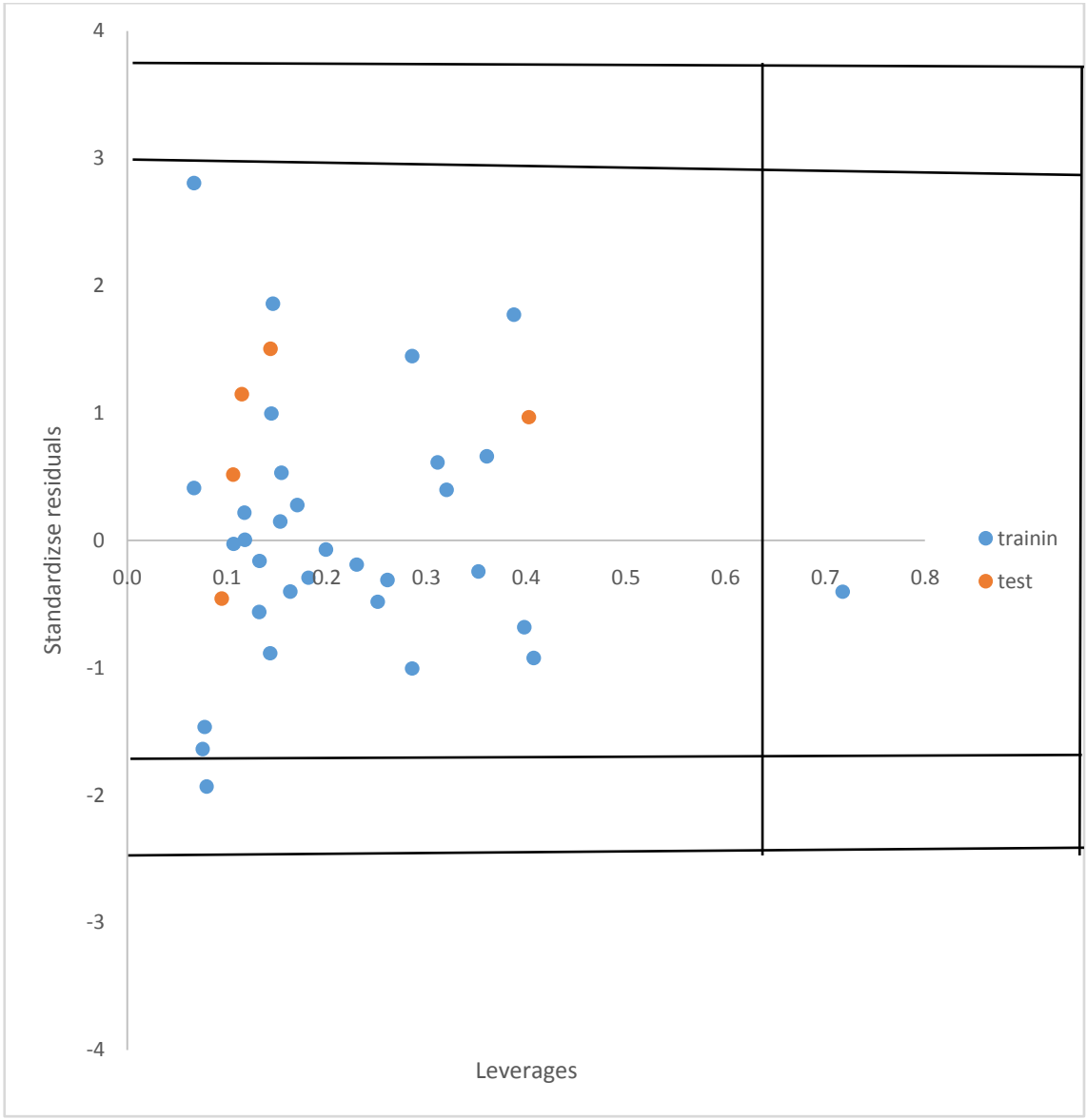


Fig 4.15: The william's plotApplicability Domain of model 7

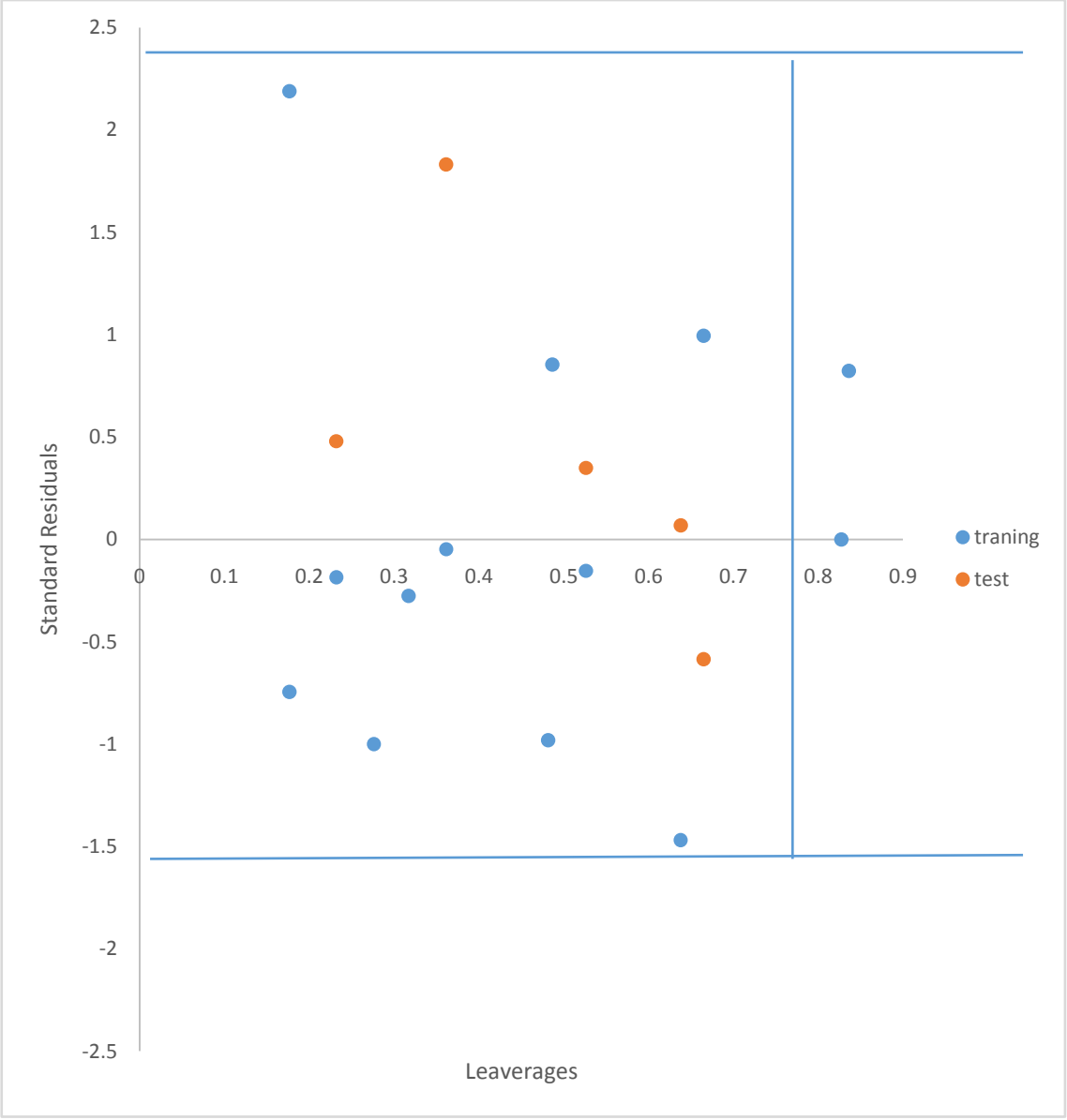


Fig 4.16: The William's plot Applicability Domain of model 10

CHAPTER FIVE

5.0 DISCUSSION OF RESULTS

5.1 Molecular Optimization and Descriptor Calculation

The optimization of the molecules used for this study were successfully carried out. The optimization procedure was carried out in the following order :Semi-empirical (PM3) then follow by Density Functional Theory (B3LP/6-3LG). The molecular properties (descriptors) computed from each optimized structure include 0D, 1D, 2D, and 3D descriptors with the aid of Spartan 14 V.1.1.0 software, Chm 3D Pro 12.0.1V software, Padel descriptor tool kit listed in section 3.2.4. The successful optimization of the studied molecules indicates that their structures correspond their natural geometry. Thus, all the descriptors derived from these structures were good and reliable.

5.2 GFA Derived QSAR Model for anti-*Candida albicans* compounds of Model-1 Compounds (set 1)

Model-1 was developed to predict the antifungi activity of the compound against *Candida albicans*. The result of the GFA QSAR model is in conformity with the standard shown in Table 3.2 as R^2 = 0.9864, R^2_{adj} = 0.9762, $R^2_{(cv)}$ = 0.9280, $R^2_{pred.}$ = 0.9725. This confirmed the robustness of the model. It could be seen from Model-1that the accuracy of the model, indicated by the R^2 value, was reasonably high therefore the predictive power of the model, as indicated by the R^2_{adj} and R^2_{cv} values, was also high and the regression was significant according to F-test. The Friedman's lack-of-fit (LOF) score which evaluated the QSAR model by considering the number of descriptors as well as the quality of fitness, was chosen: the lower the LOF, the less likely it is that GFA model will fit the data. The significant regression is given by F-test, and the higher the value, the better the model.

Table 4.1 showed a univariate analysis for the activity data and contains several statistical measures which described the activity data. The most vital parameters in Table 4.1 are the skewness and kurtosis. Skewness is the third moment of the distribution, which indicated the symmetry of the distribution. Kurtosis is the fourth moment of the distribution, which indicated the profile of the column of data relative to a normal distribution (Khaled and Abdel-Shafi, 2011; Khaled and El-Sherik, 2013; Khaled, 2011).

The agreement between the experimental pMIC values of molecules used in the training set and the predicted values by the optimization model 1 was presented in the Fig. 4.1. The high linearity (R^2 = 0.9864) of the plot of observed pMIC against predicted pMIC, also showed the high predictability of the model. The high Linearity of this plots indicated high predictive power of this models.

To ascertain whether there exists a systematic error in the model development, observed pMIC was plotted against residual pMIC. Figure 4.5 shows the model did not show any proportional and systematic error, because the propagation of the residuals on both sides of zero is random (Beheshti *et al*., 2012).

The comparison of observed and predicted antifungi activities of the compounds is presented in Table 4.5. The predictability of model 1 was evidenced by the low residual values observed in the Table.

Tables 4.9 give the Golbraikh and Tropsha criteria for the optimization models 1. The results showed that the QSAR models met the Golbraikh and Tropsha criteria for robust and stable QSAR model shown in Table 3.7. These further confirm the robustness and stability of the generated models.

In order to check for multi-collinearity among the descriptors used in the model, the corresponding VIF values of the six descriptors used in the optimization model (Model

1) were calculated and presented in Table 4.13. All the variables have VIF values of less than 5 indicating that the model obtained was statistical significance and the descriptors were found to be reasonably orthogonal (Jaiswal *et al*., 2004; Shapiro *et al*., 1998).

Fig 4.9i and 4.9ii contains two charts : (i) Contains the residual values plotted against the predicted values and (ii) displays the residual values plotted against number of compounds in the training set of table 3.2. Each chart contains a dotted line that indicates the critical threshold of two standard deviations beyond which a value may be considered an outlier. Inspection of Fig 4.9i and 4.9ii showed that there are no points appearing outside the dotted lines which make the QSAR model acceptable.

The William's plot for model 1 (set 1) was given in Fig. 4.13. The leverage values was calculated for every compound and plotted against standardized residuals, and it allowed a graphical detection of both the outliers and the influential chemicals in the model . The influence of compounds on model could be given when the leverage value is greater than warning leverage (in this case h* = 0.70) . This plot showed that all the compounds of the training set and test set were inside of this square area and all the chemicals have a leverage lower than the warning h* value of 0.70. All results confirm that model 1 was a valid model and could be utilized to provide a better understanding of the anti-*Candida albicans* activity.

5.2.1 Significance of the descriptors in model 1

The descriptors molecular weight, connolly accessible area, cLogp and molar refractivity all have negative coefficients, indicating that the magnitude of the MIC of these compounds against *Candida albicans* increases with decrease in the values of these descriptors. Since the activity of compounds varies inversely with its minimum inhibitory concentration (MIC), the lower the values of these descriptors in a molecule, the lower the activity of the molecule against *Candida albicans* because activity is

directly proportional to descriptors. Also for positive coefficient of the descriptors molar volume and polarizability indicating that the size of the MIC of these compounds against *Candida albicans* increase with increase with the values of these descriptors.

Molar refractivity and cLogp are Thermodynamic descriptors. The negative sign of the thermodynamic parameters, indicating that the molecular structure need lower-cost energy to remove electrons from the molecule to favour the biological activity, after all the molar refractivity is contributing as a steric descriptor of the substituent (Ameji *et al*., 2015; Nawale and Ashtekar, 2014).

Further more, the molar refractivity is a parameter that is dependent on the spatial array of the aromatic ring in the synthesized compounds (Ameji *et al*., 2015). The spatial arrangement is important or comes into play when a drug molecule interacts with a receptor. The molar refractivity was negatively correlated with the activity against *Candida albicans*. This indicated that as activity increase molar refractivity also increase and vice-versa.

Connolly's accessible area (caa), a steric descriptor, represented the surface area that is in contact with the solvent. The descriptor bears negative coefficient in the model, suggesting increase in the bulkiness of the substituents as such as descriptor increase activity of the compound increase (Choudhary *et al*., 2005).

Molar volume (mv) is a descriptors of geometrical, characterizing the shape and extent of the molecule in terms of its 3D coordinates. The positive sign of molecular volume indicates that the magnitude of the MIC of these compounds against *Candida albicans* increases with increase in the values of these descriptors and vice versa (Vadlamudi and Kulkarni, 2003).

Polarizability is a quantum Chemical descriptor. The positive sign of quantum chemical parameter indicates the influence on the model owing to their significance in the selected model (Arthur *et al*., 2015).

5.3 GFA Derived QSAR Model for anti-*Candida albicans* compounds of model-4 Compounds (set 2)

Model-4 was developed to predict the antifungi activity of the compound against the *Candida albicans* (set 2). The result of the GFA QSAR model is in conformity with the standard shown in Table 3.3 as R^2 = 0.9064, R^2_{adj} = 0.8830, $R^2_{(cv)}$ = 0.7321, R^2_{pred}. = 0.8722. This confirmed the robustness of the model. It could be seen from Model-4that the accuracy of the model, indicated by the R^2 value, was reasonably high therefore the predictive power of the model, as indicated by the R^2_{adj} and R^2_{cv} values, was also high and the regression was significant according to F-test. The Friedman's lack-of-fit (LOF) score which evaluated the QSAR model by considering the number of descriptors as well as the quality of fitness, was chosen: the lower LOF, the less likely it is that GFA model will fit the data. The significant regression is given by F-test, and the higher the value, the better the model.

Table 4.2 showed a univariate analysis for the activity data and contains several statistical measures which described the activity data. The most vital parameters in Table 4.2 are the skewness and kurtosis. Skewness is the third moment of the distribution, which indicated the symmetry of the distribution. Kurtosis is the fourth moment of the distribution, which indicated the profile of the column of data relative to a normal distribution.

The agreement between the experimental pMIC values of molecules used in the training set and the predicted values by the optimization model 4 was presented in the Fig. 4.2. The high linearity (R^2 = 0.9064) of the plot of observed pMIC against predicted pMIC,

showed a high predictability of the model. The high linearity of this plots indicated high predictive power of this models.

Furthermore, the plot of observed pMIC versus residual pMIC (Figure 4.6) indicated that there was no systemic error in model development as the propagation of residuals was observed on both sides of zero (Ameji *et al*., 2015; Nawale and Ashtekar, 2014).

The comparison of observed and predicted antifungi activities of the compounds is presented in Table 4.6. The predictability of model 4 was evidenced by the low residual values observed in the Table.

Tables 4.10 give the Golbraikh and Tropsha criteria for the optimization models 4. The results showed that the QSAR models met the Golbraikh and Tropsha criteria for robust and stable QSAR model shown in Table 3.7. These further confirm the robustness and stability of the generated models.

In other to check for multi-collinearity among the descriptors used in the model, the corresponding VIF values of the four descriptors used in the optimization model (Model 4) was calculated and presented in Table 4.14. All the variables have VIF values of less than 5 indicating that the obtained model has statistical significance, and the descriptors were found to be reasonably orthogonal (Jaiswal *et al*., 2004; Shapiro *et al*., 1998).

Fig 4.10i and 4.10ii contains two charts : (i) Contains the residual values plotted against the predicted value and (ii) displays the residual values plotted against number of compounds in the training set of Table 3.3. Each chart contains a dotted line that indicates the critical threshold of two standard deviations beyond which a value may be considered to an outlier. Inspection of Fig 4.10i and 4.10ii showed that there is no points appeared outside the dotted lines which make the QSAR model acceptable.

The William's plot for model 4 was given in Fig. 4.14. The leverage values was calculated for every compound and plotted against standardized residuals, and it

allowed a graphical detection of both the outliers and the influential chemicals in the model . The influence of compounds on model can be given when the leverage value is greater than warning leverage (in this case h* = 0.65) .This plot showed that all the compounds of the training set and test set are inside of this square area with the exception of a compound ,all the chemicals have a leverage lower than the warning h* value of 0.65. All results confirm that model 4 was a valid model and can be utilized to provide a better understanding of the anti-*Candida albicans* activity (Beheshti *et al*., 2012).

5.3.1 Significance of the Descriptors in the Model 4

Model 4 indicated that the MIC was correlated to SHBint5 , WD unity, wlambda3.polar and wlambda2.eneg descriptors.A positive correlation of anti-fungal activity against *C.albicans* with SHBint5 (electrotopological state) descriptor indicated the strength for potential hydrogen bonds of path lenght5 in the molecular structure of the compound , WD unity (A non-directional WHIM) descriptor indicated weighted by unit weights and wlambda3.polar (Non-directional WHIM) descriptor indicated weighted by atomic polarizabilitiesindicated that the magnitude of the pMIC of these organisms was directly proportional to values of these descriptors in the molecules. The higher the values of these descriptors in a molecule, the more the activity of the molecule against *Candida albicans* and vice versa.

Also, anti *Candidaalbicans* pMIC of the organism is negatively correlated to their wlambda2.eneg (A non-directional WHIM) descriptor indicated weighted by mulliken atomic electronegativities. It implies that the higher the value of this descriptor, the higher the antifungal activity against *Candida albicans* and vice versa.

5.4 GFA Derived QSAR Model for anti-*Candida albicans* Compounds of Model- 7 Compounds (set 3)

Model-7 was developed to predict the antifungi activity of the compound against *Candida albicans*. The result of the GFA QSAR model is in conformity with the standard shown in Table 3.4 (a and b) as R^2 = 0.8085, R^2_{adj} = 0.7625, $R^2_{(cv)}$ = 0.6578, R^2_{pred}. = 0.7941. This confirmed the robustness of the model. It could be seen from Model-7that the accuracy of the model, indicated by the R^2 value, was reasonably high therefore the predictive power of the model, as indicated by the R^2_{adj} and R^2_{cv} values, was also high and the regression was significant according to F-test. The Friedman's lack-of-fit (LOF) score which evaluated the QSAR model by considering the number of descriptors as well as the quality of fitness, was chosen: the lower the LOF, the less likely it is that GFA model will fit the data. The significant regression was given by F-test, and the higher the value, the better the model.

Table 4.3 showed a univariate analysis for the activity data. It contains several statistical measures which described the activity data. The most important parameters in Table 4.3 are the skewness and kurtosis. Skewness is the third moment of the distribution, which indicated the symmetry of the distribution. Kurtosis is the fourth moment of the distribution, which indicated the profile of the column of data relative to a normal distribution.

The agreement between the experimental pMIC values of molecules used in the training set and the predicted values by the optimization model 7 was presented in the Fig. 4.3. The high linearity (R^2 = 0.8085) of the plot of observed pMIC against predicted pMIC, showed a high predictability of the model. The high Linearity of this plots indicated high predictive power of this models.

Furthermore, the plot of observed pMIC versus residual pMIC (Figure 4.7) indicated that there was no systemic error in model development as the propagation of residuals was observed on both sides of zero (Ameji *et al*., 2015; Nawale and Ashtekar, 2014).

The comparison of observed and predicted antifungi activities of the compounds is presented in Table 4.7. The predictability of model 7 was evidenced by the low residual values observed in the Table.

Tables 4.11 give the Golbraikh and Tropsha criteria for the optimization models 7. The results showed that the QSAR models met the Golbraikh and Tropsha criteria for robust and stable QSAR model shown in Table 3.7. These further confirm the robustness and stability of the generated models.

In order to check for multi-collinearity among the descriptors used in the model, the corresponding VIF values of the six descriptors used in the optimization model (Model 7) were calculated and presented in Table 4.15. All the variables have VIF values of less than 5 indicating that the obtained model has statistical significance, and the descriptors were found to be reasonably orthogonal (Jaiswal *et al*., 2004; Shapiro *et al*., 1998).

Fig 4.11i and 4.11iicontains two charts. (i) Contains the residual values plotted against the predicted values and (ii) displays the residual values plotted against number of compounds in the training set of Table 3.4. Each chart contains a dotted line that indicates the critical threshold of two standard deviations beyond which a value may be considered to an outlier. Inspection of Fig. 4.11i and 4.11iishowed that there is no points appeared outside the dotted lines which make the QSAR model acceptable.

The Williams plot for the developed model-7 was given in Fig. 4.15. The leverage values was calculated for every compound and plotted against standardized residuals, and it allowed a graphical detection of both the outliers and the influential chemicals in the model . The influence of compounds on model could be given when the leverage

value was greater than warning leverage (in this case h* = 0.656) .This plot showed that all the compounds of the training set and test set were inside of this square area with the exception of a compound, i.e with the exception of a compound, all the chemicals have a leverage lower than the warning h* value of 0.656. All results confirm that the build model was a valid model and could be utilized to provide a better understanding of the anti-*Candida albicans* activity.

5.4.1 Significance of the Descriptors in the Model 7

A positive correlation of anti-fungal activity against *C.albicans* with PNSA-3 (Charge Polar surface area-3), MOMI-XZ (moment of inertial), and max electpot (Quantum Electronics) in model 7 indicated that the magnitude of the pMIC of these compounds were directly proportional to values of these descriptors in the molecules. The lower the values of these descriptors in a molecule, the more the activity of the molecule against *Candida albicans* and vice versa.

Also, anti- *Candida albicans* pMIC of the organism was negatively correlated to their BCUTw-11 (BCUT), Polar surface area (Dimensional) and logp (Lipophilicity). It implies that the higher the value of this descriptor, the higher the antifungal activity against *Candida albicans* and vice versa.

PNSA-3 is a Charge weighted partial negative surface area of the molecule (Ivakumar *et al*, 2007). This descriptor provide important information concerning features of the molecules that were responsible for strong intermolecular interactions and in turn important for helping to correlate structural features to physical properties of the molecules (Stanton and Jurs, 1990).

MOMI-XZ this is a moment of inertial types of descriptor that describes the contour of the projection of the molecule on the plane defined by two principal axes of the molecule XZ ratio of x to y of the atoms in the molecule.

Maximum electrostatic potential defined as the sum of surface contributions of polar atoms (usually oxygens, nitrogens and attached hydrogens) in a molecule, has been shown to correlate well with drug transport properties (Gilson *et al*., 1988).

BCUTw-11 this is an eigenvalue based descriptor (n high lowest atom weighted BCUTS). The descriptor was based on a weighted version of the burden matrix which takes into account both the connectivity as well as atomic properties of the molecules. The weights were variety of atom properties placed along the diagonal of the Burden matrix of the Coumarin molecule (Pearlman & Smith, 1999).

Polar surface area is a descriptor that was shown to correlate well with passive molecular transport through membranes and, therefore, allows prediction of transport properties of drugs. It combine shape and electronic information to characterize atoms in the molecule (Hall, 1990).

cLogp is a dimensional descriptor which describe the logarithm of the partition coefficient between n-octanol and water (Baggiani *et al*., 2007; Golmohammadi and Dashtbozorgi, 2010). The clogP value is a property of lipophilicity. The property indicate the lipid solubility and affinity of compounds toward lipid. The clopP value has negative correlation with the activity; this indicates that the substituents which increase solubility in water will lead to decreased activity. The decrease in solubility in water or increase in the lipophillic nature of compounds will lead to increased penetration of the compounds across the cell wall as shown by the work of (Lipinski *et al*., 2001).

5.5 GFA Derived QSAR Model for anti-*Epidermophyton floccosum* Compounds of Model-10 Compounds (set 4)

Model-10 was developed to predict the antifungi activity of the compound against *Epidermophyton floccosum*. The result of the GFA QSAR model was in conformity with the standard shown in Table 3.7 as R^2 = 0.9428, R^2_{adj} = 0.8954, R^2(cv) = 0.8196, R^2_{pred}. = 0.7202. This confirmed the robustness of the model. It could be seen from Model-10that the accuracy of the model, indicated by the R^2 value, was reasonably high therefore the predictive power of the model, as indicated by the R^2_{adj} and R^2_{cv} values, was also high and the regression was significant according to F-test. The Friedman's lack-of-fit (LOF) score which evaluated the QSAR model by considering the number of descriptors as well as the quality of fitness, was chosen: the lower the LOF, the less likely it is that GFA model would fit the data. The significant regression was given by F-test, and the higher the value, the better the model.

Table 4.4 showed a univariate analysis for the activity data. Table 4.4contains several statistical measures which described the activity data. The most important parameters in Table 4.4are the skewness and kurtosis. Skewness is the third moment of the distribution, which indicated the symmetry of the distribution. Kurtosis is the fourth moment of the distribution, which indicated the profile of the column of data relative to a normal distribution.

The agreement between the experimental pMIC values of molecules used in the training set and the predicted values by the optimization model 4 was presented in the Fig. 4.4. The high linearity (R^2 = 0.9428) of the plot of observed pMIC against predicted pMIC, showed a high predictability of the model. The high Linearity of this plots indicated high predictive power of this models.

Furthermore, the plot of observed pMIC versus residual pMIC (Figure 4.8) indicated that there was no systemic error in model development as the propagation of residuals was observed on both sides of zero (Ameji *et al*., 2015; Nawale and Ashtekar, 2014).

The comparison of observed and predicted antifungi activities of the compounds is presented in Table 4.8. The predictability of model 4 was evidenced by the low residual values observed in the Table.

Fig 4.12i and 4.12iicontains two charts. (a) Contains the residual values plotted against the predicted values and (b) displays the residual values plotted against number of compounds in the trainig set of Table 3.5 (a, b and c). Each chart contains a dotted line that indicates the critical threshold of two standard deviations beyond which a value may be considered to an outlier. Inspection of Fig. 4.12i and 4.12iishowed that there was no points appeared outside the dotted lines which make the QSAR model acceptable.

Tables 4.12 give the Golbraikh and Tropsha criteria for the optimization models 10. The results showed that the QSAR models met the Golbraikh and Tropsha criteria for robust and stable QSAR model shown in Table 3.7. These further confirm the robustness and stability of the generated models.

In order to check for multi-collinearity among the descriptors used in the model, the corresponding VIF values of the four descriptors used in the optimization model (Model 10) were calculated and presented in Table 4.16. All the variables have VIF values of less than 5 indicating that the obtained model has statistical significance, and the descriptors were found to be reasonably orthogonal (Jaiswal *et al*., 2004; Shapiro *et al*., 1998).

The Williams plot for the developed model was given in Fig. 4.16. The leverage values was calculated for every compound and plotted against standardized residuals, and it allowed a graphical detection of both the outliers and the influential chemicals in the model . The influence of compounds on model could be given when the leverage value was greater than warning leverage (in this case h* = 1.154) . This plot showed that the leverage values were less than the warning value of 1.154. Also, from Fig. 4.16, it was obvious that there were no outlier compounds with standard residuals >3d for both the training and test sets. All results confirm that the build model was a valid model and could be utilized to provide a better understanding of the anti-*Epidermophyton floccosum*.

5.5.1 Significance of the Descriptors in the Model 10

A positive correlation of anti-fungal activity against *Epidermophyton floccosum*with H^o (Thermodynamic) indicated that the magnitude of the pMIC of these complexes is directly proportional to values of this descriptor in the molecules. Since the activity of drug varies inversely with its minimum inhibitory concentration (MIC), the lower the values of these descriptors in a molecule, the more the activity of the molecule against *Epidermophyton floccosum* and vice versa. Also, anti- *Epidermophyton floccosum* pMIC of the organism is negatively correlated to their Solvation energy (Quantum electronics), Maximum Elpotential (Dimensional) and G^o (Thermodynamic). It implies that the higher the value of this descriptor, the higher the antifungal activity against *Epidermophyton floccosum* and vice versa.

Standard enthalpy (H^o)is a thermodynamic type of descriptors which describe the molecule atoms that represent a set of diverse points in space and the atomic property

and function as evaluated at those points as shown in model-10.This is in agreement with similar work carried out by (Geerlings *et al*., 2003).

Solvation Energy is a Quatum electronics type of descriptor which describe the minimum energy of the molecular conformation.

Maximum electrostatic potential is a dimensional type of descriptor that describe the van der waals surface area of the molecule.

Standard Gibbs Energy (G^o) is a thermodynamic type of descriptor that depends on the spatial array of the aromatic ring in the synthesized compounds also necessary to study the interaction of the ligand with the receptor.

CHAPTER SIX

6.0 SUMMARY, CONCLUSION AND RECOMMENDATION

6.1 Summary

The generated optimum QSAR models performed to explore the structural requirements controlling the observed biological activities of Coumarin and Neolignans Compouds are represented by models 1 (set 1), 4 (set 2), 7 (set 3) and 10 (4). Models 1, 4 and 7 give the predictive models for MIC anti-*Candida albicans* molecules while model 10 give the predictive model for MIC of anti-*Epidermophyton floccosum* molecules.

The results from this study were significant as they involve handling of cell-based activity data (MIC) with good activity predictions. Results from these studies were complimentary and could be used as guide for developing molecules with better activity.

6.2 Conclusion

QSAR analysis of 108 molecules for their activity against *Candida albicans* and *Epidermophyton floccosum* were performed using various physicochemical descriptors and GFA technique. The molecules were divided into four classes according to chemical structures. Separate analysis was performed for each chemical data set and QSAR equations were generated. These equations were then analyzed for their statistical significance and making test set predictions. The results indicated that for Coumarin and Neolignans derivatives, models 1, 4, 7 and 10 gives good prediction

The LOO and LNO cross-validation methods and the external validation indicated that the models are significant, robust and have good internal and external predictability. The inhibitory activity of the investigated compounds was described based on descriptors. Thermodynamic, dimensional and steric properties played a significant role in explaining the activity of the dataset. The results indicated that the activity against

strains of *Candida albicans* and *Epidermophyton floccosum* were favoured by smaller molecular volume, higher molecular length, the electron donating capacity, increased radius of the substituents in ring of compounds, and electrophiles centers.

From the above, these results suggested derivatives of Coumarin and Neolignans as promising compounds for the development of new anti-*Candida albicans* and anti-*Epidermophyton floccosum agents*. This work fully supports and extends the results obtained by other workers.

6.3 Recommendations

Based on this studies, the following recommendation can be drawn, this can be done by

i. Using large experimental data of these important compounds activities and structural studies would be performed to reach important compound families individually.

ii. Selected properties and one possible mechanism of action for the compounds studied.

iii. Understanding of the mechanisms involved in the immune response to fungal infection that can aid the development of effective vaccines.

iv. Focus on developing models with data sets that are larger and built around more diverse collections of compounds with a wide range of chemical functionalitites.

REFERENCES

Abi-said, D., Anaissie, E., Uzun, O., Raad, I., Pinzcowski, H., and Vartivarian, S. (1997). The Epidemiology of Hematogenous Candidiasis Caused by Different Candida Species. *Clinical Infectious Diseases*, *24*(June), 1122–1128.

Achkar, J. M., and Fries, B. C. (2010). Candida Infections of the Genitourinary Tract. *Clinical Microbiology Reviews*, *23*(2), 253–273. doi.org/10.1128/CMR.00076-09

Alanis, A. J. (2005). Resistance to Antibiotics□ : Are We in the Post-Antibiotic Era□ ? *Archives of Medical Research*, *36*, 697–705. doi.org/10.1016/j.arcmed.2005.06.009

Alexander, D., Tropsha, A., and Winkler, D. A. (2015). Beware of R 2□ :Simple , Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models.doi.org/10.1021/acs.jcim.5b00206

Ameji, J. P., Uzairu, A., and Idris, S. O. (2015). Quantum modeling of the toxicity of selected Anti-Candida albicans Schiff bases and their Nickel (II) complexes. *Journal of Computational Methods in Molecular Design Scholars Research Library*, *5*(3), 91–103. Retrieved from http://scholarsresearchlibrary.com/archive.html

Antoniadou, A., Torres, H. A., Lewis, R. E., Thornby, J., Bodey, G. P., Tarrand, J. J., and Kontoyiannis, D. P. (2003). Candidemia in a Tertiary Care Cancer Center. *Medicine*, *82*(5), 309–321. doi.org/10.1097/01.md.0000091182.93122.8e

Apic, M. B and Didier, P. (2002). Guideline for hand hygiene in health-care settings Recommendations of the Healthcare Infection Control Practices Advisory. *American Journal of Infection Control*, *30*(May), 1 – 46. doi.org/10.1067/mic.2002.130391

Arora, K. (2013). QSAR Studies of Some Pyrazolones as Antimicrobial Agent. *International Journal of Pharm and Bio Sciences*, *4*(1), 657–671.

Arthur, D. E., Uzairu, A. A., and Ja ', A. M. (2015). Quantitave structure-activity relationship (QSAR) modeling of 1, 10- phenanthroline derivatives antimalarial compounds. *Journal of ComputationalMethods in Molecular Design*, *5*(3), 33–38. Retrieved from http://scholarsresearchlibrary.com/archive.html

Bag, S., Ramar, A. S., and Degani, M. S. (2009). Synthesis and biological evaluation of a, b-unsaturated ketone as potential antifungal agents. *Med Chem Res*, (18), 309–316. doi.org/10.1007/s00044-008-9128

Baggiani, C., Anfossi, L., Baravalle, P., and Giovannoli, C. (2007). Molecular recognition of polycyclic aromatic hydrocarbons by pyrene-imprinted microspheres. *Analytical Bioanalytical Chemistry*, 413–422. doi.org/10.1007/s00216-007-1318-8

Basu, S., Gugnani, H. C., Joshi, S., Gupta, N., and Candida, D. D. De. (2003). Distribution of Candida species in different clinical sources in Delhi , India , and proteinase and phospholipase activity of Candida albicans isolates de Candida

albicans. *Revista Iberoamericana de Micología*, *20*(4), 137–140.

Batovska, D., Parushev, S., Slavova, A., Bankova, V., and Tsvetkova, I. (2007). Study on the substituents ' effects of a series of synthetic chalcones against the yeast Candida albicans. *Europea Journal of Medicinal Chemistry*, *42*, 87–92. doi.org/10.1016/j.ejmech.2006.08.012

Becke, A. D., and Becke, A. D. (1993). Densityfunctional thermochemistry . III . The role of exact exchange Density-functional thermochemistry . III . The role of exact exchange. *The Journal of Chemical Physics*, *98*(7), 5648–5652. doi.org/10.1063/1.464913

Beheshti, A., Pourbasheer, E., and Nekoei, M. (2012). QSAR modeling of antimalarial activity of urea derivatives using genetic algorithm – multiple linear regressions. *Journal of Saudi Chemical Society*, *1*, 1–9. doi.org/10.1016/j.jscs.2012.07.019

Bellmann, R. (2007). Clinical Pharmacokinetics of Systemically Administered Antimycotics. *Current Clinical Pharmacology*, *2*(1), 37–58.

Bernini, R., Gualandi, G., Crestini, C., Barontini, M., Cristina, M., Willför, S., and Saladino, R. (2009). A novel and efficient synthesis of highly oxidized lignans by a methyltrioxorhenium / hydrogen peroxide catalytic system . Studies on their apoptogenic and antioxidant activity. *Bioorganic & Medicinal Chemistry*, *17*(15), 5676–5682.doi.org/10.1016/j.bmc.2009.06.010

Bisignano, G., Sanogo, R., Marino, A., Aquino, R., Angelo, V. D., and Germano, M. P. (2000). Antimicrobial activity of Mitracarpus scaber extract and isolated constituents. *Letters in Applied Microbiology*, *30*(2), 105–108.

Bisno, A. L., Rubin, F. A., Cleary, P. P., and Dale, J. B. (2005). Classification and interpretation in quantitative structure- activity relationships *(Doctoral dissertation, University of Nottingham repository)*. *Clinical Infectious Diseases*.

Bissell, V., Felix, F. D. H., Wray, D., Edinburgh, F., and Millard, E. H. D. (1993). Comparative trial of fluconazole and amphotericin in the treatment of denture stomatitis. *Oral Surgery Oral Medicine Oral Pathology*, *76*(1), 35–39.

Blanco, J. L., and Garcia, M. E. (2008). Immune response to fungal infections. *Veterinary Immunology and Immunopathology*, *125*(1), 47–70. doi.org/10.1016/j.vetimm.2008.04.020

Blot, S. I., Vandewoude, K. H., Hoste, E. A., and Colardyn, F. A. (2002). Effects of Nosocomial Candidemia on Outcomes of Critically Ill Patients. *The American Journal of Medicine.*, *113*(06), 480–485.

Blumberg, H. M., Jarvis, W. R., Soucie, J. M., Edwards, J. E., Patterson, J. E., Pfaller, M. A., and Group, S. (2001). Risk Factors for Candidal Bloodstream Infections in Surgical Intensive Care Unit Patients : The NEMIS Prospective Multicenter Study. *IEEE Transactions on Evolutionary Computation*, *5*(15 July), 17–26.

Brameier, M., and Banzhaf, W. (2001). A Comparison of Linear Genetic Programming and Neural Networks in Medical Data Mining. *IEEE Transactions on Evolutionary Computation*, *5*(1), 17–26.

Braumann, T. (1986). Determination of hydrophobic parameters by reversed-phase liquid chromatography: Theory, Experimental techniques and application in studies on QSAR. *Journal of Chromatography*, *373*, 191–225.

Brennan, B., and Leyden, J. J. (1997). Overview of topical therapy for common superficial fungal infections and the role of new topical agents. *Journal of the American Academy of Dermatology*, *36*(2), 1–8.

Bruce, C. L. (2010). Classification and Interpretation in Quantitative Structure-Activity Relationships *(Doctoral dissertation, University of Nottingham).http://eprints.nottingham.ac uk/11666/1/thesis-final.pdf*.

Calandra, T., and Marchetti, O. (2002). Antifungal prophylaxis for intensive care unit patients : let ' s fine tune it.*Intensive Care Med*, *28*, 1698–1700. doi.org/10.1007/s00134-002-1539-4

Cannon, R. D., Holmes, A. R., Mason, A. B., and Monk, B. C. (1995). Oral Candida : Clearance , Colonization , or Candidiasis ?*Journal of Dental Research*, *74*(5), 1152–1161.

Choudhary, G., Karthikeyan, C., Moorthy, N. S. H. N., Sharma, S. K., and Trivedi, P. (2005). QSAR Analysis of some Cytotoxic Thiadiazinoacridines. *Internet Electronic Journal of Molecular Design*, *4*(11), 793–802.

Clayton, Y. M., and Hospital, S. T. (1992). Clinical and mycological diagnostic aspects of onychomycoses and dermatomycoses. *Clinical and Experimental Dermatology*, *17*(Suppl. 1), 37–40.

Coopersmith, C. M., Rebmann, T. L., Zack, J. E., Ward, M. R., Corcoran, R. M., Schallom, M. E., and Fraser, V. J. (2002). Effect of an education program on decreasing catheter-related bloodstream infections in the surgical intensive care unit. *Critical Care Medicine*, *30*(1), 59–64.

Cronin, T. G. and M. T. D. (2005). Quantitative Structure–Activity Relationships in the Human Health and Environmental Sciences, Part 2. *SAR and QSAR in Environmental Research*, *16*(1-2), 171–190. doi.org/10.1080/10629360412331319808

Damme, S. Van. (2009). *Quantum chemistry in QSAR Quantum chemical descriptors Use , benefits and drawbacks (Doctoral dissertation,Ghent University Belgium). http://hdl.handle.net/1854/LU-698592*.

Degreef, H. (2008). Clinical Forms of Dermatophytosis (Ringworm Infection). *Mycopathologia*, *166*(October 2007), 257–265. doi.org/10.1007/s11046-008-9101-8

Dehghan, P. (1997). Prevalence and aetiology of dermatophytosis in Isfahan, Iran. *Dermatophytosis in Isfahan, Iran*, *40*, 321–324.

Denning, D. W. (2003). New drug classes Echinocandin antifungal drugs. *The Lancet*, *362*, 1142–1151.

Dhib, I., Fathallah, A., Yaacoub, A., Zemni, R., Gaha, R., and Said, M. B. (2013). Clinical and mycological features of onychomycosis in central Tunisia: a 22 years retrospective study (1986–2007). *Clinical and Mycological Features*, *56*, 273–280. doi.org/10.1111/myc.12016

Diekema, D. J., and Pfaller, M. A. (2004). Nosocomial Candidemia : An Ounce of Prevention Is Better Than a Pound of Cure. *Infection Control and Hospital Epidemiology*, *25*(8), 624–626.

Durdu, M., Ilkit, M., Tamadon, Y., Tolooe, A., and Rafati, H. (2017). Topical and systemic antifungals in dermatology practice. *Expert Review of Clinical Pharmacology*, *01751*(2433), 1751–2441. doi.org/10.1080/17512433.2017.1263564

Eggimann, P., Fluckiger, U., and Ruef, C. (2005). Invasive candidiasis : comparison of management choices by infectious disease and critical care specialists. *Intensive Care Med*, *31*, 1514–1521. doi.org/10.1007/s00134-005-2809-8

Ellis, D., Unit, M., and Hospital, C. (2006). Management of cutaneous fungal infections . *Journal of Medical Microbiology*, *55*(16), 1–7.

Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T. D., Mcdowell, R. M., and Gramatica, P. (2003). Regulatory Acceptance of (Q) SARs | Mini-Monograph Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs. *Environmental Health Perspectives*, *111*(10), 1361–1375. doi.org/10.1289/ehp.5758

Falahati, M., Akhlaghi, L., Lari, A. R., and Alaghehbandan, R. (2003). Epidemiology of dermatophytoses in an area south of Tehran , Iran. *Mycopathologia*, *156*, 279–287.

Fotos, P. G., and Paso, E. (1992). Oral candidosis. *Oral Surgery Oral Medicine Oral Pathology*, *74*(1), 41–49.

Fraser, V. J., Jones, M., Dunkel, J., Storfer, S., Medoff, G., and Dunagan, W. C. (1992). Candidemia in a Tertiary Care Hospital : Epidemiology , Risk Factors , and Predictors of Mortality. *Clinic Infectious Diseases*, *15*, 414–421.

Fridkin, S. K. (2005a). Candidemia is Costly — Plain and Simple. *Clinic Infectious Diseases*, *41*(1), 1240–1241.

Fridkin, S. K. (2005b). The Changing Face of Fungal Infections in Health Care Settings. *Clinic Infectious Diseases*, *41*(15), 1455–1460.

Gamboa, A. B., Mendoza, M., Fernández, A., Díaz, E., Candida, D. De, and Caracas, D. (2006). Detection of Candida dubliniensis in patients with candidiasis in Caracas , Venezuela. *Revista Iberoamericana de Micología*, *23*(2), 81–84. doi.org/10.1016/S1130-1406(06)70018-3

Geerlings, P., Proft, F. De, and Langenaeker, W. (2003). Conceptual Density Functional Theory. *Chemical Reviews*, *103*(5), 1793–1874.

Ghannoum, M. A., and Rice, L. B. (1999). Antifungal Agents : Mode of Action , Mechanisms of Resistance , and Correlation of These Mechanisms with Bacterial Resistance Antifungal Agents : Mode of Action , Mechanisms of Resistance , and Correlation of These Mechanisms with Bacterial Resistance. *Clinical Microbiology Reviews*, *12*(4), 501–517.

Gilson, M. K., Sharp, K. A., and Honig, B. H. (1988). Calculating the Electrostatic Potential of Molecules in Solution : Method and Error Assessment. *Journal of Computation Chemistry*, *9*(4), 327–335.

Giusiano, G., Delia, C., Hern, J., Eraso, E., Quind, G., Guardia, C., and Guarro, J. (2008). Terbinafine susceptibility patterns for onychomycosis-causative dermatophytes and Scopulariopsis brevicaulis. *International Journal of Antimicrobial Agents*, *31*, 540–543. doi.org/10.1016/j.ijantimicag.2008.01.023

Golmohammadi, H., and Dashtbozorgi, Z. (2010). Quantitative structure – property relationship studies of gas-to-wet butyl acetate partition coefficient of some organic compounds using genetic algorithm and artificial neural network. *Structural Chemistry*, *21*(6), 1241–1252. doi.org/10.1007/s11224-010-9669-8

González, U., Seaton, T., Bergus, G., and Jacobson, J. (2012). Systemic antifungal therapy for tinea capitis in children (Review). *Cochrane Database of Systematic Reviews*, (8), 1–76.

Gramatica, P. (2002). Beware of q2! *Journal of Molecular Graphics and Modelling*, *20*(4), 269–276. doi.org/10.1016/S1093-3263(01)00123-1

Gramatica, P., and Ester Papa, E. G. (2007). Statistical external validation and consensus modeling : A QSPR case study for KOC prediction. *Journal of Molecular Graphics and Modelling*, *25*(APRIL), 755–766. doi.org/10.1016/j.jmgm.2006.06.005

Graybill, J. R. (1992). Future directions of antifungal chemotherapy. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *14 Suppl 1*, S170–S181.

Gudlaugsson, O., Gillespie, S., Lee, K., Berg, J. Vande, Hu, J., Messer, S., and Diekema, D. (2003). Attributable Mortality of Nosocomial Candidemia , Revisited. *Clinic Infectious Diseases*, *37*(1 (November)), 1172–1177.

Guido, R. V. C., Oliva, G., and Andricopulo, A. D. (2008). Virtual Screening and Its

Integration with Modern Drug Design Technologies. *Current Medicinal Chemistry*, *15*(1), 37–46.

Guo, Y. L., Wang, B., Lee, C., Wang, J., Lee, C., Wang, B., and Wang, J. (1996). Prevalence of dermatoses and skin sensitisation associated with use of pesticides in fruit farmers of southern Taiwan. *Occupational and Environmental Medicine*, *53*, 427–431.

Guttman, L., American, S., Review, S., and Apr, N. (1944). A basis for scaling qualitative. *American Sociology Review*, *9*(2), 139–150.

Hajjeh, R. A., Sofair, A. N., Harrison, L. H., Lyon, M., Arthington-skaggs, B. A., Mirza, S. A., and Warnock, D. W. (2004). Incidence of Bloodstream Infections Due to Candida Species and In Vitro Susceptibilities of Isolates Collected from 1998 to 2000 in a Population-Based Active Surveillance Program Incidence of Bloodstream Infections Due to Candida Species and In Vitro Susc. *Journal of Clinical Microbiology*, *42*(4), 1519–1527. http://doi.org/10.1128/JCM.42.4.1519

Hall, L. B. and L. H. (1990). An Electrotopology-State Index for Atoms in Molecules. *Pharmaceutical Research*, *7*(8).

Havlickova, B., Czaika, V. A., and Friedrich, M. (2008). Epidemiological trends in skin mycoses worldwide. *Superficial Dermatomycoses Worldwide*, *51*((supplement 4)), 2–15.

He, L., and Jurs, P. C. (2005). Assessing the reliability of a QSAR model ' s predictions. *Journal of Molecular Graphics and Modelling*, *23*, 503–523. doi.org/10.1016/j.jmgm.2005.03.003

Hirashima, A., Rafaeli, A., and Kuwano, E. (1999). Three-Dimensional Quantitative Structure-Activity Studies of Octopaminergic Agonists Responsible for the Inhibition of Sex-Pheromone Production in Hercoverpa armigera. *Bioorganic & Medicinal Chemistry*, *7*(11), 2621–2628.

Ibrahim, E. H., Sherman, G., Ward, S., Fraser, V. J., and Kollef, M. H. (2000). The Influence of Inadequate Antimicrobial Treatment of Bloodstream Infections on Patient Outcomes in the ICU Setting. *Clinical Investigation in Critical Care*, *118*(1), 146–155.doi.org/10.1378/chest.118.1.146

Ivakumar, P. M. S., Kailasam, S., Abu, G. E. B., and Ukesh, D. M. (2007). QSAR Studies on Chalcones and Flavonoids as Anti-tuberculosis Agents Using Genetic Function Approximation (GFA) Method, *55*(1), 44–49.

Jaiswal, M., Khadikar, P. V, Scozzafava, A., and Supuran, C. T. (2004). Carbonic anhydrase inhibitors□ : the first QSAR study on inhibition of tumor-associated isoenzyme IX with aromatic and heterocyclic sulfonamides. *Bioorganic & Medicinal Chemistry Letters*, *14*, 3283–3290. doi.org/10.1016/j.bmcl.2004.03.099

Jayakumar Swamy, B.H.M, Nandini, H.G.M and Pramod, N. (2013). Synthesis,

Characterization and Antifungal Activity of Coumarin Carbohydrazide Containing Quinoline Derivatives. *International Journal of Research in Pharmaceutical and Nano Sciences*, *2*(2), 195–202.

Jeng, K. C. G., and Hou, R. C. W. (2005). Sesamin and Sesamolin : Nature ' s Therapeutic Lignans. *Current Enzyme Inhibition*, *1*(1), 11–20.

Johan, P. W., and Sj, R. E. (2005). Chemical studies on antioxidant mechanisms and free radical scavenging properties of lignans. *Organic and Biomolecular Chemistry*, *3*(18), 3336–3347.

Jose, M., Costa-de-oliveira, S., Cavaleiro, C., and Palmeira, A. (2006). Antifungal activity of the essential oil of Thymus pulegioides on Candida , Aspergillus and dermatophyte species. *Journal of Medical Microbiology*, *55*, 1367–1373. doi.org/10.1099/jmm.0.46443-0

Kar, S., and Roy, K. (2011). Development and validation of a robust QSAR model for prediction of carcinogenicity of drugs. *Indian Journal of Biochemistry and Biophysics*, *48*(2), 111–122.

Khaled, K. F. (2011). Modeling corrosion inhibition of iron in acid medium by genetic function approximation method: A QSAR model. *Corrosion Science*, *53*, 3457–3465. doi.org/10.1016/j.corsci.2011.01.035

Khaled, K. F., and Abdel-Shafi, N. S. (2011). Quantitative Structure and Activity Relationship Modeling Study of Corrosion Inhibitors: Genetic Function Approximation and Molecular Dynamics Simulation Methods. *Int. J. Electrochem. Sci*, *6*, 4077–4094. Retrieved from www.electrochemsci.org

Khaled, K. F., and El-Sherik, A. M. (2013). Using Molecular Dynamics Simulations and Genetic Function Approximation to Model Corrosion Inhibition of Iron in Chloride Solutions. *Int. J. Electrochem. Sci. Int. J. Electrochem. Sci*, *8*(8), 10022–10043. Retrieved from www.electrochemsci.org

Khan, F., Yadav, D. K., Maurya, A., and Srivastava, S. K. (2011). Modern Methods & Web Resources in Drug Design and Discovery. *Letters in Drug Design and Discovery*, *8*(5), 469–490.

Kiralj, R., and Ferreira, M. M. C. (2009). Basic Validation Procedures for Regression Models in QSAR and QSPR Studies: Theory and Application. *Journal of Brazil Chemical Society*, *20*(4), 770–787.

Kiyama, K. A., Aruyama, M. M., Amauchi, S. Y., Akashima, Y. N., Akato, T. N., Ago, R. T., and Oba, Y. K. (2007). Antimicrobiological Activity of Lignan : Effect of Benzylic Oxygen and Stereochemistry of 2 , 3-Dibenzyl-4-butanolide and 3 , 4-Dibenzyltetrahydrofuran Lignans on Activity. *Bioscience, Biotechnology, Biochemistry*, *71*(7), 1745–1751. doi.org/10.1271/bbb.70168

Kovalishyn, V., Kopernyk, I., Chumachenko, S., Shablykin, O., Kondratyuk, K., Pil, S.,

and Metelytsia, L. (2014). QSAR studies , design , synthesis and antimicrobial evaluation of azole derivatives. *Computational Biology and Bioinformatics*, *2*(2), 25–32. doi.org/10.11648/j.cbb.20140202.12

Kumar, K. A., Singh, S. K., Kumar, B. S., and Doble, M. (2007). Synthesis , anti-fungal activity evaluation and QSAR studies on podophyllotoxin derivatives. *Central European Journal of Chemistry*, *5*(3), 880–897. doi.org/10.2478/s11532-007-0036-6

Lacy, A., and Kennedy, R. O. (2004). Studies on Coumarins and Coumarin-Related Compounds to Determine their Therapeutic Role in the Treatment of Cancer. *Curr Pharmaceutical Design*, *10*(30), 3797–3811.

Lakshmipathy, D. T., and Kannabiran, K. (2010). Review on dermatomycosis□ : pathogenesis and treatment. *Natural Science*, *2*(7), 726–731.

Laverdiere, M., Hoban, D., Restieri, C., & Habel, F. (2002). In vitro activity of three new triazoles and one echinocandin against Candida bloodstream isolates from cancer patients. *Journal of Antimicrobial Chemotherapy*, *50*(1), 119–123. doi.org/10.1093/jac/dkf074

Leach, A. R., and Gillet, V. J. (2003). Book review, 2003.

Lee, W., Webster, C., and Williams, J. D. (1985). Outbreak of systemic Candida albicans in intensive care unit caused by cross infection. *British Medical Journal (Clinic Reserve Ed)*, *290*(6470), 746–748.

Licciardi, P. V, and Underwood, J. R. (2011). International Immunopharmacology Plant-derived medicines□ : A novel class of immunological adjuvants. *International Immunopharmacology*, *11*(3), 390–398. doi.org/10.1016/j.intimp.2010.10.014

Lin, M. Y., Carmeli, Y., Zumsteg, J., Flores, E. L., Tolentino, J., Sreeramoju, P., and Weber, S. G. (2005). Prior Antimicrobial Therapy and Risk for Hospital-Acquired Candida glabrata and Candida krusei Fungemia□ : a Case-Case-Control Study. *Antimicrobial Agents and Chemotheraphy*, *49*(11), 4555–4560. doi.org/10.1128/AAC.49.11.4555

Lin, S., Schranz, J., and Teutsch, S. M. (2001). Aspergillosis Case-Fatality Rate□ : Systematic Review of the Literature. *Clinic Infectious Diseases*, *32*((1 February)), 358–366.

Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development. *Advanced Drug Delivery Reviews*, *46*, 3–26. Retrieved from www.elsevier.com

Lipsett, P. A. (2004). Clinical Trials of Antifungal Prophylaxis among Patients in Surgical Intensive Care Units□ : Concepts and Considerations. *Clinic Infectious Diseases*, *39*(Suppl 4), 193–199.

Longbottom, J. L., and Pepys, J. (1964). Pulmonary aspergillosis□ : diagnostic and immunological significance of antigens and c-substance in aspergzllus fumzgatus xxxviiI in the serum of patients suffering from pulmonary aspergillosis , was reported by Pepys et al . (1959). The present report deals with further studies of (1) the antigens in different cultures of A . fumigatus and other with pulmonary disorders , with or without evidence of infection by , *88*(3).

Ma, K., Wang, P., Fu, W., Wan, X., Zhou, L., Chu, Y., and Ye, D. (2011). Rtional design of 2-pyrrolinones as inhibitors of HIV-1 integrase. *Bioorganic & Medicinal Chemistry Letters*, *21*, 6724–6727. doi.org/10.1016/j.bmcl.2011.09.054

Maraki, S., Nioti, E., Mantadakis, E., and Tselentis, Y. (2007). A 7-year survey of dermatophytoses in Crete , Greece. *Mycoses*, *50*(April), 481–484. doi.org/10.1111/j.1439-0507.2007.01403.x

Marchisio, V. F., Preve, L., and Tullio, V. (1996). Fungi responsible for skin mycoses in Turin (Italy) Das Erregerspektrurn der Hautmykosen in Turin (Italien). *Mycoses*, *150*, 141–150.

Marco, F., Pfaller, M. A., Messer, S. A., and Jones, R. N. (1998). Activity of MK-0991 (L-743 , 872), a New Echinocandin , Compared with Those of LY303366 and Four Other Antifungal Agents Tested Against Blood Stream Isolates of Candida spp . *Diagnostic Microbiology and Infectious Disease*, *32*(1), 33–37.

Marques, S. A., Tortorano, A. M., Mendes, R. P., Dermatologia, D. De, Medicina, F. De, Unesp, D. B., and Hospital, M. (2000). Mycoses associated with AIDS in the Third World. *Medical Mycology*, *38*(Supplement 1), 269–279.

Marr, K. A., Seidel, K., and White, T. C. (1997). Candidemia in Allogeneic Blood and Marrow Transplant Recipients□ : Evolution of Risk Factors after the Adoption of Prophylactic Fluconazole. *Journal of Infectious Diseases*, *181*(1), 309–316.

Matthews, E. J., and Contrera, J. F. (2007). In silico approaches to explore toxicity end points□ : issues and concerns for estimating human health effects.*Expert Opinion Drug Metabolic Toxicology*, *3*(1), 125–134.

Mcneil, M. M., Nash, S. L., Hajjeh, R. A., Phelan, M. A., Conn, L. A., Plikaytis, B. D., and Warnock, D. W. (2001). Trends in Mortality Due to Invasive Mycotic Diseases in the United States , 1980 – 1997. *Clinic Infectious Diseases*, *33*((1 September)), 641–647.

Mercader, A. G., and Pomilio, A. B. (2012). 2D- and 3D-QSAR Studies of Flavonoids , Biflavones and Chalcones□ : Anti- viral , Antibacterial , Antifungal , and Antimycobacterial Activities. *Anti-Infective Agents*, *10*(1), 41–54.

Miller, D. J., and Mejicano, G. C. (2001). Vertebral Osteomyelitis due to Candida Species□ :Case Report and Literature Review. *Clinic Infectious Diseases*, *33*(July), 523–530.

Millikan, L. E. (2010). Current concepts in systemic and topical therapy for superficial mycoses. *Clinics in Dermatology*, *28*(2), 212–216. doi.org/10.1016/j.clindermatol.2009.12.016

Morrell, M., Fraser, V. J., Kollef, M. H., Morrell, M., Fraser, V. J., and Kollef, M. H. (2005). Delaying the Empiric Treatment of Candida Bloodstream Infection until Positive Blood Culture Results Are Obtained□ : a Potential Risk Factor for Hospital Mortality Delaying the Empiric Treatment of Candida Bloodstream Infection until Positive Blood Culture. *Antimicrobial Agents and Chemotheraphy*, *49*(9), 3640–3645.doi.org/10.1128/AAC.49.9.3640

Moss, M. (2003). The Epidemiology of Sepsis in the United States from 1979 through 2000. *The New England Journal of Medicine*, *17*(4), 1546–1554.

Mota, S. G. R., and Castilho, M. S. (2009). 2D QSAR Studies on a Series of Bifonazole Derivatives with Antifungal Activity. *Journal of Brazil Chemical Society*, *20*(3), 451–459.

Motta, L. F., and Almeida, W. P. (2011). Quantitative Structure-Activity Relationships (QSAR) of a Series of Ketone Derivatives as anti-Candida albicans. *International Journal of Drug Discovery*, *3*(2), 6–9.

Mun, P., Burillo, A., and Bouza, E. (2000). Criteria used when initiating antifungal therapy against Candida spp . in the intensive care unit. *International Journal of Antimicrobial Agents*, *15*, 83–90.

Murtaza, G., and Mukhtar, M. (2015). A Review□ : Antifungal Potentials of Medicinal Plants. *Journal of Bioresource Management*, *2*(2), 23–31.

Nawale, R. B., and Ashtekar, S. S. (2014). AS ANTIFUNGAL AGENTS. *International Journal of Research in Pharmacy and Chemistry*, *4*(1), 26–33.

Netzeva, T. I., Worth, A. P., Aldenberg, T., Benigni, R., Mark, T. D., Gramatica, P., and Perkins, R. (2005). Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure – Activity Relationships. *ATLA*, *33*, 155–173.

Nicosia, S., Capra, V., and Rovati, G. E. (2001). Leukotrienes as Mediators of Asthma. *Pulmonary Pharmacology and Therapeutics*, *14*(1), 3–19.

Niimi, M., Firth, N. A., and Cannon, R. D. (2010). Antifungal drug resistance of oral fungi. *Odontology*, *98*, 15–25. doi.org/10.1007/s10266-009-0118-3

Noverr, M. C., Noggle, R. M., Toews, G. B., and Huffnagle, G. B. (2004). Role of Antibiotics and Fungal Microbiota in Driving Pulmonary Allergic Responses. *Infection and Immunity*, *72*(9), 4996–5003. doi.org/10.1128/IAI.72.9.4996

Noverr, M. C., Phare, S. M., Toews, G. B., Coffey, M. J., and Huffnagle, G. B. (2001). albicans Produce Immunomodulatory Prostaglandins. *Infection and Immunity*, *69*(5), 2957–2963. doi.org/10.1128/IAI.69.5.2957

Odds, F. C. (2003). Antifungal agents : their diversity and increasing sophistication. *Mycologist*, *17*(May), 51–55. doi.org/10.1017/S0269915X03002064

Ojala, T. (2001). *Biological Screening of Plant Coumarins*. Retrieved from https://www.google.com.ng/search?tbm=bks&hl=en&q=Ojala+T+(2001)+Biological+Screening+of+Plant+Coumarins.+PhD+Thesis,+University+of+Helsinki,+Helsinki,+Finland.

Oprea, T. I., and Matter, H. (2004). Integrating virtual screening in lead discovery. *Current Opinion in Chemical Biology*, *8*(4), 349–358. doi.org/10.1016/j.cbpa.2004.06.008

Pappas, P. G., Kauffman, C. A., Andes, D., Benjamin, D. K., Calandra, T. F., Edwards, J. E., and Sobel, J. D. (2009). Clinical Practice Guidelines for the Management of Candidiasis : 2009 Update by the Infectious Diseases Society of America,*11*, 503–535.doi.org/10.1086/596757

Pappas, P. G., Kauffman, C. A., Andes, D. R., Clancy, C. J., Marr, K. A., Ostrosky-zeichner, L., and Sobel, J. D. (2016). Clinical Practice Guideline for the Management of Candidiasis : 2016 Update by the Infectious Diseases Society of America. *Clinic Infectious Diseases*, *62*(4), 409–417. doi.org/10.1093/cid/civ1194

Pappas, P. G., Rex, J. H., Sobel, J. D., Filler, S. G., Dismukes, W. E., Walsh, T. J., and Edwards, J. E. (2004). Guidelines for Treatment of Candidiasis. *Clinical Infectious Diseases*, *38*(15 January), 161–189.

Patel, H. M., Noolvi, M. N., and Sharma, P. (2014). Quantitative structure – activity relationship (QSAR) studies as strategic approach in drug discovery. *Medicinal Chemistry Research*, *23*(12), 4991–5007. doi.org/10.1007/s00044-014-1072-3

Pathare, S. S., Bhansali, S. G., Mahadik, K. R., and Kulkarni, V. M. (2015). Pharmacophore modeling and atom-based 3D-QSAR studies of antifungal benzofurans. *International Journal of Pharmacy and Pharmaceutical Sciences*, *7*(3), 453–458.

Pearlman, R. S., and Smith, K. M. (1999). Metric Validation and the Receptor-Relevant Subspace Concept. *Journal of Chemical Information and Computer Sciences*, *39*(1), 28–35.

Peter G. Pappas, Barbara D. Alexander, David R. Andes, Susan Hadley, Carol A. Kauffman, Alison Freifeld, Elias J. Anaissie, Lisa M. Brumble, Loreen Herwaldt, James Ito, Dimitrios P. Kontoyiannis, G. Marshall Lyon, Kieren A. Marr, Vicki A. Morrison, Benjam, and T. M. C. (2010). Invasive Fungal Infections among Organ Transplant Recipients: Results of the Transplant-Associated Infection Surveillance Network (TRANSNET). *Clinical Infectious Diseases*, *50*(8), 1101–1111. doi.org/10.1086/651262

Pfaller, M. A., Diekema, D. J., Pfaller, M. A., and Diekema, D. J. (2007). Epidemiology of Invasive Candidiasis : a Persistent Public Health Problem Epidemiology of

Invasive Candidiasis : A Persistent Public Health Problem. *Clinical Microbiology Reviews*, *20*(1), 133–163. doi.org/10.1128/CMR.00029-06

Pfaller, M. A., Pappas, P. G., and Wingard, J. R. (2006). Invasive Fungal Pathogens : Current Epidemiological Trends. *Clinical Infectious Diseases*, *43*(Suppl 1), 3–14.

Piarroux, R., Grenouillet, F., Balvay, P., and Tran, V. (2004). Assessment of preemptive treatment to prevent severe candidiasis in critically ill surgical patients. *Critical Care Medicine*, *32*(12), 2443–2449. doi.org/10.1097/01.CCM.0000147726.62304.7F

Pinheiro, a. a C., Borges, R. S., Santos, L. S., and Alves, C. N. (2004). A QSAR study of 8.O.4′-neolignans with antifungal activity. *Journal of Molecular Structure: THEOCHEM*, *672*(1-3), 215–219. doi.org/10.1016/j.theochem.2003.11.027

Puranik, P., and Kulkarni, Æ. V. M. (2009). QSAR analysis of N-myristoyltransferase inhibitors : antifungal activity of benzofurans. *Medicinal Chemistry Research*, *18*(July), 206–220. doi.org/10.1007/s00044-008-9120-5

Puzyn, T., Leszczyński, J., and Cronin, M. (2010). *Recent Advances in QSAR Studies: Methods and Applications*. *Springer Science and Business Media*. doi.org/10.1007/978-1-4020-9783-6

Rathod, A. K. (2011). Antifungal and Antibacterial activities of Imidazolylpyrimidines derivatives and their QSAR Studies under Conventional and Microwave-assisted. *International Journal of PharmTech Research*, *3*(4), 1942–1951.

Rentz, A. M., Halpern, M. T., & Bowden, R. (1998). The Impact of Candidemia on Length of Hospital Stay , Outcome , and Overall Cost of Illness. *Clinical Infectious Diseases*, *27*(October), 781–788.

Rohini, K and Srikumar, P. (2014). Therapeutic Role of Coumarins and Coumarin-Related Compounds. *Journal of Thermodynamic and Catalysis*, *5*(2), 5–7. doi.org/10.4172/2157-7544.1000130

Sardari, S., Mori, Y., Horita, K., and Micetich, R. G. (1999). Synthesis and Antifungal Activity of Coumarins and Angular Furanocoumarins. *Bioorganic and Medicinal Chemistry*, *7*, 1933–1940.

Satoshi Yamauchi, Tomoko Ina, Takuya Kirikihira, and T. M. (2004). Synthesis and Antioxidant Activity of Oxygenated Furofuran Lignans. *Bioscience, Biotechnology, Biochemistry*, *68*(1), 183–192. doi.org/10.1271/bbb.68.183

Seebacher, C.,Bouchara, J. and Mignon, B. (2008). Updates on the Epidemiology of Dermatophyte Infections. *Mycopathologia*, *166*, 335.doi.org/10.1007/s11046-008-9100-9

Segal, E. (2005). Candida , still number one – what do we know and where are we going from there? *Mycoses*, *48*(Supplement 1), 3–11.

Seifert, M. H. J., Wolf, K., and Vitt, D. (2003). Virtual high-throughput in silico screening. *Biosilico*, *1*(4), 143–149.

Shan, B., Cai, Y., Brooks, J. D., and Corke, H. (2007). The in vitro antibacterial activity of dietary spice and medicinal herb extracts, *117*, 112–119. doi.org/10.1016/j.ijfoodmicro.2007.03.003

Shapiro, S., Guggenheim, B., and Zu, C.-. (1998). Inhibition of Oral Bacteria by Phenolic Compounds . Part 2 . *Quantitative Structural Activity Relationship*, *17*, 338–347.

Sharma, V., Kumawat, T. K., Sharma, A., and Seth, R. (2015). Dermatophytes□ : Diagnosis of dermatophytosis and its treatment. *African Journal of Microbiology Research*, *9*(19), 1286–1293.doi.org/10.5897/AJMR2015.7374

Singh, S., Khadikar, P. V, Scozzafava, A., Supuran, C. T., Control, P., and Bioinorganica, C. (2009). QSAR studies for the inhibition of the transmembrane carbonic anhydrase isozyme XIV with sulfonamides using PRECLAV software. *Journal of Enzyme Inhibition and Medicinal Chemistryp*, *24*(2), 337–349. doi.org/10.1080/14756360802187489

Sobel, J. D. (2003). Management of Patient with Recurrent Vulvovaginal Candidiasis. *Drugs*, *63*(11), 1059–1066.

Sobel, J. D., and Rex, J. H. (2001). Invasive Candidiasis□ : TurningRisk into a Practical Prevention Policy□ ?*Clinical Infectious Diseases*, *33*(15 july), 187–190.

Stanton, D. T., and Jurs, P. C. (1990). Development and Use of Charged Partial Surface Area Structural Descriptors in Computer-Assisted Quantitative Structure-Property Relationship Studies. *Analytical Chemistry*, *62*(21), 2323–2329.

Tong, W., Hong, H., Xie, Q., Shi, L., Fang, H., and Perkins, R. (2005). Assessing QSAR Limitations – A Regulatory Perspective. *Current Computer - Aided Drug Design*, *1*(2), 195–205.

Traynor, T. R., and Huffnagle, G. B. (2001). Role of chemokines in fungal infections. *Medical Mycology*, *39*(July), 41–50.

Ursu, O., Costescu, A., Diudea, M. V, and Parv, B. (2006). QSAR Modeling of Antifungal Activity of Some Heterocyclic Compounds. *Croatica Chemica Acta*, *79*(3), 483–488.

Vadlamudi, S. M., and Kulkarni, V. M. (2003). 3D-QSAR of Protein Tyrosine Phosphatase 1B Inhibitors by Genetic Function Approximation 3D-QSAR of Protein Tyrosine Phosphatase 1B Inhibitors by Genetic Function Approximation. *Internet Electronic Journal of Molecular Design*, *2*(October), 1–28. doi.org/10.1080/14756360601051274

Vermes, A., Guchelaar, H., and Dankert, J. (2000). JAC Flucytosine□ : a review of its

pharmacology , clinical indications , pharmacokinetics , toxicity and drug interactions. *Journal of Antimicrobial Chemotherapy*, 171–179.

Vianna, D. R., Hamerski, L., Figueiró, F., Bernardi, A., Visentin, L. C., Pires, E. N. S., and Pinto, A. C. (2012). Selective cytotoxicity and apoptosis induction in glioma cell lines by 5-oxygenated-6 , 7-methylenedioxycoumarins from Pterocaulon species. *Europea Journal of Medicinal Chemistry*, *57*, 268–274. doi.org/10.1016/j.ejmech.2012.09.007

Vincent, J.L. Anaissie, E., and Bruining, H. (1998). Epidemiology, diagnosis and treatment of systemic Candida infection in surgical patients under intensive care. *Intensive Care Med*, (24), 206.doi.org/10.1007/S001340050552

Votano, J. R., Parham, M., Hall, L. H., and Kier, L. B. (2004). New predictors for several ADME / Tox properties : Aqueous solubility , human oral absorption , and Ames genotoxicity using topological descriptors. *Molecular Diversity*, *8*, 379–391.

Wang, J., Krudy, G., Xie, X., Wu, C., and Holland, G. (2006). Genetic Algorithm-Optimized QSPR Models for Bioavailability , Protein Binding , and Urinary Excretion. *Journal of Chemical Information and Modeling*, *46*, 2674–2683.

Wey, S. B., Mori, M., Pfaller, M. A., Woolson, R. F., and Wenzel, R. P. (2015). Hospital-Acquired Candidemia The Attributable. *Clinical Epidermiology*, *148*, 2642–2645.

White, T. C., Oliver, B. G., Gra, Y., & Henn, M. R. (2008). Generating and Testing Molecular Hypotheses in the Dermatophytes. *Eukaryotic Cell*, *7*(8), 1238–1245. doi.org/10.1128/EC.00100-08

Wilson, L. S., Reyes, C. M., Stolpman, M., Speckman, J., Allen, K., and Beney, J. (2002). The Direct Cost and Incidence of Systemic Fungal Infections. *Value in Health*, *5*(1), 26–34.

Wisplinghoff, H., Bischoff, T., Tallent, S. M., Seifert, H., Wenzel, R. P., and Edmond, M. B. (2004). Nosocomial Bloodstream Infections in US Hospitals : Analysis of 24 , 179 Cases from a Prospective Nationwide Surveillance Study. *Clinical Infectious Diseases*, *39*(3), 309–317.

Wu, J., Pan, J., Chen, S., Yang, M., Wu, J., and Zou, K. (2009). An update on lignans : natural products and synthesis. *The Royal Society of Chemistry*, *26*(10), 1251–1292.doi.org/10.1039/b910940d

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., and Dan, J. H. (2008). *Top 10 algorithms in data mining*. doi.org/10.1007/s10115-007-0114-2

Yamauchi, S., Hayashi, Y., Nakashima, Y., Kirikihira, T., and Yamada, K. (2005). Effect of Benzylic Oxygen on the Antioxidant Activity of Phenolic Lignans. *Journal of Natural Products*, *68*(10), 1459–1470.

Yanai, T., Tew, D. P., and Handy, N. C. (2004). A new hybrid exchange – correlation functional using the. *Chemical Physics Letters*, *393*(July), 51–57. doi.org/10.1016/j.cplett.2004.06.011

Zheng, W., and Tropsha, A. (2000). Novel Variable Selection Quantitative Structure - Property Relationship Approach Based on the k -Nearest-Neighbor Principle. *Journal of Chemical Information and Computer Science*, *40*(1), 185–194.