Modeling the aromatase inhibitor activity of indole-imidazole derivatives: Quantitative structure activity relationship and molecular docking study

نمذجة نشاط تثبيط الأروماتيز لمشتقات إندول-إيميدازول: دراسة كمية لعلاقة البنيات بالفعالية ودراسة الترسية الجزيئية

A dissertation submitted in partial fulfillment for the requirement of M.sc degree in chemistry

BY

Rawan Mustafa Ali Massad

(B.Sc in chemistry)

Supervisor

Prof. Dr. Ahmed Elsadig Mohammed Saeed

March, 2021

Holy Quran Verse

قال تعالى:

((اقْرَأْ بِاسْمِ رَبِّكَ الَّذِي خَلَقَ * خَلَقَ الْإِنْسَانَ مِنْ عَلَقٍ * اقْرَأْ وَرَبُّكَ الْأَكْرَمُ * الَّذِي عَلَّمَ بِالْقَلَمِ * عَلَّمَ الْإِنْسَانَ مَا لَمْ يَعْلَمْ))

[العلق: 1 - 5]

DEDICATION

To my beloved family and my friends

ACKNOWLEDGEMENT

A lot of thanks to Allah for giving me the strength and health to complete this work.

I would like to express my gratitude to my supervisor: Prof.Dr. Ahmed Elsadig Mohammed Saeed for guiding me during this research and answering my question.

I would also like to express my sincere appreciation to Dr. Tawasul Taj Elsir Hassan Hag Alsiddig for her invaluable advice and support.

My thanks also go to

Esraa Mohammed for her useful help.

My brother for giving me an internet access which help me a lot in the writing of this research.

Anyone who help me during my work .

Abstract

Aromatase, a cytochrome P450 enzyme complex present in breast tissues, plays a significant role in the biosynthesis of estrogen from androgen. Estrogen is a significant factor in the maintenance and progression of hormone-dependent breast cancer. Therefore, inhibition of aromatase is a promising target for hormone-dependent breast cancer therapy.

In this study in order to obtain the structure requirements for the inhibition of the active sites of aromatase enzyme, the quantitative structure-activity relationships(QSAR) of 19 indole-imidazole derivatives was investigated by the PLS method.

The dataset was split randomly into training set (15 compounds) to build the QSAR model and test set (3 compounds) for the external validation of the model. As a result, a model with only three descriptors (diameter, petitjean, Q_VSA_FPNEG) was found to be robust enough for prediction of the aromatase inhibitor activity of the new designed indole-imidazole derivatives, with an R^2 of 0.892 and Q^2 of 0.741.

A series of a new 112 compounds were modeled and designed, out of these, only 18 compounds were found to have biological activity more than those of letrozole (reference compound); these compounds were docked into the active site of aromatase to understand their inhibition action and their binding energy toward aromatase enzyme.

Analysis of energy of the eighteen compound-aromatase complexes revealed that compound **CVIII** has a low binding energy (strong binding affinity) to the aromatase as compared to letrozole; the energy of this compound is less by 8 units than those of letrozole, this compound is enhanced by electron withdrawing group (COOH) at meta position of the phenyl ring of indole.

المستخلص

الموجود في أنسجة الثدي دورًا مهمًا في التخليق (p450) يلعب الاروماتيز وهو مركب انزيم السيتوكروم هرمون الاستروجين هو عامل مهم في معالجة سرطان الثدي المعتمد. الحيوي للإستروجين من الأندروجين على الهرمونات . لذلك ، فإن تثبيط الأروماتيز هو هدف واعد لعلاج سرطان الثدي المعتمد على الهرمونات.

في هذه الدراسة من أجل الحصول على المتطلبات الهيكلية لتثبيط المواقع النشطة لإنزيم الأروماتيز ، تم فحص العلاقة الكمية للبنية بالفعالية لـ 19 مشتقًا من مشتقات الإندول-إيميدازول بطريقة PLS.

ومجموعة QSAR تم تقسيم مجموعة البيانات عشوائيًّا إلى مجموعة تدريب (15 مركبًا) لبناء نمّوذج اختبار (3 مركبات) للتحقق الخارجي من النموذج. نتيجة لذلك ، تم العثور على نموذج يحتوي على ثلاثة واصفات فقط petitjean ،Q_VSA_FPNEG (القطر ، ليكون قويًا بما يكفي للتنبؤ بنشاط تثبيط 0.741. مقدارها Q² و 0.892 مقداره R² الأروماتيز لمشتقات الإندول-إيميدازول المصممة الجديدة ، مع

تمت نمذجة وتصميم سلسلة من 112 مركبًا جديدًا ،من جميع هذه المركبات وجد أن 18 مركبًا فقط لها نشاط تم وضع هذه المركبات في الموقع النشط. بيولوجي أكثر من تلك الموجودة في ليتروزول (مركب مرجعي) للأروماتيز لفهم تأثير تثبيطها وطاقة ربطها تجاه إنزيم الأروماتيز.

له طاقة ربط CVIII كشف تحليل الطاقة لجميع معقدات الأروماتيز مع الثمانية عشرمركب أن المركب طاقة هذا المركب أقل بمقدار 8 وحدات من. منخفضة (تقارب ارتباط قوي) بالأروماتيز مقارنة مع ليتروزول في COOH)تلك الموجودة في ليتروزول. يتم تعزيز هذا المركب عن طريق مجموعة ساحبة للإلكترون(الموضع ميتا من حلقة الفينيل في الإندول.

TABLE OF CONTENT

Title	page
Holy Quran Verse	I
Dedication	II
Acknowledgment	III
Abstract(English)	IV
Abstract(Arabic)	V
Table of content	VI
List of tables	X
List of figures	XI
List of Appendix	XIII
List of Scheme	XV
List of Abbreviation	XVI

CHAPTER ONE

Introduction

CHAPTER TWO

MATERIALS AND METHODS

2.1	Software programs	19
2.1.1	ACD lab software	19
2.1.2	MOE software	19
2.1.3	SPSS software	19
2.2	Computational method	19
2.2.1	Dataset	19
2.2.2	Molecular descriptor	20
2.2.3	QSAR study	20
2.2.3.1	QSAR model	20
2.2.3.2	Validation of the QSAR model	21
2.2.3.2	Evaluating the performance of QSAR models	21
2.2.4	Modeling of indole-imidazole derivatives	21
2.2.5	Crystal structure from PDB	21
2.2.6	Database generation	22
2.2.7	Molecular docking	22
2.2.7.1	Validation of the molecular docking procedure	22
2.2.7.2	Molecular docking study	22

CHAPTER THREE

DISCUSSION

NO	Title	Page
3.	Discussion	31
3.1	QSAR study	31

CHAPTER FOUR

List of tables

List of figures

List of Appendixes

LIST OF SCHEMES

LIST OF ABBREVIATIONS

RMSE	Root mean square of error
CADD	computer Aided Drug Design
SEE	Standard Error of Estimate
QSAR	Quantitative Structure Activity Relationship
SPSS	Statistical Package for Social Science
RP	Relative Potency
r	Correlation Coefficient
r^2	square of correlation coefficient
PDB	Protein Data Bank
PLS	Partial Least Square
Arg	Arginine
LOO	Leave One Out
MOE	Molecular Operator Environment
AI	Aromatase Inhibitor
AIs	Aromatase inhibitors
Q^2	cross-validated r-squared
SBDD	Structure based drug design
LBDD	Ligand based drug design
ER+	Estrogen receptor-positive
P	Signature of the model
F	Fischer statistic
ACD	Advance chemistry development
IC_{50}	Concentrations corresponding to 50% activity inhibition
RMSD	Root mean square deviation

Introduction

1.1 Introduction

With its origins rooted in organic synthesis and medicinal chemistry, heterocyclic compounds present themselves as a fundamental division of organic chemistry. Defined by IUPAC as "cyclic compounds having as ring member's atoms of at least two different elements".(Martins *et al*., 2015)

The chemistry of heterocyclic compounds is one of the most complex branches of organic chemistry. It is equally interesting for its theoretical implications, for the diversity of its synthetic procedures, and for the physiological and industrial significance of heterocyclic compounds(Hofmann,1953). Thename heterocyclic comes from the Greek word "heteros" which means "different.",the familiar hetero atoms are nitrogen, oxygen and sulphur and other variety of atoms including Se, P, Si, B hetero atoms are also widely known.

Heterocyclic compounds are of enormous interest in our daily life. Amongst the approximately 20 million chemical compounds identified by the end of the second millennium. More than two-thirds are fully or partially aromatic and approximately half are heterocycles. Today, gigantic numbers of heterocyclic compounds are known and this number is increasing rapidly(Narnaware and Shende, 2018).

Compounds classified as heterocyclic probably constitute the largest and most varied family of organic compounds. After all, every carbocyclic compound, regardless of structure and functionality, may in principle be converted into a collection of heterocyclic analogues by replacing one or more of the ring carbon atoms with a different element. Even if the consideration to oxygen, nitrogen and sulphur (the most common heterocyclic elements) were restricted, the permutations and combinations of such a replacement are numerous.(Hote and Bhoyar, 2014).

Heterocycles, as privileged structures in drug discovery, constitute one of the mostsignificant areas of research in medicinal chemistry (Fan *et al*.,2018) Most of the synthetic heterocyclic compounds act as a drug is used asanticonvulsants, hypnotics, antineoplastic, antiseptics, antihistaminics, antiviral, anti-tumour etc. In every year large number of heterocyclic drugs is being introduced in pharmacopeia's(Hossain and Nanda, 2018).

1.1.1History

The history of heterocyclic chemistry began in the 1800s, in step with the development of organic chemistry.

Some noteworthy developments:

1818: Brugnatelli isolates alloxan from uric acid

1832: Dobereiner produces furfural (afuran) by treating starch with sulfuric acid

1834: Runge obtains pyrrole ("fiery oil") by dry distillation of bones

1906:Friedlander synthesizes indigo dye, allowing synthetic chemistry to displace a large agricultural industry

1936: Treibs isolates chlorophyll derivatives from crude oil, explaining the biological origin of petroleum.

1951: Chargaff's rules are described, highlighting the role of heterocyclic compounds (purines and pyrimidines) in the genetic code (Aljamali,2015)

1.2 Indole

Indole (Figure 1.1) is the parent substance of a large number of important compounds that occur innature (Kaushik *et al.*, 2013), an indole is an aromatic heterocyclic (El-Sayed *et al.*, 2015).It is a planarmolecule comprising of benzene ring fused to a pyrrole ring at -2 & -3 position(Naim *et al.*, 2016). It is one of the most abundant heterocycles found in natural products and biologically active molecules. In fact, it can be regarded as the most important of all the privileged structures in medicinal chemistry.
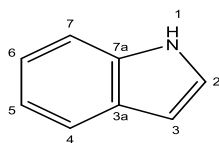
indole

Figure 1.1 Chemical structure of 1H-indole

The discovery and structure elucidation of indole dates from 1866, when Adolf vonBaeyer synthesized indole by zinc-dust pyrolysis of oxindole, which had been obtained by reduction of isatin, a product of the oxidation of the natural blue pigment Indigo.Consequently, the name Indole derives from thatof Indigo(Barluenga and Valdes, 2011). Chemical structure of oxindole, indigo and isatin was shown in figure1.2.
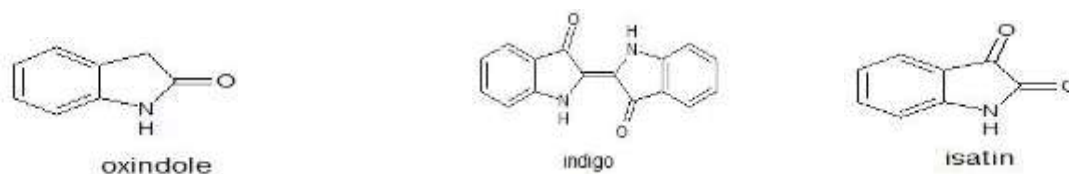
Figure 1.2 Chemical structure of oxindole, indigo and isatin

The indole ring system is probably one of the most ubiquitous heterocycles in nature that occurs in many biologically natural and synthetic compounds. Indole is a popular component of fragrances and the precursor to many pharmaceuticals. Indoles are among the most abundant and important classes of *N*-heterocycles. (Geovanni *et al*., 2010),it is a heterocyclic system with 10 electrons (Sharma *et al*., 2010)So it is electron rich system. It is resonance energy is 47-49 Kcal/mole. The electrons on nitrogen are involved in aromatic sextet; hence it is very weak base with pKa value -3.5 (Sundberg, 2000)

1.2.1 indole synthesis

The synthesis of indoles has been an active research field due to their structural diversity as well as numerous applications of natural and synthetic indole derivatives. (Hao *et al*., 2014)

Indoles are usually prepared from non - heterocyclic precursors by cyclisation reactions on suitably substituted benzenes; they can also be prepared from pyrroles by construction of the homocyclic aromatic ring, and from indolines by dehydrogenation(Joule and Mills, 2010) and those are the most used method in the synthesis of indole:

1.2.1.1General method for Indole synthesis

Indole syntheses almost universally involve annelation of the five-membered ring to an existing benzene ring bearing the appropriate functionality. These can be divided into those in which two substituents in an *ortho* relationship are connected together (Scheme 1.1 A), and those in which a single substituent is cyclized directly onto the aromatic ring (Scheme 1.1B) (Inman and Moody, 2013).

A.

X = NH2, NHCOR, NO$_2$

$Y = I, CH_3, CH_2R^2, COR^2, CH_2CH=CH_2, C=CR^1$

B.

$Z = NHNH_2, NH_2, NHR, NO_2, CHO,$ halide

Scheme 1.1 General methods for indole synthesis

1.2.1.2 Fischer indole synthesis:

Fischer indole synthesis (Scheme 1.2)is one of the oldest and most convenient, and it is currentlyused for a broad range of applications.(Murakami, 2012)and it was represented as one of the most powerful and versatile routes for the synthesis of indole heterocycles(Hao *et al.*, 2014).This reaction was first reported by Fischer and Jourdan in 1883(Wang, 2010)by treatment of pyruvic acid 1-methylphenylhydrazone with alcoholic hydrogen chloride.(Robinson,1963)
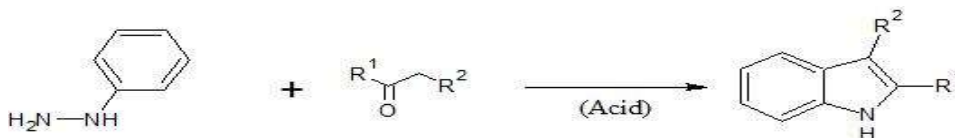
Scheme 1.2 Fischer indole synthesis

1.2.1.3 Bichler Indole synthesis

First reported in 1892, the Bischler indole synthesis(Scheme1.3) involves alkylation of an aniline with anα-haloketone, followed by acid-catalyzed ring closure(Inman and Moody, 2013)
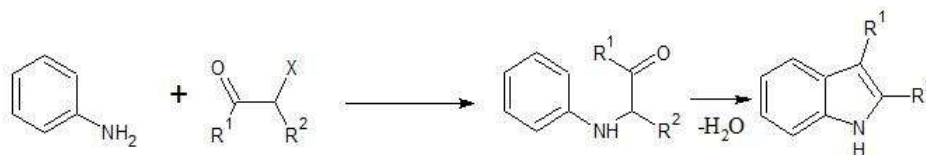
Scheme 1.3 Bischler indole synthesis

1.2.2 Physical propertise of indole

Indole is a colourless crystalline solid in the form of the shinning leaflets.it melts at 52°C, and boil at 253°C. It is voatile with steam.It is soluble in hot water and in

hot alcohol,and is also soluble in ether and in benzen.The compound in an impure state has an unpleasant odor.(Vanolder and Lindwall,1942)

1.2.3 Chemical properties of indole

1.2.3.1 Electrophilic aromatic Substitution

The π-excessive character of the pyrrole ring makes the indole ring susceptible to electrophilic attack(Sundberg, 2000). Unlike pyrrole, addition of electrophiles takes place preferentially at C3. A simple explanation for this can be deduced by analysis of the Wheland intermediates resulting from the attack of a nucleophile at C2 and C3(Barluenga and Valdes, 2011).

1.2.3.1.1Halogenation

3-Chloroindole can be obtained by chlorination with either hypochlorite ion or with sulfuryl chloride. In the former case the reaction proceeds through a 1-chloroindole intermediate. 3-Chloroindole is quite unstable to acidic aqueous solution, in which it is hydrolyzed to oxindole. 3-Bromoindole has been obtained from indole using pyridinium tribromide as the source of electrophilic bromine. Indole reacts with iodine to give 3-iodoindole. Both the 3-bromo and 3-iodo compounds are susceptible to hydrolysis in acid but are relatively stable in base(Sundberg, 2000).

1.2.3.1.2Sulfonation

Sulfonation of indole, at C - 3, is achieved using the pyridine – sulfur trioxide complex in hot pyridine. Gramine is sulfonated in oleum to give 5 - and 6 - sulfonic acids, attack being on a diprotonated (C - 3, side chain - N) salt(Joule and Mills, 2010).

1.2.3.1.3N-Alkylation

1-Substitution is favoured when the indole ring is deprotonated and the reaction medium promotes the nucleophilicity of the resulting indole anion. Conditions which typically result in *N*-alkylation are generation of the sodium salt by sodium amide in liquid ammonia, use of sodium hydride or a similar strong base in *N, N*-dimethylformamide or dimethyl sulfoxide, or the use of phase-transfer

conditions(Sundberg,2000).

1.3 Imidazole

The name "imidazole" was coined in 1887 by the German chemist Arthur Rudolf Hantzsch (1857–1935)(Manocha *et al.*,2016). It is a planar cyclic molecule consisting of a five membered ring. It contains two nitrogens and three carbons, with the nitrogens occupying the first and third positions (Sujatha *et al.*, 2016). The chemical structure of imidazole was shown in Figure1.3.
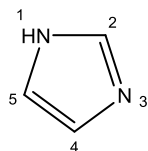
imidazole

Figure 1.3 Chemical structure of imidazole

The systemic name for the compound is 1, 3 diazole, one of the annular N bear a H atom and can be regarded as a pyrrole type. N. (Jawaharmal *et al.*,2012) The solubility of imidazole is high in polar and low in nonpolar solvent(Hofmann,1953).These compounds are aromatic, each carbon atom and one nitrogen atomparticipating with one electron to the conjugated p system while the other heteroatomparticipates with a lone pair of electrons(Revuelta *et al.*, 2011).

It exists in two equivalent tautomeric forms because the hydrogen atom can be located on either of the two nitrogen atoms(Verma et al., 2013;Zhang *et al.*, 2013) Furthermore, the electron-rich nitrogen heterocycle could not only readily accept or donate proton, but also easilyform diverse weak interactions. (Zhang *et al.*, 2013)

It is amphoteric compound (Pozharskii et al., 1966) susceptible to electrophilic and nucleophilic attack(Ghosh and Biplab, 2013; Romero *et al.*, 2014).This property is explained by the resonance interactions, which increases the basicity of the two N-atom.Resonance structures of imidazole are shown in Figure1.4 (Daraji *et al.*, 2019).
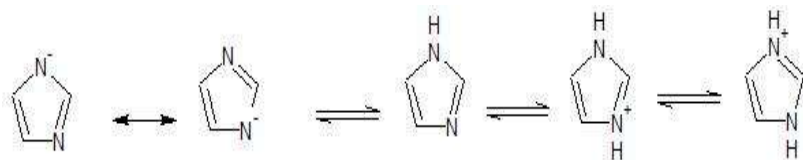
Scheme 1.4 Resonance structure of imidazole

1.3.1 Synthesis of imidazole
1.3.1.1 Debus synthesis

Imidazole was first reported for Debus *et al.*, in 1858 from glyoxal and formaldehyde (Ali *et al.*, 2017) . This reaction provide 2-mono substituted and 2, 4, 5- homo tri substitutedImidazole. (Baroniya *et al.*, 2010).Debus reaction was shown in Scheme 1.5
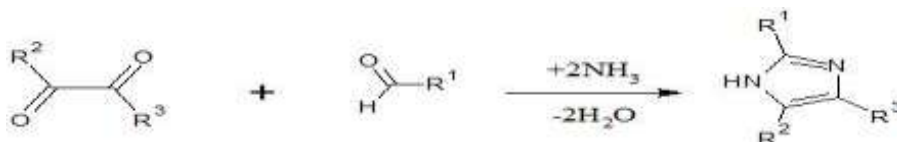
Scheme 1.5 Debus synthesis of imidazole

1.3.1.2 Radziszewski Synthesis

An imidazole synthesis involving the condensation of a dicarbonylcompound with an aldehyde and ammonia (Scheme 1.6) was discovered almost simultaneouslyby Japp and Radziszewski (Hofmann, 1953)
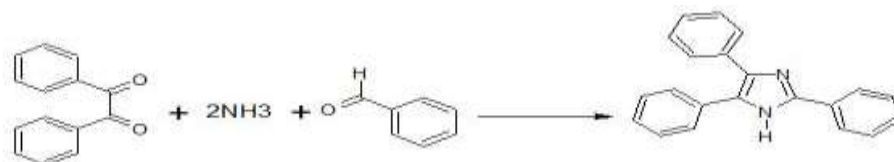
Scheme 1.6 Radziszewski synthesis of imidazole

1.3.1.3 Van Leusen synthesis

Van Leusen imidazole synthesis (Scheme 1.7) based on TosMICs (tosylmethylisocyanides), which is the cycloaddition reaction, is one of the most convenient and attractive protocols for the preparation of imidazole-based small molecules, due to its excellent advantages like simple manipulation, easily obtained raw materials and a wide range of substrates, which has been developed rapidly in the past decades (Zheng *et al.*, 2020).

Scheme 1.7 The first example of Van Leusen imidazole synthesis

1.3.1.4 Wallach synthesis

The cyclization ofN, N′disubstituted oxamides with PCl_5 (Scheme 1.8) to afford 1-substituted5-chloroimidazoles (Revuelta *et al*., 2011). Chlorine containing compound is obtained which on reduction with hydroiodic acid give N- methyl imidazole. Under the same condition N, N-diethyloxamide is converted to a chlorine.compound, which on reduction gives 1- ethyl –2- methyl imidazole. (Chawla*et al*., 2012).
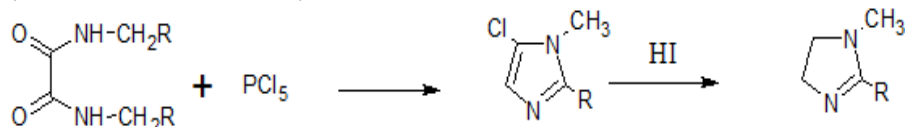
Scheme 1.8 Wallach synthesis of imidazole

1.3.2 Physical properties

Imidazole is a colourless organic compound having melting point 89-91 °C and boiling point is 256 °C. It has high boiling point as compared by the other five membered heterocyclic compounds. In marked contrast to imidazole, the boiling point of 1- methyl imidazole is comparatively low. It demonstrates that hydrogen bonding exists in imidazole ring and may consists up to 20 molecules(Baroniya *et al*.,2010).

1.3.3 Chemical properties of imidazole

1.3.3.1 Reaction with acids

Imidazole is a mono acidic base. It forms crystalline salts with acids and also possesses weakly acidic property(Chaudhury *et al*., 2015). The reaction of imidazole with acid was shown in Scheme 1.9
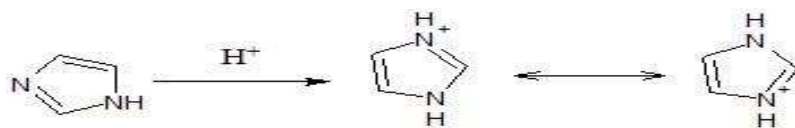
Scheme 1.9 Reaction of imidazole with acids

1.3.3.2 **Phosphorylation**

N-phosphorylimidazoles (Scheme 1.10) play an important role in enzymic transphosphorylation.They form in the reaction of two moles of imidazole orbenzimidazole with one mole of a dialkyl(diaryl)phosphoric acid chloride. (Pozharskii *et al*., 1966).
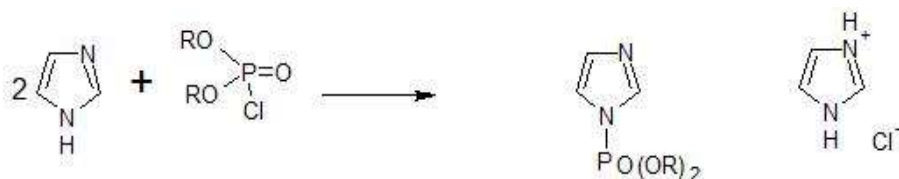
Scheme1.10Phosphorylation of imidazole

1.3.3.3 **Quaternization**

Quaternization (Scheme 1.11) of the nitrogen atom of the imidazole is normally achieved by the reaction of alkyl halides or dialkyl sulfates in an organic solvent under strongly basic conditions. Alkylation of imidazoles is achieved by heating 1-carboethoxyimidazoles.(Baroniya*et al*., 2010)
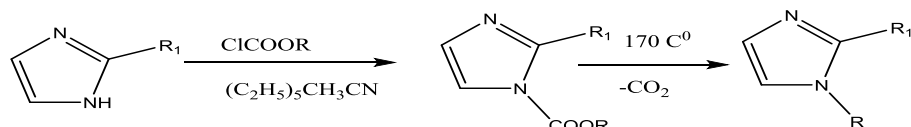
Scheme 1.11 Quaternization of imidazole

1.4 Indole and Imidazole biological activity

Constructing novel pharmacologically interesting hybrid compounds for drug discovery was an important strategy in designing novel active leads (Yousif *et al*., 2019).Combination of structural features of two or more functionally active substances into one molecule may lead to synergism, enhancement or modulation of the desired characteristics of individual components(Singla *et al*., 2018).

Imidazole and indole residue are probably the most well- known heterocycles which are common and important features of a variety of natural products and medicinal agent (Benkli *et al*.,2004). It is evident from the literature that conjugated indole-imidazole derivatives known to be associated with a broad spectrum of biological activities(Naureen *et al*., 2019). However**,** biological

activity of the indole and azoleconjugates are lack of investigation(Sumiya *et al*., 2017)

The imidazole heterocycle which acts as a structural subunit of more complex natural products and drugs is ubiquitous in nature (Fan *et al*., 2018).

Structural characteristics of imidazole ring arebeneficial for its derivatives to readily bind with a variety of enzymes and receptors in biological systems via hydrogen bonds, coordination, ion–dipole, cation–π, π–π stacking, hydrophobic effects, van derWaals forces, and so on, thereby exhibiting broad bioactivities(Zhang *et al*., 2013)

Imidazole and its derivatives have gained remarkable importance due to their widespread biological activities and their use in synthetic chemistry. Imidazole derivatives possess a broad spectrum of pharmacological activities (Vijesh *et al*., 2013) such as,antifungal agent (Odds,1980; Uno *et al*., 1983; Khabnadideh *et al*.,2009; Yakovychuk *et al*., 2018),anti-tubercular agents (Fan *et al*.,2018),antiHIV agent(;Rashamuse *et al*,2019 Rashamuse *et al*.,2020),antimicrobial agent(Valdez *et al*., 2009; Wunk *et al*., 2013; El-aal *et al*., 2015), anti-Inflammatory activity (Che *et al*., 2010),antiplasmodial activity (Kondaparla *et al*., 2018),anticonvulsant activity (Nardi et al., 1981;Mariappan *et al*., 2013) antibacterial activity (Gomleksiz*et al*., 2013; Sun *et al*.,2017),anticancer agent(Gore,2016;Elahian *et al*., 2014; Ali et al.,2017;Zhang *et al*.,2019),antimycobacterial activity (Zampieri *et al*.,2007),antineoplastic activity (Zhao et al.,2018)**,**antispasmodic and antidiarrheal activities (Lakshmanan *et al*.,2011), and antichagasic agents(de Araújo *et al*., 2019)Also indole nucleus is well known for its pharmaceutical and biochemical activities (El-ezbawy *et al*.,1989), and it has been becoming an important structural component in many pharmaceutical agent (Liu *et al*., 2018).Both naturally occurring and synthetic indole containing molecules produced by many synthetic approaches have important uses and potential as drugs with a broad spectrum of applications in many therapeutic categories, including non-steroidic anti-inflammatory, (Palmisano *et al*.,2010)antibacterial activity(Mielczarek *et al*., 2014; Oh *et al*., 2016; Sewalad and Shaaban et al., 2017;Hong *et al*., 2017),antimicrobial activity(El-zebawy *et al*.,1989;Quazi   *et al*.,2017;Natraj *et al.,*2010),antifungal activity(Ashok   *et al*.,2013),anti-inflammatory   activity(Badiger   *et   al*., 2009),antimalarial activity(Yadav *et al*., 2016).

Heterocyclic compounds appear to be most effectiveagainst various cancers.Around 60% of the medicationsused for cancer are based on heterocyclic

moieties. Amongthe various heterocyclic moieties, nitrogen-based compounds have been found to be effective against different types of cancers (Ali *et al*., 2018)

Cancer is one of the crucial public health problems in the world. The corresponding incidence and mortality statistics shows that it is growing in the economically developing and developed countries(Liu *et al*., 2018). It is the most complex disease with many types and subtypes affecting different body organs (Patil *et al*., 2017).Actually, the inducement of cancer is greatly complicated, involving many molecular mechanisms(Wang *et al*., 2015).

As a compound incorporating both imidazole and indole motifs, Trachycladindole(Figure 1.4) is an indolyl-imidazole derivative exhibiting potent cytotoxic activities against leukemia cells and lung cancer cells(Yousif *et al*.,2019).

Figure 1.4 Chemical structure of Trachycladindole

Many research revealed that imidazole derivatives as anticancer drugs possess considerable potentiality. Generally, imidazole could interfere with DNA synthesis via weak interactions such as hydrogen bonds, coordination bonds, and $\pi-\pi$ stacking and then halt cell growth and division. Meanwhile, imidazole could easily bind to protein molecules compared with the other heterocyclic rings.Furthermore, some imidazole drugs, at high concentrations, could directly inhibit the synthesis of essential cell membrane components without interference with sterols and sterol esters(Zhang *et al*., 2013).

Breast cancer is the second most prevalent cancer worldwide(Bae *et al*., 2018).It is still the most frequently identified cancer in women(Semire and Oyebamiji, 2017). It continues to be a major cause of cancer deaths(Saberi *et al*., 2005).It is also a predominantly post-menopausal disease (Benson and Raviskar, 2007).

Approximately two-thirds of breast cancers are hormone-dependent(Gobbi *et al*., 2014). This hormone dependency of breast cancer interacts with environmental and genetic factors to determineincidence and progression of the disease.(Benson and

Raviskar, 2007).It is widely accepted that estrogen plays an important role inthe genesis and evolution of breast tumors (Wang *et al*., 2013).

Aromatase is a cytochrome P-450 dependent enzyme (Xie *et al*., 2014; Pingaew *et al*., 2017) that catalyzes the aromatization of androgens to estrogens(Xie *et al*.,2014; Marchand *et al*., 2003)and is a particularly attractive target in the treatment of estrogen receptor positive breast cancer. Inhibitors of this enzyme are potential therapeutics for estrogen dependent breast cancers.Aromatase inhibitors can be both steroidal and non-steroidal compounds (Roy and Roy, 2010).

Steroidal aromatase inhibitor(AI) (e.g., exemestane in Fig.1.5) derived from the substrate androstenedione interacts with aromatase through chemical actions, resulting in an irreversible binding process of the species; while non-steroidal AI (e.g., anatrozole and letrozole in Fig1.5) binds to enzyme through non-covalent interactions, resulting in a reversible binding process(Kang *et al*.,2018).

Exemestane   Anastrozole   Letrozole

Figure 1.5 Chemical structures for steroidal (exemestane) and nonsteroidal (anastrozole, letrozole) aromatase inhibitors(AIs).

Exploration of the binding characteristics of aromatase inhibitors in the active site as well as the properties important for binding, are of importance in designing more selective aromatase inhibitors(Roy and Roy, 2010).

The indole moiety accounts for fitting the binding site of aromatase, (Kang *et al*., 2018) and one of the most important feature for strong inhibitor binding to CYP enzymes is the capability to interact as the sixth ligand with the iron atom of the haem group, always present in this enzyme family. This coordination is excellently performed by the lone pair carried on the sp2 hybridized imidazole nitrogen but other electronrich heterocycles might be accepted in this position as well(Castellano *et al*., 2008).

Many indoles containing-molecules have been reported to display potent aromatase inhibitory activity (Pingaew *et al*., 2017).Some indole derivatives such as azolylmethyl-1H-indoles, azolylbenzyl-1H-indoles,5-aryl (imidazol-1-yl) methyl-

1H-indoles, and 4-aroylindoles,22 indole-3-carbinols, indole-imidazoles, and 2- and 3-aryl(azolyl)methylindoles have been previously synthesized and evaluated as potential anti-aromatase agents.(Prior *et al*., 2017) Le Borgne et al. reported letrozole analogues having an arylindole moiety with imidazole(Dorion *et al*., 2011).

1.5 computational chemistry*:*

The term computational chemistry is generally used when a mathematical method is sufficiently well developed that it can be automated for implementation on a computer(Young, 2001).

In 1951 an international conference was held at Shelter Island near Long Island in New York, N.Y. Most of the leading figures in quantum chemistry were present. Two persons there symbolized the phasing out of desktop mechanical calculators (Prof. Kotani from Japan) and the phasing in of electronic digital computers (Prof. Roothaan of the United Sates). That was the first major conference with a focus on the emerging computer in theoretical chemistry(Lewars, 2011).

Initially, the focus was on force-field based methodologies for studying the structures, dynamics and interactions of biomolecules as such, and thedevelopment of accurate models for the main biological solvent, water. With the emergence of accurate quantum chemical techniques suitable for studying (from a quantum chemistry perspective) large systems, density functional theory entered the stage in the 1990s as the key approach for investigating enzymatic mechanisms or properties and reactions of small, but biologically relevant, molecules. The combined use of these tools, so-called QM/MM and QM/MM-MD techniques enables precise descriptions of biological phenomena and reactions (Genheden *et al*., 2017).

Computational chemistry is an exciting and fast-emerging discipline which deals with the modeling and the computer simulation of systems(Ramachandran *et al*.,2008).
 The field of computational chemistry has become an extremely valuable research tool in chemistry, physics, and biology(Hango *et al*., 2014).
It can also be described as computation methods aimed at understanding molecules and materials(Sadiku *et al*,. 2017). It uses methods of theoretical chemistry, incorporated into efficient computer programs, to calculate the structures and properties of molecules and solids. However, more frequently the use of computers

to make chemical predictions. Sometimes computational chemistry is used to predict new molecules or new reactions which are later investigated experimentally(Khilari and Kadam, 2017).

Computational methodologies also used for prediction and deep understanding of complex scientific facts has accelerated our knowledge gain and had tremendous impact in our societal growth(Mithri *et al*., 2016).It is also focused on obtaining results relevant to chemical problems, not directly at developing new theoretical methods. There is of course a strong interplay between traditional theoretical chemistry and computational chemistry(Jensen, 2007).

So Computational results normally complement the information obtained from theoretical chemistry and chemical experiments(Matthew et al., 2017).

Computational Science typically unifies three distinct elements:
- Modeling, algorithms and simulations
- Software developed to solve natural science, social science, engineering, and medical problems
- Computer and information science that develops and optimizes advanced hardware systems, software, networking and data management components. (Eweas *et al*., 2014)

The incorporation of molecular and computational strategies to drug development along with organic synthesis approaches has led to a sharp increase in the availability of biological, structural and chemical data.

Computational chemistry is a highly skilled scientific field that has delivered proven results for drug discovery (Slater, 2014). The scientific literature points to several successful examples in the use of computational chemistry on the design and optimization of lead compounds, with application in pharmaceutical sciences. (Alencar Filho, 2017)

Computer-aided drug discovery (CADD) is one strategy that aims to streamline the drug development process. CADD's great potential is evident by its extensive application in novel drug discovery, rational drug design and optimization of drug candidates. Specifically, CADD is used to explain molecular mechanisms of drug activities and to discover drugs with improved activity and reduced side effects(Wang *et al*., 2015).CADD can not only preliminarily predict the activity of inhibitors, but also save experimental costs and provide a guidance for designing more effective inhibitors by exploring the reaction mechanism of inhibitors at the molecular level (Xu *et al*., 2020).

Based on the availability of experimentally determined 3D structures of target proteins. CADD can generally be categorized into structure-based drug design (SBDD) and ligand-based drug design (LBDD)(Wang *et al*., 2015).

Widely used SBDD technologies, X-ray crystallography, molecular docking and structure-based virtual screening have provided important insights into ligand-receptor molecular recognition. On the other side, LBDD techniques, such as pharmacophore modeling, quantitative structure-activity relationships (QSAR), and ligand-based virtual screening, have been broadly used to probe small-molecule virtual databases and identify correlations between biological activity and chemical structure(Ferreira *et al*., 2018).

1.6 Molecular modeling:

Molecular modelingmeans the generation, manipulation, and/or representation of realistic molecular structures and associated physicochemical properties. As generally used by scientists in industry, organic chemistry, and other fields, the term molecular modeling encompasses a number of techniques associated with computational chemistry(Lipkowitz and Boyd,1992).

It is used to model or mimic the behaviour of molecules (Eweas *et al*., 2014).It is also generally implies a graphics 'presentation of the geometry or the configuration of the atoms of the molecule(Dugas,1992).

Molecular modeling is a particularly useful tool because it is an efficient method of preliminary testing for further research. This modeling system can cut down costs that wet laboratories require as well as return approximate results in a speedy and efficient manner. (Hango *et al*., 2014)

1.7 Quantitative structure activity relationship:

First suggested by Hansch and Fujita in the 1960s, SAR modeling is a computer-aided method to describe the mathematical relationship between structural attributes and specific activities of a set of selected chemicals(Wang *et al*.,2015).

The quantitative structure-activity relationship (Leonard and Roy,2006; Fassihi and sabet, 2008; Du *et al*., 2019)is an area of computational chemistry that builds statistical models to predict quantities such as binding affinity, acute toxicity or pharmacokinetic parameters of existing or hypothetical molecules.

In more detail, QSAR is a mathematical relationship between the biological activity of a molecule and its geometric and physicochemical characteristics and

this relationship can be used to evaluate the activity of new molecules (Park *et al*., 2007).

After the earlier QSAR studies by Hansch, who showed a correlation between biological activity and octanol-water partition coefficient, it is now assumed that the sum of substituent effects on the steric, electronic and hydrophobic interaction of compounds with their receptor determines their biological activity (Fassei and Sabet,2008).TheQSAR models are useful for various purposes including the prediction of activities of untested chemicals. The challenge, therefore, is to select the group of descriptors that describe the most critical structural and physicochemical features associated with activity(Leonard and Roy, 2006).The result of the treatment was an equation which describes, in a quantitative manner, the relationship between biological activity of a compound and its chemical structure. (Reed *et al*.,1984)

Effective descriptor or variable selection is an integral part of the QSAR modeling process (Leonard and Roy, 2006)Thus, the development of a good and effective QSAR model is a function of higher quality data and the choice of descriptors (Semire and Oyebamiji, 2017).

The process of QSAR model development can be generally divided into three stages: data preparation, data analysis, and model validation. These steps represent a standard practice of any QSAR modeling, and the researcher's interests,experience, and software availability often determine their implementations(Leonard and Roy, 2006).

1.8 Molecular Docking

Molecular docking is one of the most frequently used methods SBDD because of its ability to predict, with a substantial degree of accuracy, the conformation of small-molecule ligands within the appropriate target binding site (Ferreira *et al*., 2015).

Molecular docking is mainly used to predict the stable drug interactions by inspecting and modeling drug molecular interactions between drug molecule and target receptor molecules. They are used to develop several ligand conformations and orientations and the most suitable ones are selected for further research (Mithiri *et al*., 2016).

The process begins with the application of docking algorithms that pose small molecules within the active site of the target. Algorithms are complemented by

scoring functions that are designed to predict the biological activity through the evaluation of interactions between compounds and potential targets. (Prada-Gracia *et al*., 2016)

1.9 Aim and Objectives

The main aim of this study is to build quantitative structure activity relationship(QSAR) model able to predict the biological activity (Aromatase inhibition) of the indole-imidazole derivatives

Specific objectives

The specific objectives of the present work:

- Identify the ability of the QSAR model by use it to predict the aromatase inhibitor activity of new designedindole imidazole derivatives.
- Docking of the new designed indole-imidazole derivatives (whose have biological activities more or equal to those of letrozole) into the binding site of aromatase protein with iron porphyrin
- Understand the interaction of the compounds with the protein.

Materials and Methods

2.1 Software programs

2.1.1 ACD lab software

Advanced Chemistry Development (Version 12.01), Inc., (ACD/Labs) develops and commercializes informatics solutions for chemical, biochemical, and pharmaceutical R&D. Copyright © 1997–2013 Advanced Chemistry Development, Inc. All rights reserved.

2.1.2 MOE software

Molecular Operating Environment (Version10) software package (MOE, 2009.10, Chemical Computing Group, Montreal, QC, Canada) is a software system designed to support Cheminformatics, Molecular Modelling, Bioinformatics, Virtual Screening, Structure-based-design and can be used to build new applications based on SVL (Scientific Vector Language).

2.1.3 SPSS software

SPSS "**S**tatistical **P**ackage for the **S**ocial **S**ciences" (SPSS, Version 16.0) is a software package used for interactive, or batched, statistical analysis.

2.2 Computational method

2.2.1 Dataset

A dataset of 19 compounds of indole-imidazole derivatives as aromatase inhibitors were obtained from the literature (Wang *et al*., 2013). Their chemical structure and biological data were shown in Table 2.1

The data set was divided into the two subsets, training set of 15 compounds (78.94 %) and test set of 4 compounds (21.05 %). Training set was used to build a regression model, and the test set was used to evaluate the predictive ability of the model obtained.

All compounds were sketched using ACD/ChemSketch Freeware and it were saved as mol. File format. And then it was opened by MOE software and their energies were minimized. Both the training and test sets were stored as a dataset in mdb files, and the relative potency (RP) value for each compound was added manually. The chemical structures of the compounds and their biological activities were shown in Table 2.1

2.2.2 Molecular descriptors

MOE offers a wide range of 2D and 3D molecular descriptors in order to calculate the molecular properties of compounds(Haidar *et al*.,2020). In this work, 2D descriptors were calculated for all compounds in the training set.

The value of the descriptors that use in the QSAR model was shown in the Table 2.2

2.2.3 QSAR study

The QSAR model was developed by selecting the activity (relative potency) as "dependent variable" and the descriptors as model fields. After performing the regression analysis for the training set, the root mean square error (RMSE) and r^2 values of the fit were calculated and the model was saved as a QSAR fit model, which was then used for the prediction of activities of the test data set. The QSAR fit was then used for the validation and cross-validation. Plotting the tested relative potency (RP)values (X-axis) versus the predicted (PRED) values (Y-axis) was performed to assess the predictive ability of the model. The correlation coefficient (r^2) was determined for the test set using the QSAR fit model and was used to predict the activity of the new designed indole-imidazole derivatives as well. The predicted RP was shown in table 2.1

2.2.3.1 QSAR model

In this work, the QSAR models constructed by partial least square (PLS); PLS regression is a technique that generalizes and combines features from principal component analysis and multiple regressions. It is particularly useful when there is a need to predict a set of dependent variables (Y, bioactivity in QSAR) from a very large set of independent variables (X, molecular descriptors specifically in QSAR). It used to build a linear model and to find the relations between two matrices by the following Eq.1.

$$Y = XB + E \qquad\qquad\qquad\qquad Eq.1.$$

where Y is an n by m variables response matrix, X is an n by p variables predictor matrix, B is a p by m regression coefficient matrix, n is the number of samples, m is the number of dependent variables, p is the number of independent variables, and E is an error for the model, which has the same dimensions as Y(Park *et al*.,2008).

The minimum ratio between the number of compounds and the number of the descriptors was five(Chayawan *et al*.,2020).

2.2.3.2 Validation of the QSAR model

There are two main types of validation required to establish a final QSAR model: i) internal validation and ii) external validation. Leave-one-out (LOO) cross-validation and a variant called k-fold cross-validation are the two most widely used internal validation methods]. In brief, any one of the components used as validation data will in turn form the training set to estimate the coefficients of the QSAR model and to predict the bioactivity of the test compound. The predictive power of the model is characterized by the cross validated Q^2 by Equation 2:

$$Q^2 = 1 - \Sigma \, (Ypred - Yobs)^2 \, / \, \Sigma \, (Ypred - Ymean)^2 > 0.5 \qquad \text{Eq.2.}$$

External validation methods are designed to predict the activity of the test set molecules not used for model construction. This method is very rigorous and is considered the most conclusive proof of reliability for the developed model. (Wang et al.,2015)

2.2.3.3 Evaluating the performance of QSAR models

Six statistical parameters such as square of correlation coefficient (r^2) root mean square of error (RMSE), Fischer's value (F), cross-validated correlation coefficient (q^2), standard error of estimate (SEE)and significance of the model(P)were used to assess predictive performance of the constructed QSAR model.

2.2.4 Modeling of new indole-imidazole derivatives

A new set of 112 indole-imidazole derivatives were designed by using ACD/ChemSketch and the structure of the all compounds were adjusted then it was saved as mol. file format. It was then opened by MOE software and their energy was minimized. The QSAR model-1 was selected to predict the biological activity of designed compounds. Structure and the predicted biological activity were shown in Table2.5.

2.2.5 Crystal Structure from PDB

Structure of aromatase containing an iron porphyrin is available in Protein Data Bank (PDB entry: 3EQM) with a resolution of 2.9 Å(Ghosh *et al*., 2009). The structure was optimized by using MOE software.3D protonation was done to

change the state into an ionization level. Then water molecules, androstenedione and phosphate ions were removed from the original PDB file, and reserved the remaining 452 amino acids and an iron porphyrin for subsequent computational study.

2.2.6 Data base generation

From the 112 new designed compounds only 18 compounds have a relative potency equal or more than those of letrozole.

All of the 18 compounds were optimized by the MOE software. The energy of the compounds was minimized using the following parameters gradient: 0.05, Force Field: MMFF94X. It was then saved as database for the docking step.

2.2.7 Molecular docking

2.2.7.1 Validation of the docking procedure

The co-crystallized ligand (Haem) was re-docked into the prepared aromatase protein structure binding pocket. RMSD value of 0.3319 was obtained indicating a successful docking procedure.

2.2.7.2 Molecular docking study

The docking of the new designed indole-imidazole derivatives into the active site of the aromatase enzyme (3EQM) was achieved using MOE-Dock implemented in MOE. The docking parameters were set as Rescoring 1: London dG, Placement: triangle matcher, Retain10, Refinement Force field, and Rescoring 2: None, Retain 10. The docking part of MOE can give the correct conformation of the ligand to obtain a minimum energy structure. The top conformation for each compound was selected based on the S score, and visual inspection in 2D and 3D was carried out using MOE.

Table2.1 Structures and *in vitro* CYP19 inhibitory activities of indole-imidazole derivatives (Wang et al.,2013) and the predicted RP from QSAR study
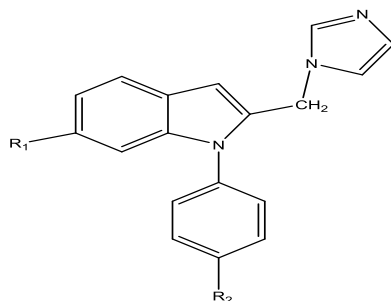
Compounds	R_1	R_2	$IC_{50}{}^a$ (nM)	$RP^b{}_{exp}$	RP_{pre}	Residual
10a[T]	H	F	9.01 ± 12.61	1.8346	0.2348	1.5998
10b	H	CH_3	148.93 ± 12.61	0.1110	0.2013	-0.0903
10c	H	OCH_3	77.36 ± 6.31	0.2137	0.1557	0.0580
10d[T]	H	CF_3	4.93± 0.23	3.3529	0.9769	2.2376
10e	CH_3	H	16.58±1.39	0.997	0.9077	0.0893
10f	CH_3	H	21.39± 1.76	0.7728	0.8621	-0.0893
10g	CH_3	CH_3	164.01± 14.63	0.1008	0.1607	-0.0599
10h	CH_3	OCH_3	138.72±11.46	0.1192	0.1155	0.0037
10i	CH_3	F	56.83±5.18	0.2909	0.1905	0.1004
10j	OCH_3	CN	27.01±1.92	0.6120	0.4973	0.1147
10k[T]	OCH_3	H	6.23±0.51	2.6576	0.2269	2.4307
10l	OCH_3	F	48.93±4.36	0.3378	0.2162	0.1216
10m[T]	OCH_3	CF_3	25.56±2.14	0.6467	0.8837	-0.237
10n	OCH_3	CN	46.92±4.13	0.3523	0.5048	-0.1525
10o	OCH_3	CH_3	57.43±4.96	0.2878	0.1867	0.1011
10p	OCH_3	OCH_3	111.09±10.47	0.1488	0.1393	0.0095
10q	Cl	OCH_3	203.34±18.91	0.0813	0.1147	-0.0334
10r	Cl	CH3	235.33±22.49	0.0702	0.1599	-0.0897
10s	Cl	Cl	217.43±20.67	0.0760	0.1591	-0.0831
Letrozole	_	_	16.53±1.24	1.00	_	_

a Values are the mean of at least three experiments performed in triplicate.

b Relative potency RP = IC_{50} (letrozole)/IC_{50}(tested compound).

T Test set compounds

Table 2.2The value of chemical descriptors used in the QSAR models

No	Compounds	Q_VSA_FPNEG	Diameter	Petitjean	LogS	BcutSlogp0	VSA_hyd	logP(o/w)
1	10a	0.1015	10	0.5000	-4.1067	-2.2412	246.4117	4.5795
2	10b	0.0972	10.0000	0.5000	-4.2856	-2.2441	257.0992	4.7245
3	10c	0.1005	11.0000	0.4545	-3.8621	-2.2570	269.5991	4.3825
4	10d	0.1015	11.0000	0.4545	-4.8682	-2.2424	272.4621	5.3613
5	10e	0.1031	9.0000	0.444	-3.8117	-2.2408	241.3390	4.4265
6	10f	0.0972	9.0000	0.4444	-4.2856	-2.2417	257.0992	4.7615
7	10g	0.0919	10.0000	0.5000	-4.7595	-2.2450	272.9077	5.0595
8	10h	0.0953	11.0000	0.4545	-4.3360	-2.2575	285.4076	4.7175
9	10i	0.0958	10.0000	0.5000	-4.5806	-2.2421	262.2202	4.9145
10	10j	0.1449	11.0000	0.4545	-4.6365	-2.2420	243.4992	4.4215
11	10k	0.1005	10.0000	0.5000	-3.8621	-2.2495	269.5991	4.4195
12	10l	0.0991	10.0000	0.5000	-4.157	-2.2497	274.7201	4.5725
13	10m	0.1952	11.0000	0.4545	-4.9186	-2.2504	300.7706	5.3543
14	10n	0.1459	11.0000	0.4545	-4.2130	-2.2496	255.992	4.0795
15	10o	0.0953	10.0000	0.5000	-4.3360	-2.2513	285.4076	4.7175
16	10p	0.0984	11.0000	0.4545	-3.9125	-2.2593	287.8923	4.3755
17	10q	0.0952	11.0000	0.4545	-4.5964	-2.2569	257.0992	5.0115
18	10r	0.0918	10.0000	0.5000	-5.0199	-2.2440	269.5991	5.3535
19	10s	0.0917	10.0000	0.5000	-5.2803	-2.2403	241.3390	5.6475

Table 2.3 QSAR model equations

Equation number	Equation	R²	RMSE
1	RP = 5.73683 -0.33652 diameter -5.83615 petitjean+7.69314 Q_VSA_FPNEG	0.892	0.0889
2	RP = 36.96524+14.36218BCUT_SLOGP_0-0.19678diameter-0.49586 logP(o/w)	0.742	0.1373
3	RP = 73.97961 +31.71359 BCUT_SLOGP_0 -0.00605 diameter +0.52610 logS	0.690	0.1507
4	RP = 39.73608+14.95553BCUT_SLOGP_0-0.17991diameter-8.35299 petitjean	0.785	0.1253
5	RP= -46.48105-22.46454BCUT_SLOGP_0-0.50992diameter+14.75699 Q_VSA_FPNEG	0.772	0.1292
6	RP =74.72306+31.99324BCUT_SLOGP_0 -0.16082logP(o/w)+0.39305logS	0.705	0.147
7	RP =74.13977 +31.01140 BCUT_SLOGP_0 -0.28704 logP(o/w) - 5.81027petitjean	0.791	0.1236
8	RP =76.30279+32.29670BCUT_SLOGP_0 -0.62022logP(o/w) - 4.36607Q_VSA_FPNEG	0.657	0.1585
9	RP =81.79346+34.36055BCUT_SLOGP_0 +0.36196 logS-5.58242 petitjean	0.862	0.1005
10	RP = 71.30983+30.73617BCUT_SLOGP_0+0.50056 logS+2.89421 Q_VSA_FPNEG	0.721	0.1429
11	RP =4.43770 -0.00076 BCUT_SLOGP_0 +0.23154 logS-0.01155 vsa_hyd	0.643	0.161
12	RP =5.86637-0.30526 logS-0.74205 logP(o/w) -0.32977 diameter	0.729	0.1408

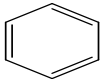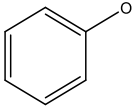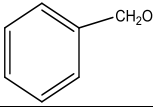| 13	RP =7.18458 -0.27606 logP(o/w)-0.30215 diameter-5.20218 petitjean	0.823	0.1137
14	RP =4.38073-0.29641logP(o/w)-0.32749 diameter+6.73046Q_VSA_FPNEG	0.784	0.1258
15	RP = 5.74647-0.33848logP(o/w)-0.21198diameter-0.00615 vsa_hyd	0.767	0.1304
16	RP =6.72504 -0.27134 diameter+0.08791logS -6.84484 petitjean	0.733	0.1398
17	RP = 3.62909 -0.31439 diameter +0.25352logS+9.91441 Q_VSA_FPNEG	0.807	0.1186
18	RP =5.27331-0.13168 diameter +0.25260 logS-0.00929 vsa_hyd	0.735	0.1392
19	RP =7.16269-0.19911 diameter -6.15301 petitjean -0.00703 vsa_hyd	0.849	0.1051
20	RP = 4.47192 +0.16796 logP(o/w)+0.36758logS-0.01241vsa_hyd	0.657	0.1584
21	RP = 5.2526 -0.02740 logP(o/w) -0.01083 vsa_hyd -4.00102petitjean	0.665	0.1565
22	RP = 5.27523 -3.02213 petitjean +0.13603logS -0.01090vsa_hyd	0.693	0.1499
23	RP = 5.39799 -4.37681 petitjean -0.01108 vsa_hyd -0.27721 Q_VSA_FPNEG	0.669	0.1556
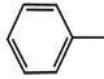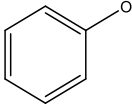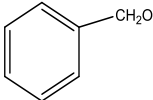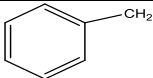
Table2.4 Statistical parameters for the best QSAR model equations
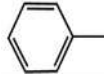
	NO. of training set	No. of test set	r^2	Q^2	r^2_{pre}	RMSE	SEE	F value	P value
Equation 1	15	3	0.892	0.741	0.84	0.0907	0.139	30.305	< 0.005
Equation 9	15	3	0.862	0.754	0.950	0.1005	0.1174	22.934	< 0.005
Equation19	15	3	0.849	0.712	0.997	0.1051	0.122	20.6	< 0.005
Equation 13	15	3	0.823	0.715	0.916	0.1137	0.1328	17.09	<0.005

Table 2.5 Chemical structure and the predicted relative potency of the new designed indole imidazole derivatives
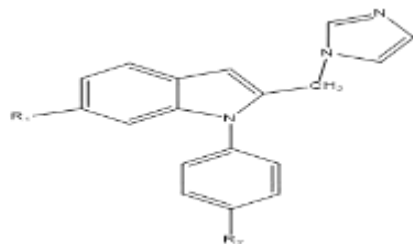
NO.	Compounds	R_1	R_2	RP^a_{pre}	RP^b_{pre}
1	I	CN	CF_3	1.2699	-0.4138
2	II	NO_2	CF_3	1.6107	-0.7736
3	III	NH_2	CF_3	1.0648	-0.5270
4	IV		CF_3	-0.0302	-1.0582
5	V	CH_3	CF_3	0.8936	-0.5477
6	VI	Cl	CF_3	0.8920	-4.5441
7	VII	Br	CF_3	0.8401	-0.3568
8	VIII		CF_3	-0.6402	-4.9066
9	IX		CF_3	-0.7877	-0.5138
10	X	OCOCH3	CF_3	0.4761	-2.6040
11	XI	COH	CF_3	1.1836	-0.5820
12	XII	CF3	CF_3	1.6359	-0.2071
13	XIII	F	CF_3	0.9547	-1.7754
14	XIV	OCH_2CH_3	CF_3	0.2105	-0.6402
15	XV	OH	CF_3	1.1124	-0.7291
16	XVI	CH_2CH_3	CF_3	0.8187	-0.8150
17	XVII	$NHCOCH_3$	CF_3	0.4144	-1.8619
18	XVIII	CH_2ph	CF_3	-0.6471	-2.3765
19	XIX	H	NO_2	0.9443	0.2721
20	XX	H	OH	0.4241	0.8131
21	XXI	H	COOH	0.6167	0.3887
22	XXII	H	$OCOCH_3$	-0.1867	-0.0346
23	XXIII	H	Br	0.1723	0.3754
24	XXIV	H	Cl	0.2005	0.4612
25	XXV	H	$COCH_3$	0.4033	0.2524
26	XXVI	H	COH	0.4611	0.4474

NO.	Compounds	R_1	R_2	RP^a_{pre}	RP^b_{pre}
27	XXVII	H	CH_2CH_3	0.0894	-4.2558
28	XXVIII	H	OCH_2CH_3	-0.4863	-3.9535
29	XXIX	H	CN	0.5600	0.5725
30	XXX	H	SO_3H	1.0250	0.5363
31	XXI	H	$NHCOCH_3$	-0.2487	0.1688
32	XXXII	H	N_2	0.6651	0.7163
33	XXXIII	H	CCl_3	1.9233	-0.6112
34	XXXIV	H	NH_3	0.2300	-0.2497
35	XXXV	H	$COOCH_3$	-0.1867	0.0336
36	XXXVI	H	$COOCH_2CH_3$	-0.3441	-4.1321
37	XXXVII	(phenyl ring)	CH_3	-0.6141	-1.2283
38	XXXVIII	(phenyl ring–O)	CH_3	-0.7497	-1.3535
39	XXXIX	OCH_2CH_3	CH_3	0.0793	-3.8992
40	XL	COH	CH_3	0.4746	0.1466
41	XLI	(phenyl ring–CH_2O)	CH_3	-1.2878	-1.8815
42	XLII	(phenyl ring–CH_2)	CH_3	-0.7531	-2.0712
43	XLIII	Br	CH_3	0.1346	-0.0833
44	XLIV	F	CH_3	0.1905	0.1841
45	XLV	CF_3	CH_3	0.9649	-0.4382
46	XLVI	OH	CH_3	0.3701	0.3750
47	XLVII	NH_2	CH_3	0.3309	0.5246
48	XLVIII	$COOCH_2CH_3$	CH_3	-0.2826	-4.0445
49	XLIX	CH_2OCH_3	CH_3	0.0793	-2.1274
50	L	CH_2CH_3	CH_3	0.1242	-4.2017
51	LI	CN	CH_3	0.5685	0.2598
52	LII	NO_2	CH_3	0.9335	0.0586
53	LIII	$COCH_3$	CH_3	0.4227	0.0860
54	LIV	$OCOCH_3$	CH_3	0.3648	-0.1847
55	LV	Cl	CN	0.4961	0.2602
56	LVI	Cl	NO_2	0.8606	-0.0302

NO.	Compounds	R_1	R_2	RP^a_{pre}	RP^b_{pre}
57	LVII	Cl	SO_3H	0.9429	0.2410
58	LVIII	Cl	N_2	0.5948	0.4040
59	LIX	Cl	NH_3	0.1864	-0.5762
60	LX	Cl	$OCOCH_3$	-0.2380	-0.3271
61	LXI	Cl	CH_2CH_3	0.0523	-4.5693
62	LXII	Cl	OH	0.3690	0.4869
63	LXIII	Cl	F	0.1890	0.3050
64	LXIV	Cl	Br	0.1339	0.0492
65	LXV	Cl	CCl_3	1.8040	-0.9065
66	LXVI	Cl	COOH	0.5514	0.0865
67	LXVII	Cl	OCH_2CH_3	-0.5232	-4.2554
68	LXVIII	Cl	COH	0.4022	-0.0497
69	LXIX	Cl	$COCH_3$	0.3504	0.1283
70	LXX	Cl	OSO_2NH_2	0.3414	-2.6280
71	LXXI	Cl	CH_2OCH_3	-0.5232	-4.4257
72	LXXII	Cl	H	0.8612	0.5821
73	LXXIII	CH_2ph	CN	-0.9771	-1.8784
74	LXXIV	$COCH_3$	CN	0.7284	0.3877
75	LXXV	COOH	CN	0.9337	0.4680
76	LXXVI	$COOCH_3$	CN	0.1240	0.2194
77	LXXVII	CH_2OCH_3	CN	-0.1514	-2.0417
78	LXXVIII	CCl_3	CN	2.1235	-0.4236
79	LXXIX	OH	CN	0.7117	0.6332
80	LXXX		CN	0.9892	0.6864
81	LXXXI	$COOCH_2CH_3$	CN	-0.0494	-3.9315
82	LXXXII	F	CN	0.5433	0.4420
83	LXXXIII		CN	-0.7263	-0.9129
84	LXXXIV	COCl	CN	1.3852	0.0936
85	LXXXV		CN	-0.3742	0.5725
86	LXXXVI	Oph	CN	-0.9719	-0.8587
87	LXXXVII	CF_3	CN	1.2699	-0.9633
88	LXXXVIII	CN	CN	0.8891	0.5469
89	LXXXIX	NH_2	CN	0.6681	0.7832
90	XC	CH_3	OH	0.3701	0.5796
91	XCI	CH_3	NO_2	0.8622	0.0621
92	XCII	CH_3	COH	0.4033	0.2279

NO.	Compounds	R_1	R_2	RP^a_{pre}	RP^b_{pre}
93	XCIII	CH_3	COOH	0.5526	0.1790
94	XCIV	CH_3	CH_2CH_3	0.0530	-4.4477
95	XCV	CH_3	$COCH_3$	0.3514	0.0415
96	XCVI	CH_3	SO_3H	0.9445	0.3315
97	XCVII	CH_3	Cl	0.1599	0.2284
98	XCVIII	CH_3	Br	0.1346	0.1426
99	XCIX	CH_3	$COOCH_3$	-0.2370	-0.1690
100	C	CH_3	NH_3	0.1877	-0.4701
101	CI	CH_3	CH_2OCH_3	-0.5225	-2.5094
102	CII	CH_3	OCH_2CH_3	-0.5225	-4.1332
103	CIII	CH_3	$COOCH_2CH_3$	-0.3947	-4.3032
104	CIV	CH_3		0.5962	0.4974
105	CV	CH_3	OSO_2NH_2	0.3429	0.2199
106	CVI	CH_3	CF_3	0.8936	-0.6402
107	CVII	CH_3	COCl	1.0212	-1.863
108	CVIII	COOH	CF_3	1.4611	-0.5229
109	CIX	OSO_2NH_2	CF_3	1.0062	-0.2670
110	CX	SO_3H	CF_3	1.7938	-0.1680
111	CXI	SO_3H	CH_3	1.1647	0.4520
112	CXII	SO_3H	CN	1.4483	0.7598

a predicted RP by equation 1

b predicted RP by equation 13

Table 2.6 Binding energy, bond of interaction and Amino acid interaction of the new designed indole imidazole derivatives with the active site of aromatase

No	Compounds number	Bond energy(kcal/mole)	Interacted amino acid	Interacted group	Type of interaction	Bond length
1	I	-24.70	NO INTERACTION			
2	II	-21.13	NO INTERACTION			
3	III	-22.51	NO INTERACTION			
4	XI	-21.34	Arg115	Imidazole ring	π-cation	_
5	XII	-24.5	NO INTERACTION			
6	XXX	-21.14	Arg115	SO_3H	Hydrogen bond(acceptor)	2.00
7	XXXIII	-20.27	NO INTERACTION			
8	LXV	-20.87	NO INTERACTION			
9	LXXVIII	-19.58	NO INTERACTION			
10	CVII	-22.64	NO INTERACTION			
11	XV	-21.88	NO INTERACTION			

12	LXXIV	-23.77	NO INTERACTION			
13	CVIII	-27.46	Arg145	OH-carboxylic acid	Hydrogen bond (acceptor)	2.19
			Arg115	OH-carboxylic acid	Hydrogen bond (acceptor)	1.83
			Arg435	OH-carboxylic acid	Hydrogen bond (acceptor)	2.31
14	CIX	-21.88	Ser314	NH_2-OSO_2NH_2	Hydrogen bond (donor)	2.14
15	CX	-20.08	Arg145	SO_3H	Hydrogen bond	3.13
			Arg115	SO_3H	Hydrogen bond (acceptor)	1.85
16	CXI	-17.73	Arg145	SO_3H	Hydrogen bond(acceptor)	1.99
			Arg115	SO_3H	Hydrogen bond (acceptor)	1.85
			Arg115	SO_3H	Hydrogen bond (acceptor)	2.33
17	CXII	-19.32	Arg115	Substituted phenyl on indole ring	π –cation	_
			Arg115	SO_3H	Hydrogen bond (acceptor)	2.57
			Arg115	SO_3H	Hydrogen bond(acceptor)	1.93
			Arg145	SO_3H	Hydrogen bond (acceptor)	2.38
18	LXXVII	-19.088	_	Imidazole ring	π –cation	_
19	Letrozole	-19.49	No interaction			

Table 2.7 Binding energy, bond of interaction and Amino acid interaction of indole imidazole derivatives with the active site of aromatase

No	Compounds number	Bond energy	Interacted amino acid	Interacted group	Type of interaction	Bond length
1	10a	-20.04	Arg115	Imidazole ring	π –cation	—
2	10d	-21.51	No interaction			
3	10k	-22.95	Arg115	Imidazole ring	π –cation	—

Discussion

3.Discussion

Aromatase, a cytochrome P450 enzyme complex present in breast tissues, plays a significant role in the biosynthesis of estrogen from androgen. Inhibition of the aromatase is a crucial strategy to prevent the growth stimulation effect of estrogens in breast cancer tissues in postmenopausal women. So, aromatase inhibitors (AIs) have been developed and used as anti-breast cancer agents. (Pingaew *et al*.,2017)

It was reported that indole-imidazole derivatives were used as aromatase inhibitors, the indole moiety accounts for fitting the binding site of aromatase, (Kang *et al*.,2018) and one of the most important feature for strong inhibitor binding to CYP enzymes is the capability to interact as the sixth ligand with the iron atom of the haem group, always present in this enzyme family. This coordination is excellently performed by the lone pair carried on the sp^2 hybridized imidazole nitrogen but other electron rich heterocycles might be accepted in this position as well. (Castellano *et al*.,2008)

A combination of experimental and computational approaches facilitates researchers to understand the molecular mechanism of inhibitory action and discover more potent non-steroidal AIs against aromatase, thereby opening up a novel therapeutic strategy for hormone dependent breast cancer. (Kang *et al*.,2018)

The knowledge of both a leader structure from QSAR and its interaction type at the active site of the protein target from docking techniques represents a promising path in the search for and the development of new series of molecules active against a particular disease (Márquez *et al*., 2019)

3.1 QSAR study

QSARs are invaluable in the development of new agents as they permit interpretation of structure-activity data and prediction of compounds with some desirable pharmacological profiles (Hasegawa *et al*.,1997). The QSAR approach can be generally described as an application of data analysis methods and statistics to developing models that could accurately predict biological activities or properties of compounds based on their structures(Tropsha,2010) it has been widely used to provide useful information for guiding the design and discovery of many classes of drugs, including aromatase inhibitors(Pingaew *et al*., 2017).

In the present study a dataset of 19 compounds was obtained from the literature (Wang *et al*., 2013)15 compounds were used as a training set to build the QSAR model and four compounds were used as a test set to validate the model.

The QSAR model was obtained by using PLS method by MOE software, the descriptors that used in the PLS model were refined by the following principles (1) the descriptors with standard deviation less than 0.0001 were excluded; (2) the descriptors with at least one missing value were deleted; (3) the descriptors with (abs)pair correlation larger than or equal to 0.8 were excluded; and (4) the descriptors that Pearson correlation coefficients ($|r|$) between descriptors and relative potency of indole-imidazole derivatives lower than 0.3 were deleted. The remaining descriptors were used for the further analysis(Qin *et al*., 2017).

The descriptors that used for building the PLS models are listed in table 3.1.

Table 3.1 The definition of descriptors that used in PLS model

Descriptor	Definition	Class
Diameter	Largest value in the distance matrix [Petitjean ,1992].	Physical properties
Petitjean	Value of (diameter - radius) / diameter.	Physical properties
Q_VSA_FPNEG	Fractional negative polar van der Waals surface area. This is the sum of the v_i such that q_i is less than -0.2 divided by the total surface area.	Partial Charge
Q_VSA_HYD	Total hydrophobic van der Waals surface area. This is the sum of the v_i such that $|q_i|$ is less than or equal to 0.2.	Partial Charge
VSA hyd	Approximation to the sum of van der Waalssurface areas of hydrophobic atoms (Å^2).	Pharmacophore Feature
BCUTSLOGP0	The BCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge.	Adjacency and Distance Matrix
LogS	Log of the aqueous solubility (mol/L). This property is calculated from an atom contribution linear atom type model [Hou et al., 2004]	Physical properties
LogP(o/w)	Log of the octanol/water partition coefficient (including implicit hydrogens). This property is calculated from a linear atom type model [Labute, 1998] with $r^2 = 0.931$, RMSE=0.393 on 1,827 molecules.	Physical properties

q_i is the partial charge of atom i.

$v_i$is the van der Waals surface area

Refer to table 2.3a number of QSAR equations were obtained the best of them are listed below:

RP = 5.73683 -0.33652 diameter -5.83615 petitjean+7.69314 Q_VSA_FPNEG

$N = 15$ $R^2 = 0.892$, F = 30.305, $Q^2 = 0.74$……………………………………. (1)

RP = 81.79346+0.36196 logS -5.58242 petitjean +34.36055 BCUT_SLOGP_0

$N = 15$, $R^2 = 0.862$, F= 22.93, $Q^2 = 0.754$……………………………………. (9)

RP = 7.16269 -0.19911 diameter -0.00703 vsa_hyd -6.15301 petitjean

$N=15$, $R^2=0.849$, F=20.6, $Q^2 =0.712$……………………………………. (19)

RP = 7.18458 -0.27606 logP(o/w)-0.30215 diameter-5.20218 petitjean

$N = 15$, $R^2 =0.823$, F=17.09, $Q^2=0.715$……………………………………. (13)

Were RP is the relative potency of the training set compounds

In principle, q^2 and r^2 should have higher values while SEE should have smaller value. High q^2 and r^2(in general, $q^2>0.5$ and $r^2>0.6$) are two important indicators that the model is highly predictive. (Li *et al*., 2018)

Fischer's value (F), is the Fisher ratio, reflects the ratio of the variance explained by the model and the variance due to the error in the regression. High values of the *F* -test indicate that the model is statistically significant. (Khamouli *et al*., 2018) and p-value should not be higher than 0.05 (Golbraikh *et al*., 2017)

The best model, selected according to the statistical robustness (where $q^2>0.5$ and $r^2>0.6$indicates predictive models), is model 1, with an R^2 of 0.892, a high Fisher ratio value of 30.02, and a great cross validated correlation coefficient (Q^2) of 0.741. The difference between R^2 and Q^2 values was 0.1507 unit, it ensures that model 1 does not present data overfitting(Qin *et al*., 2013).

Equation (1) is dependent on three descriptors, namely Fractional negative polar van der Waals surface area; the increase in the value of this descriptor will increase the relative potency, diameter and petitjean the decrease in the value of these two descriptors will increase the RP.

Equation 13 also applied to predict the biological activity of the new designed indole-imidazole derivatives but it show a low value.

Plots of experimental versus predicted RP values, experimental versus X - Predicted (For the cross validation LOO) and experimental versus predicted for test

set (external validation) obtained from model 1 was shown in figure 3.1,3.2 and 3.3 consecutively
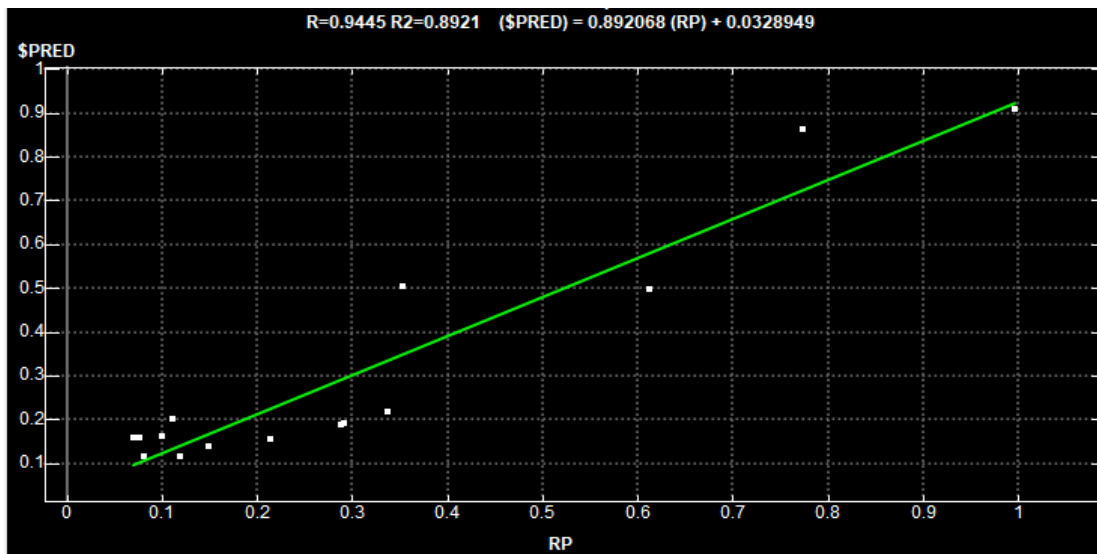

Figure 3.1 Correlation plot between the experimental and predicted RP of training set
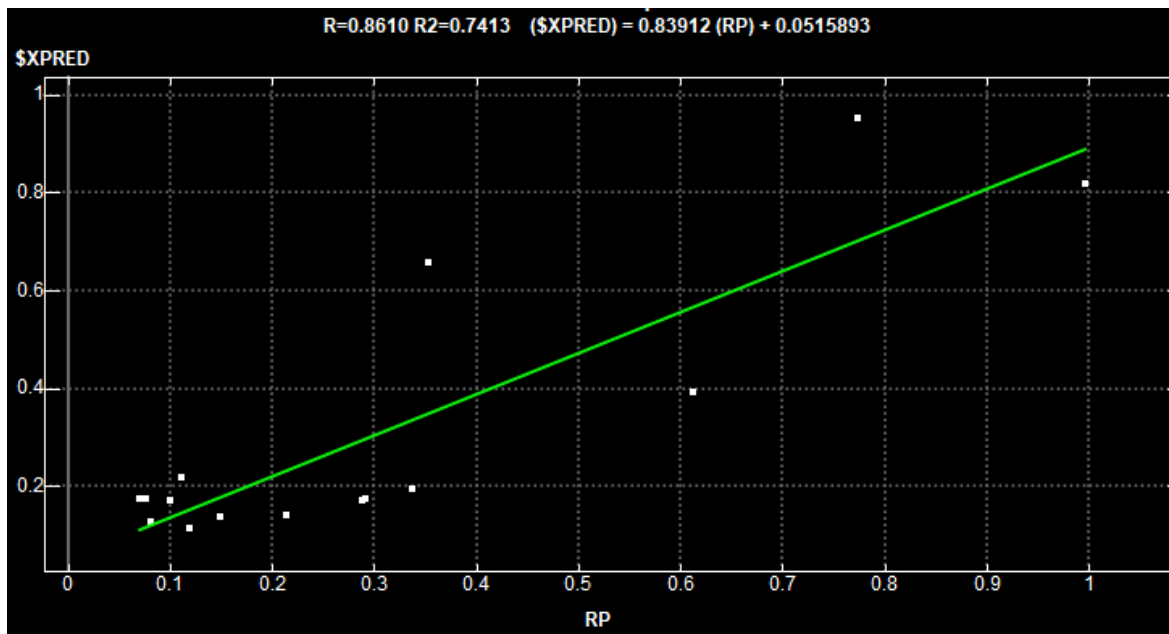

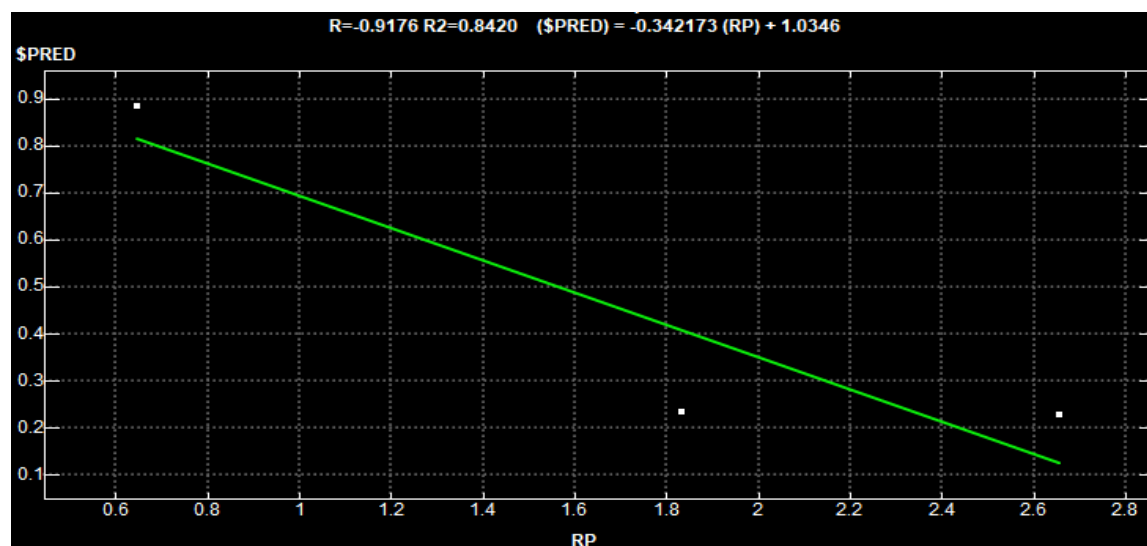Figure 3.2 Cross validation plot of the experimental RP versus predicted

Figure 3.3 Correlation plot between the experimental and predicted RP of test set

The correlation between the biological activity of the indole-imidazole compounds and the descriptor that use in the best model were obtained from the correlation matrix (Figure 3.4) all of the descriptors found to have correlation with the biological activity more than 0.3

	1	2	3	4	5	6	7
1. RP	100	44	-53	44	-52	36	-73
2. BCUT_SLOGP_0	44	100	-70	-44	32	10	-70
3. diameter	-53	-70	100	4	-24	40	39
4. logS	44	-44	4	100	-55	15	-14
5. petitjean	-52	32	-24	-55	100	-42	22
6. Q_VSA_FPNEG	36	10	40	15	-42	100	-55
7. vsa_hyd	-73	-70	39	-14	22	-55	100

Figure 3.4 Correlation Matrix between the descriptors

Wang *et al.*,2013 observed that proton or a small electron-withdrawing groups on the C-4 position of the phenyl ring, connected to the indole nitrogen, might enhance AI activities, and any bulky group should be avoided in order to keep a relative small value for these kinds of molecules. So in this study the designing of

the new indole-imidazole derivatives depending on this observation but also some molecules with an electron donating group at the C-4 position were designed to make a comparison between the two type of the substituents.

A group of 112 compounds were designed and modeled, and their biological activity were predicted by using Model -1.

All of the compounds that exhibit high inhibitor activity with relative potency (RP) from(2.12 – 1.02) have an electron withdrawing group in the meta position of the indole ring R_1.

Compound XXVII and LXII which possess an electron donating group at R_2 show a low inhibitor activity of 0.0894 and 0.3690 consecutively.

all of the compounds that exhibit high inhibitor activity toward aromatase (more than letrozole) have an electron withdrawing group at para position of the phenyl ring R_2. So R_2 favorite electron withdrawing group and these support the observation of Wang *et al*.,2013.

The most potent compounds in the new designed compounds are LXXVIII, XXXIII, LXV, XII, CX, II and CVIII (RP: 2.12,1.92,1.80,1.63,1.79,1.61,1.46), which are more powerful in the inhibition of the aromatase as compared to letrozole.

3.1.1 Application domain of the QSAR model

The application domain of the QSAR model was defined by the leverage approach from the hat matrix (h_i in Equation (1)), which is calculated from the descriptors of chemicals, and by identification of chemicals with LOO cross-validated standardized residuals greater than 2.0 standard deviation units.

An outlier in the QSAR model is defined as h_i value larger than the warning leverage h* and LOO standardized residuals greater than 2.0, which is graphically depicted in the Williams plot. The warning leverage h* is fixed at (3k)/n, where k is the number of model parameters and n is the number of the objects used to calculate the model.

$$h_i = x_i^T (X^T X)^{-1} x_i \ (i = 1, \ldots, n)$$ (1)

where x_i is the descriptor row vector of the query compound; X is the n × k matrix of k model descriptorvalues for n training set compounds and the superscript T refers to the transpose of the matrix/vector(Qin *et al*., 2017).
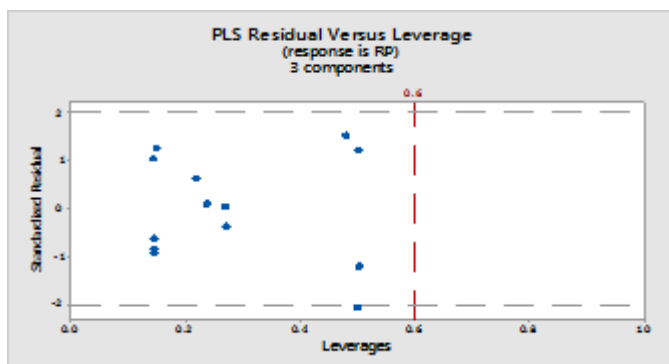
Figure 3.5 Williams plot of the model- 1

The control leverage h* for model 1 fixed at 0.6. As can be seen from figure 3.4 model -1 has no outlier.

3.2 Molecular docking study

Molecular docking is widely used in the field of drug screening and design. The theoretical basis is that the process of ligand and receptor recognition relies on spatial shape matching and energy matching, which is the theory of "inducing fit". Determining the correct binding conformation of small molecule ligands and protein receptors in the formation of complex structures is the basis for drug design and studying its action mechanism(Lin *et al*., 2020).

In order to understand the binding mode of active compounds in the active site pocket of aromatase (Figure 3.), docking study was performed using MOE-Dock.

The specific cleft in which the ligands bind (within 4 Å) contains both polar (Arg115, Arg375, Asp309, Asp371, Ser478, Thr310, Asp371, Glu302) and nonpolar (Ala306, Ala307, Ile133, Ile305, Leu477, Met374, Phe134, Phe221, Trp224, Val369, Val370, Val373) amino acids. (Roy and Roy,2010)the currently accepted binding modes of non-steroidal aromatase inhibitors (e.g., letrozole and anastrozole) to aromatase have been learned from computational studies (Prior *et al*., 2017)
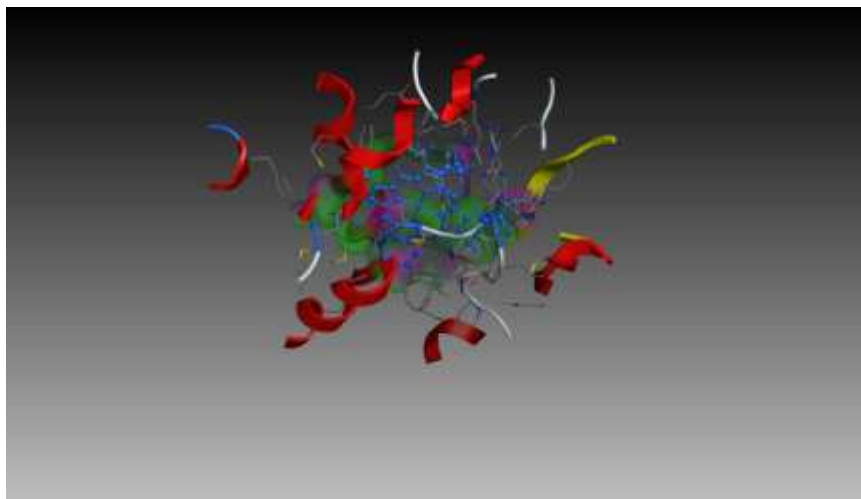
Figure3.6 Interaction of haem with the active site of aromatase

From the 112 new designed compounds only 18 compounds have biological activity (RP) more than letrozolewere docked into the active site of the aromatase to figure out their binding energy and the mode of interaction.

Compound CVIII show a low binding energy (high stability) which it is lower than those of letrozole by 8 units; it interacted with aromatase by 5 bonds three of them are hydrogen bonds with Arg145, Arg115 and Arg435 with bond angles of 2.19,1.83,2.31respectively. and metal/ion contact with Arg145, Arg115 with bond angles of 2.87and 2.75 respectively. The 2Dand 3D contact style of compound CVIII with aromatase was shown in figure 3.7
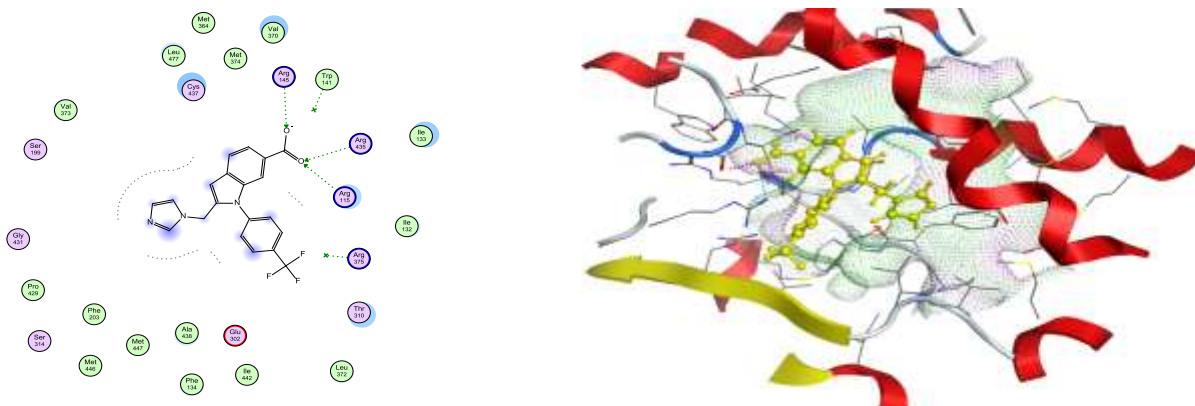

Figure 3.7 2D and 3D binding mode of compound CVIII with the active site of the aromatase enzyme

Figure 3.8 2D and 3D binding mode of compound XXX with the active site of the aromatase enzyme

Figure 3.9 2D and 3D binding mode of compound CX with the active site of the aromatase enzyme

Also compounds CXI and CXII show a high interaction with aromatase enzyme (6 interactions and 7 interactions respectively) but it shows a high binding energy of -17.73, -19.32 Kcal/mole respectively.

Conclusion and Recommendation

4. Conclusion and Recommendations

In this study 112 indole-imidazole derivatives were designed and their biological activity were evaluated by using a QSAR model (with three descriptors Q_VSA_FPNEG, diameter, Petitjean). These models were building by using PLS method and it was validated internally and externally.

After applying the QSAR model 18 compounds showed relative potency (RP) higher than letrozole.These compounds were used in the docking study to understand their mode of interaction and binding energy.

From docking study compound CVIII showed high binding energy (-27.46 Kcal/mole) which higher than those of letrozole (-19.49) by 8 units and it show a high interaction with aromatase enzyme (3 interactions)

Further research can be carried out for compound (CVIII, XXX, CX), synthesis and *in vitro* investigation of their aromatase inhibitir activity.

Reference

Alencar Filho, E. B. (2017). Computational Medicinal Chemistry: A Useful Tool for Pharmaceutical Sciences and Drug Development. *Organic and Medicinal Chemistry International Journal*, 4(3), 44-45.

Ali, I.,Mukhtar, S. D., Hsieh, M. F., Alothman, Z. A., and Alwarthan, A. (2018) Facile synthesis of indole heterocyclic compounds based micellar nano anti-cancer drugs. RSC advances, 8(66), 37905-37914.

Ali, I., Lone, M. N., & Aboul-Enein, H. Y. (2017). Imidazoles as potential anticancer agents. *MedChemComm*, *8*(9), 1742-1773.

Aljamali, N.M. (2015). Review paper on heterocyclic compounds.Journal ofPlastic and Polymer Technology (JPPT), 1(1), 49-64.

Ashok, P., Raju, G., Mallikarjunarao, R., Gopal, K. V., Sreeramulu, J., Reddy, D. M., ... & Reddy, S. R. (2013). Synthesis and biological activity of some new indole derivatives containing pyrazole moiety. *J. Chem. Pharm. Res*, *5*, 21-27.

Badiger, J., Manjulatha, K., Girish, M., Sharif, A., & Purohit, M. G. (2009). Synthesis and biological evaluation of some N-substituted indoles. *ARKIVOC: Online Journal of Organic Chemistry*.

Bae, S. H., Park, J. H., Choi, H. G., Kim, H., & Kim, S. H. (2018). Imidazole antifungal drugs inhibit the cell proliferation and invasion of human breast cancer cells. *Biomolecules & therapeutics*, *26*(5), 494.

Barluenga Mur, J., & Valdés Gómez, A. C. (2012). Five-membered heterocycles: Indole and related systems. *ChemInform*.

Baroniya, S., Anwer, Z., Sharma, P. K., Dudhe, R., & Kumar, N. (2010). Recent advancement in imidazole as anticancer agents: A review. *Der Pharmacia Sinica*, *1*(3), 172-182.

Benkli, K., Demirayak, S., Gundogdu-Karaburun, N., Kiraz, N., Iscan, G., & Ucucu, U. (2004). Synthesis and antimicrobial activities of some imidazole substituted indoles.

Benson, J. R., & Ravisekar, O. (2007). Aromatase inhibitors for treatment of breast cancer. *Current Cancer Therapy Reviews*, *3*(1), 67-79.

Castellano, S., Stefancich, G., Ragno, R., Schewe, K., Santoriello, M., Caroli, A., ... & Sbardella, G. (2008). CYP19 (aromatase): exploring the scaffold flexibility for novel selective inhibitors. *Bioorganic & medicinal chemistry*, *16*(18), 8349-8358.

Chaudhury, D., Banerjee, J., Sharma, N., & Shrestha, N. (2015). Routes of synthesis and biological significances of Imidazole derivatives: Review. *World Journal of Pharmaceutical Sciences*, *3*(8), 1668-1681.

Chawla, A., Sharma, A., & kumar Sharma, A. (2012). A convenient approach for the synthesis of imidazole derivatives using microwaves. *Synthesis*, *5*(6), 7.

Chayawan, C., Toma, C., Benfenati, E., & Caballero Alfonso, A. Y. (2020). Towards an Understanding of the Mode of Action of Human Aromatase Activity for Azoles through Quantum Chemical Descriptors-Based Regression and Structure Activity Relationship Modeling Analysis. *Molecules*, *25*(3), 739.

Che, H., Tuyen, T. N., Kim, H. P., & Park, H. (2010). 1, 5-Diarylimidazoles with strong inhibitory activity against COX-2 catalyzed PGE2 production from LPS-induced RAW 264.7 cells. *Bioorganic & medicinal chemistry letters*, *20*(14), 4035-4037.

Chi Indices I: Substituted 2-Sulfapyridines. *Journal of pharmaceutical science*, 73(2), 237-240.

Daraji, D. G., Prajapati, N. P., & Patel, H. D. (2019). Synthesis and Applications of 2- Substituted Imidazole and Its Derivatives: A Review. *Journal of Heterocyclic Chemistry*, *56*(9), 2299-2317.

de Araújo, J. S., Garcia-Rubia, A., Sebastian-Perez, V., Kalejaiye, T. D., da Silva, P. B., Fonseca-Berzal, C. R., ... & Gil, C. (2019). Imidazole derivatives as promising agents for the treatment of Chagas disease. *Antimicrobial Agents and Chemotherapy*, *63*(4).

Doiron, A., Richard, R., Cuperlovic-Culf, M., Robichaud, G. A., & Touaibia, M. (2011). Synthesis and structure activity relationship of 1-and 2-substituted-1, 2, 3-triazole letrozole-based analogues as aromatase inhibitors. *European Journal of Medicinal Chemistry*, *46*, 4010e4024.

Du, M., Zhang, D., Hou, Y., Zhao, X., & Li, Y. (2019). Combined 2D-QSAR, Principal Component Analysis and Sensitivity Analysis Studies on

Fluoroquinolones' Genotoxicity. *International journal of environmental research and public health*, *16*(21), 4156.

Dugas, H. (1992). Teaching molecular modeling: An introductory course for chemists, implemented at the Universite de Montreal. *Journal of chemical education*, *69*(7), 533.

El-Aal, E. A., Fattah, H. A., Osman, N., & Seliem, I. (2015). Synthesis of novel imidazole and fused imidazole derivatives as cytotoxic and antimicrobial agents: Molecular docking and biological evaluation. *Int J Pharm Sci*, *7*(10), 36-45.

Elahian, F., Akbari, M., Ghasemi, M., Behtooee, N., Taheri, M., & Amini, M. (2014). Synthesis and anticancer activity of 2, 4, 5-triaryl imidazole derivatives. *Letters in Drug Design & Discovery*, *11*(7), 840-843.

El-Ezbawy, S. R., & Abdel-wahab, A. M. A. (1989). Synthesis and biological activities of new Indole derivatives containing sulfide and/or sulfone moieties. Part I. *Phosphorus, Sulfur, and Silicon and the Related Elements*, *44*(3-4), 285-289.

El-sayed, M. T., Hamdy, N. A., Osman, D. A., & Ahmed, K. M. (2015). Indoles as anticancer agents. *Advances in Modern Oncology Research*, *1*(1), 20-35.

Eweas, A. F., Maghrabi, I. A., & Namarneh, A. I. (2014). Advances in molecular modeling and docking as a tool for modern drug discovery. *Der Pharma Chemica*, *6*(6), 211-228.

Fan, Y. L., Jin, X. H., Huang, Z. P., Yu, H. F., Zeng, Z. G., Gao, T., & Feng, L. S. (2018). Recent advances of imidazole-containing derivatives as anti-tubercular agents. *European journal of medicinal chemistry*, *150*, 347-365.

Fantacuzzi, M., De Filippis, B., Gallorini, M., Ammazzalorso, A., Giampietro, L., Maccallini, C., ... & Cataldi, A. (2020). Synthesis, biological evaluation, and docking study of indole aryl sulfonamides as aromatase inhibitors. *European journal of medicinal chemistry*, *185*, 111815.

Fassihi, A., & Sabet, R. (2008). QSAR Study of p56lck protein tyrosine kinase inhibitory activity of flavonoid derivatives using MLR and GA-PLS. *International Journal of Molecular Sciences*, *9*(9), 1876-1892.

Ferreira, L. G., Dos Santos, R. N., Oliva, G., & Andricopulo, A. D. (2015). Molecular docking and structure-based drug design strategies. *Molecules*, *20*(7), 13384-13421.

Ferreira, L. G., Oliva, G., & Andricopulo, A. D. (2018). From medicinal chemistry to human health: current approaches to drug discovery for cancer and neglected tropical diseases. *Anais da Academia Brasileira de Ciências*, *90*(1), 645-661.

Genheden, S., Reymer, A., Saenz-Méndez, P., & Eriksson, L. A. (2017). Computational Chemistry and Molecular Modelling Basics.

Ghosh, D., Griswold, J., Erman, M., & Pangborn, W. (2009). Structural basis for androgen specificity and oestrogen synthesis in human aromatase. Nature, 457(7226), 219-223.

Ghosh, R., & Biplab, D. (2013). Review on: synthesis, chemistry and therapeutic approaches of imidazole derivatives. *ChemInform*, *45*(50).

Gobbi, S., Rampa, A., Belluti, F., & Bisi, A. (2014). Nonsteroidal aromatase inhibitors for the treatment of breast cancer: an update. *Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents)*, *14*(1), 54-65.

Golbraikh, A., Wang, X. S., Zhu, H., & Tropsha, A. (2017). Predictive QSAR modeling: methods and applications in drug discovery and chemical risk assessment. In *Handbook of computational chemistry* (pp. 1309-1342). Springer Netherlands.

Gomleksiz, M., Alkan, C., & Erdem, B. (2013). Synthesis, characterization and antibacterial activity of imidazole derivatives of 1, 10-phenanthroline and their Cu (II), Co (II) and Ni (II) complexes. *South African Journal of Chemistry*, *66*, 107-112.

Gore, R. P. (2016). New developments in anticancer activities of imidazole and it is deriviatives. International Journal of Advance Research. 4(12), 2256-2262

Haidar, S., Marminon, C., Aichele, D., Nacereddine, A., Zeinyeh, W., Bouzina, A., ... & Jose, J. (2019). QSAR Model of Indeno [1, 2-b] indole Derivatives and Identification of N-isopentyl-2-methyl-4, 9-dioxo-4, 9-Dihydronaphtho [2, 3-b] furan-3-carboxamide as a Potent CK2 Inhibitor. *Molecules*, *25*(1), 97.

Hango, C. R., Sigalovsky, D., Lu, J., & Devaney, K. J. (2014). Computational Chemistry in the High School Classroom.

Hao, W., Jiang, Y., & Cai, M. (2014). Synthesis of indolyl imidazole derivatives via base-promoted tandem reaction of N-[2-(1-alkynyl) phenyl] carbodiimides with isocyanides. *The Journal of organic chemistry*, *79*(8), 3634-3640.

Hasegawa, K., Miyashita, Y., & Funatsu, K. (1997). GA strategy for variable selection in QSAR studies: GA-based PLS analysis of calcium channel antagonists. *Journal of Chemical Information and Computer Sciences*, *37*(2), 306-310.

Hofmann, C. (1953). Chemistry of Heterocyclic Compounds: A Series of Monographs

Hong, W., Li, J., Chang, Z., Tan, X., Yang, H., Ouyang, Y., ... & Wang, H. (2017). Synthesis and biological evaluation of indole core-based derivatives with potent antibacterial activity against resistant bacterial pathogens. *The Journal of Antibiotics*, *70*(7), 832-844.

Hossain, M., & Nanda, A. K. (2018). A Review on Heterocyclic: Synthesis and Their Application in Medicinal Chemistry of Imidazole Moiety. *Science*, *6*(5), 83-94.

Hote, S. V., & Bhoyar, S. P. (2014). Heterocyclic Compound–A Review. *IOSR Journal of Applied Chemistry*, 43-46.

Hou, T.J., Xia, K., Zhang, W., Xu, X.J.; ADME Evaluation in Drug Discovery. 4. Prediction of Aqueous Solubility Based on Atom Contribution Approach; *J. Chem. Inf. Comput. Sci.44* (**2004**) 266–275.

Hu, Q., Yin, L., & Hartmann, R. W. (2012). Selective dual inhibitors of CYP19 and CYP11B2: targeting cardiovascular diseases hiding in the shadow of breast cancer. *Journal of medicinal chemistry*, *55*(16), 7080-7089.

Inman, M., & Moody, C. J. (2013). Indole synthesis–something old, something new. *Chemical Science*, *4*(1), 29-41.

Jawaharmal, H. S., Narwal, S., Singh, G., Saini, D. R., Kaur, A., & Narwal, S. (2012). Synthesis of novel imidazole compounds and evaluation of their antimicrobial activity. *Indo Global Journal of Pharmaceutical Sciences*, *2*(2), 147-156.

Jensen, F. (2007). *Introduction to computational chemistry*,2nd (ED). John wiley & sons.

Joule, J., & Mills, K. (2010). Heterocyclic Chemistry, 5th (Ed.), A John Wiley & Sons Publication Ltd.

Kang, H., Xiao, X., Huang, C., Yuan, Y., Tang, D., Dai, X., & Zeng, X. (2018). Potent aromatase inhibitors and molecular mechanism of inhibitory action. *European Journal of Medicinal Chemistry*, *143*, 426-437.

Khabnadideh, S., Rezaei, Z., Khalafinezhad, A., Pakshir, K., Heiran, M. J., & Shobeiri, H. (2009). Design and synthesis of 2-methyl and 2-methyl-4-nitro imidazole derivatives as antifungal agents.

Khamouli, S., Beladi, S., Beladi, H., & Belkhiri, L. (2018). QSAR Studies of amino-pyrimidine derivatives as Mycobacterium Tuberculosis Protein Kinase B inhibitors. *Turkish Computational and Theoretical Chemistry*, *2*(2), 16-27.

Khilari, S., & Kadam, S. (2017). Knowledge management in computational chemistry: A literature review, 8(1), 156-162.

Kondaparla, S., Manhas, A., Dola, V. R., Srivastava, K., Puri, S. K., & Katti, S. B. (2018). Design, synthesis and antiplasmodial activity of novel imidazole derivatives based on 7-chloro-4-aminoquinoline. *Bioorganic chemistry*, *80*, I204-211.

Labute, P.; MOE LogP(Octanol/Water) Model *unpublished*. Source code in $MOE/lib/svl/quasar.svl/q_logp.svl (1998).

Lakshmanan, B., Mazumder, P. M., Sasmal, D., & Ganguly, S. (2011). Synthesis, antispasmodic and antidiarrheal activities of some 1-substituted imidazole derivatives. *Acta Pharmaceutica*, *61*(2), 227-236.

Leonard, J. T., & Roy, K. (2006). On selection of training and test sets for the development of predictive QSAR models. *QSAR & Combinatorial Science*, *25*(3), 235-251.

Lewars, E. G. (2011). Computational chemistry. *Introduction to the theory and applications of molecular and quantum mechanics*, 318.

Li, R., Du, Y., & Shen, J. (2018). Molecular modelling studies on cinnoline-based BTK inhibitors using docking and structure-based 3D-QSAR. *SAR and QSAR in Environmental Research*, *29*(11), 847-873.

Lin, X., Li, X., & Lin, X. (2020). A review on applications of computational methods in drug screening and design. *Molecules*, *25*(6), 1375.

Lipkowitz, K. B., and Boyd, B. (1992). Review in computational Chemistry, Volume Ill

Liu, J., Ming, B., Gong, G. H., Wang, D., Bao, G. L., & Yu, L. J. (2018). Current research on anti-breast cancer synthetic compounds. *RSC advances*, *8*(8), 4386-4416.

Maithri, G., Manasa, B., Vani, S. S., Narendra, A., & Harshita, T. (2016). Computational drug design and molecular dynamic studies—a review. *Int J Biomed Data Min*, *6*(01), 1-7.

Manocha, P., Wakode, D. S., Kaur, A., Anand, K., & Kumar, H. (2016). A review: Imidazole synthesis and its biological activities. *Int. J. Pharm. Sci. Res*, *1*(7), 12-16.

Marchand, P., Le Borgne, M., Palzer, M., Le Baut, G., & Hartmann, R. W. (2003). Preparation and pharmacological profile of 7-(α-Azolylbenzyl)-1H-indoles and indolines as new aromatase inhibitors. *Bioorganic & medicinal chemistry letters*, *13*(9), 1553-1555.

Mariappan, G., Alam, S., Sutharson, L., Haldar, P. K., & Nath, S. (2013). Design, synthesis and neuropharmacological activity of dihydro imidazole derivatives. Indian journal of chemistry,vol.52B,pp568-572.

Márquez, E., Mora, J. R., Flores-Morales, V., Insuasty, D., & Calle, L. (2019). Modeling the antileukemia activity of ellipticine-related compounds: QSAR and Molecular Docking Study. *Molecules*, *25*(1), 24.

Martins, P., Jesus, J., Santos, S., Raposo, L. R., Roma-Rodrigues, C., Baptista, P. V., & Fernandes, A. R. (2015). Heterocyclic anticancer compounds: recent advances and the paradigm shift towards the use of nanomedicine's tool box. *Molecules*, *20*(9), 16852-16891.

Mielczarek, M., Devakaram, R. V., Ma, C., Yang, X., Kandemir, H., Purwono, B., ... & Kumar, N. (2014). Synthesis and biological activity of novel bis-indole inhibitors of bacterial transcription initiation complex formation. *Organic & biomolecular chemistry*, *12*(18), 2882-2894.

Murakami, Y. (2012). Peculiarity of methoxy group-substituted phenylhydrazones in Fischer indole synthesis. *Proceedings of the Japan Academy, Series B*, *88*(1), 1-17.

Naim, M. J., Alam, O., Alam, J., Bano, F., Alam, P., & Shrivastava, N. (2016). Recent Review on Indole: A Privileged Structure Scaffold. *Int. J. Pharm. Sci. Res*, *7*, 51-62.

Nardi, D., Tajana, A., Leonardi, A., Pennini, R., Portioli, F., Magistretti, M. J., & Subissi, A. (1981). Synthesis and anticonvulsant activity of N-(benzoylalkyl) imidazoles and N- (. omega. -phenyl-. omega. -hydroxyalkyl) imidazoles. *Journal of Medicinal Chemistry*, *24*(6), 727-731.

Narnaware, P. H., & Shende, P. N. (2018).an overview in heterocyclic compounds and their versatile application. International Journal OF Current Engineering and Scientific Research(IJCESR), 5(4),2393-8374.

Nataraj, K. S., Rao, J. V., & Jayaveera, K. N. (2010). Synthesis and antimicrobial activity of new indole derivatives. *International Journal of Chemical Sciences*, *8*(1), 609-616.

Naureen, S., Chaudhry, F., Munawar, M. A., Ashraf, M., Hamid, S., & Khan, M. A. (2018). Biological evaluation of new imidazole derivatives tethered with indole moiety as potent α-glucosidase inhibitors. *Bioorganic Chemistry*, *76*, 365-369.

Odds, F. C. (1980). Laboratory evaluation of antifungal agents: a comparative study of five imidazole derivatives of clinical importance. *Journal of Antimicrobial chemotherapy*, *6*(6), 749-761.

Oh, k., Ishigaki, M., Hoshi, T., Youshizawa, Y., Shimura, Y. (2016). Synthesis of new indole derivatives structurally related to camlexin and biological evaluation of antibacterial activity. International Journal of Advances in Science Engineering and Technology,4(2),

Palmisano, G., Penoni, A., Sisti, M., Tibiletti, F., Tollari, S., & M Nicholas, K. (2010). Synthesis of indole derivatives with biological activity by reactions between unsaturated hydrocarbons and N-aromatic precursors. *Current Organic Chemistry*, *14*(20), 2409-2441.

Park, D. S., Kim, J. M., Lee, Y. B., & Ahn, C. H. (2007). QSID Tool: a new three-dimensional QSAR environmental tool. *Journal of computer-aided molecular design*, *22*(12), 873-883.

Patil, S. A., Addo, J. K., Deokar, H., Sun, S., Wang, J., Li, W., ... and Buolamwini, J. K.(2017). Synthesis, Biological Evaluation and Modeling Studies of New

Pyrido[3,4-b] indole Derivatives as Broad-Spectrum Potent Anticancer Agents. *Drug designing: open access*, 6(1).

Petitjean, M. (1992). Applications of the Radius-Diameter Diagram to the Classification of Topological and Geometrical Shapes of Chemical Compounds; *J. Chem. Inf. Comput. Sci. 32* 331–337.

Pingaew, R., Mandi, P., Prachayasittikul, V., Prachayasittikul, S., Ruchirawat, S., & Prachayasittikul, V. (2017). Synthesis, molecular docking, and QSAR study of sulfonamide-based indoles as aromatase inhibitors. *European Journal of Medicinal Chemistry*, *143*, 1604-1615.

Pozharskii, A. F., Garnovskii, A. D., & Simonov, A. M. (1966). Advances in the chemistry of imidazole. *Russian Chemical Reviews*, *35*(2), 122.

Prada-Gracia, D., Huerta-Yépez, S., & Moreno-Vargas, L. M. (2016). Application of computational methods for anticancer drug discovery, design, and optimization. *Boletín Médico Del Hospital Infantil de México (English Edition)*, *73*(6), 411-423.

Prior, A. M., Yu, X., Park, E. J., Kondratyuk, T. P., Lin, Y., Pezzuto, J. M., & Sun, D. (2017). Structure-activity relationships and docking studies of synthetic 2-arylindole derivatives determined with aromatase and quinone reductase 1. *Bioorganic & medicinal chemistry letters*, *27*(24), 5393-5399.

Qin, L. T., Liu, S. S., Chen, F., Xiao, Q. F., & Wu, Q. S. (2013). Chemometric model for predicting retention indices of constituents of essential oils. *Chemosphere*, *90*(2), 300-305.

Qin, L., Zhang, X., Chen, Y., Mo, L., Zeng, H., & Liang, Y. (2017). Predictive QSAR models for the toxicity of disinfection byproducts. *Molecules*, *22*(10), 1671.

Quazi, I., Sastry, V. G., & Ansari, J. A. (2017). Synthesis and Antimicrobial Activity of Indole Derivative Bearing the Pyrazole Moiety. *International Journal of Pharmaceutical Sciences and Research*, *8*(3), 1145.

Ramachandran, K. I., Deepa, G., & Namboori, K. (2008). *Computational chemistry and molecular modeling: principles and applications*. Springer Science & Business Media.

Rashamuse, T. J., Harrison, A. T., Mosebi, S., van Vuuren, S., Coyanis, E. M., & Bode, M. L. (2019). Design, synthesis and biological evaluation of imidazole and

oxazole fragments as HIV-1 integrase-LEDGF/p75 disruptors and inhibitors of microbial pathogens. *Bioorganic & Medicinal Chemistry*, *28*(1), 115210.

Rashamuse, T. J., Njengele, Z., Coyanis, E. M., Sayed, Y., Mosebi, S., & Bode, M. L. (2020). Design, synthesis and biological evaluation of novel 2-(5-aryl-1H-imidazol-1-yl) derivatives as potential inhibitors of the HIV-1 Vpu and host BST-2 protein interaction. *European Journal of Medicinal Chemistry*, *190*, 112111.

Reed, K.,Mereish, K., and Jensen, B. (1984). Quantitative Relationships Between Structure and Pharmacokinetic Parameters Using Molecular Connectivity

Revuelta, J., Machetti, F., & Cicchi, S. (2011). Five- Membered Heterocycles: 1, 3- Azoles. *Modern Heterocyclic Chemistry*, 809-923.

Robinson, B. (1963). The Fischer Indole Synthesis. *Chemical reviews*, *63*(4), 373-401.

Romero, D. H., Heredia, V. E. T., García-Barradas, O., López, M. E. M., & Pavón, E. S. (2014). Synthesis of imidazole derivatives and their biological activities. *J Chem Biochem*, *2*(2), 45-83.

Roy, P. P., & Roy, K. (2010). Docking and 3D-QSAR studies of diverse classes of human aromatase (CYP19) inhibitors. *Journal of molecular modeling*, *16*(10), 1597-1616.

Saberi, M. R., Shah, K., & Simons, C. (2005). Benzofuran-and furan-2-yl-(phenyl)-3-pyridylmethanols: Synthesis and inhibition of P450 aromatase. *Journal of enzyme inhibition and medicinal chemistry*, *20*(2), 135-141.

Sadiku, M. N. O., Musa, S. M., & Musa, O. M. (2017). Computational chemistry. Elixir Comp. Chem. 48184-48185.

Semire, B., & Oyebamiji, A. K. (2017). Theoretical studies on pyrazole derivatives as anti-breast cancer agents: DFT, QSAR and Docking methods. *Bulletin of Pharmaceutical Research*, *7*(3), 150.

Sewald, N., & Shaaban, M. (2017). Synthesis and biological activities of new bis-indole derivatives via microwave irradiation. *Z. Naturforsch*, *72*(9), 639-646.

Shaik, K. A., Ahmed, A., & Reddy, B. S. (2014). Synthesis, Characterization and Antibacterial Activity of Imidazole Derivatives of 1, 10-Phenanthroline and their

Complexes under Solvent-Free Condition*International Journal of Advanced Research in Chemical Science*, *1*(5), 1-6.

Sharma, D., Narasimhan, B., Kumar, P., Judge, V., Narang, R., De Clercq, E., & Balzarini, J. (2009). Synthesis, antimicrobial and antiviral evaluation of substituted imidazole derivatives. *European journal of medicinal chemistry*, *44*(6), 2347-2353.

Sharma, V., Kumar, P., & Pathak, D. (2010). Biological importance of the indole nucleus in recent years: a comprehensive review. *Journal of Heterocyclic Chemistry*, *47*(3), 491-502.

Slater, M.(2013). Outsourcing Computational Chemistry for Drug Discovery
Sujatha, G., Ramanathan, P., Pazhanisamy, T., & Anusuya, V. (2016). Synthesis, Spectral Characterization, Computational Studies and Antimicrobial Activities of Imidazole Derivatives. *DJ Journal Engineering Chemistry and Fuel*, *1*(4), 60-72.

Sumiya, T., Ishigaki, M., & Oh, K. (2017). Synthesis of imidazole and indole hybrid molecules and antifungal activity against rice blast. *Int J Chem Eng Appl*, *8*(3), 233-6.

Sun, X. Y., Liu, M. Y., Zhong, C. Y., Zheng, G. L., Lv, M. Y., Jing, B. T., ... & Wang, X. (2017). Synthesis and Antibacterial Evaluation of 2-Ethyl-1-(4-substituted) phenyl-1 H-imidazole Derivatives as Open-Chain Analogues of 7-Alkoxyl-4, 5-dihydro-imidazo [1, 2-a] quinolines. *Journal of the Brazilian Chemical Society*, *29*(4), 701-707.

Sundberg, R. J. (2000). Indole. *Kirk‐ Othmer Encyclopedia of Chemical Technology*.

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

Uno, J., Shigematsu, M. L., & Arai, T. (1983). Novel synergism of two antifungal agents, copiamycin and imidazole. *Antimicrobial agents and chemotherapy*, *24*(4), 552-559.

Valdez, C. A., Tripp, J. C., Miyamoto, Y., Kalisiak, J., Hruz, P., Andersen, Y. S., ... & Gillin, F. D. (2009). Synthesis and electrochemistry of 2-ethenyl and 2-ethanyl derivatives of 5-nitroimidazole and antimicrobial activity against Giardia lamblia. *Journal of medicinal chemistry*, *52*(13), 4038-4053.

Van Order, R. B., & Lindwall, H. G. (1942). Indole. *Chemical Reviews*, *30*(1), 69-96.

Verma, A., Joshi, S., & Singh, D. (2013). Imidazole: Having versatile activities. *Journal of Chemistry*, *2013*.

Vijesh, A. M., Isloor, A. M., Telkar, S., Arulmoli, T., & Fun, H. K. (2013). Molecular docking studies of some new imidazole derivatives for antimicrobial properties. *Arabian Journal of Chemistry*, *6*(2), 197-204.

Wang, F., Wang, X., Zhang, M. X., Yang, Y. H., & Zhu, H. L. (2015). Synthesis, biological evaluation and molecular modeling of 1 H-benzo [d] imidazole derivatives as novel anti-tubulin polymerization agents. *RSC advances*, *5*(91), 74425-74437.

Wang, H. (2010*). Fischer Indole Synthesis (Fischer Indolization). *Comprehencive Organic Name Reactions* (p 1069). Wiley
Wang, R., Shi, H. F., Zhao, J. F., He, Y. P., Zhang, H. B., & Liu, J. P. (2013). Design, synthesis and aromatase inhibitory activities of novel indole-imidazole derivatives. *Bioorganic & medicinal chemistry letters*, *23*(6), 1760-1762.

Wang, T., Wu, M. B., Lin, J. P., & Yang, L. R. (2015). Quantitative structure–activity relationship: promising advances in drug discovery platforms. *Expert opinion on drug discovery*, *10*(12), 1283-1300.

Wnuk, M., Marszałł, M., Zapęcka, A., Nowaczyk, A., Krysiński, J., Romaszko, J., ... & Buciński, A. (2013). Prediction of antimicrobial activity of imidazole derivatives by artificial neural networks. *Open Medicine*, *8*(1), 1-15.

Xie, H., Qiu, K., & Xie, X. (2014). Pharmacophore modeling, virtual screening, and 3D-QSAR studies on a series of non-steroidal aromatase inhibitors. *Medicinal Chemistry Research*, *24*(5), 1901-1915.

Xu, Y., He, Z., Liu, H., Chen, Y., Gao, Y., Zhang, S., ... & Liu, Y. (2020). 3D-QSAR, molecular docking, and molecular dynamics simulation study of thieno [3, 2-b] pyrrole-5-carboxamide derivatives as LSD1 inhibitors. *RSC Advances*, *10*(12), 6927-6943.

Yadav, B. P., Ahmad, I., & Thakur, M. (2016). Synthesis of some novel indole derivatives as potential antibacterial, antifungal and antimalarial agents. *IOSR J. Pharm*, *6*, 27-33.

Yakovychuk, N. D., Deyneka, S. Y., Grozav, A. M., Humenna, A. V., Popovych, V. B., & Djuiriak, V. S. (2018). Antifungal activity of 5-(2-nitrovinyl) imidazoles and their derivatives against the causative agents of vulvovaginal candidiasis. *Regulatory Mechanisms in Biosystems*, *9*(3).

Young, D. C. (2001). A Practical Guide for Applying Techniques to Real-World Problems. *Wiley eInterscience New York*.

Yousif, M. N., Hussein, H. A., Yousif, N. M., El-Manawaty, M. A., & El-Sayed, W. A. (2019). Synthesis and Anticancer Activity of Novel 2-Phenylindole Linked Imidazolothiazole, Thiazolo-s-triazine and Imidazolyl-Sugar Systems. *Journal of Applied Pharmaceutical Science*, *9*(01), 006-014.

Zampieri, D., Mamolo, M. G., Vio, L., Banfi, E., Scialino, G., Fermeglia, M., ... & Pricl, S. (2007). Synthesis, antifungal and antimycobacterial activities of new bis-imidazole derivatives, and prediction of their binding to P45014DM by molecular docking and MM/PBSA method. *Bioorganic & medicinal chemistry*, *15*(23), 7444-7458.

Zhang, L., Peng, X. M., Damu, G. L., Geng, R. X., & Zhou, C. H. (2013). Comprehensive review in current developments of imidazole- based medicinal chemistry. *Medicinal research reviews*, *34*(2), 340-437.

Zhang, M., Ding, Y., Qin, H. X., Xu, Z. G., Lan, H. T., Yang, D. L., & Yi, C. (2019). One-pot synthesis of substituted pyrrole–imidazole derivatives with anticancer activity. *Molecular Diversity*, 1-8.

Zhao, F., Lu, W., Su, F., Xu, L., Jiang, D., Sun, X., ... & Cao, F. (2018). Synthesis and potential antineoplastic activity of dehydroabietylamine imidazole derivatives. *MedChemComm*, *9*(12), 2091-2099.

Zheng, X., Ma, Z., & Zhang, D. (2020). Synthesis of imidazole-based medicinal molecules utilizing the van leusen imidazole synthesis. *Pharmaceuticals*, *13*(3), 37.

Zhuang, S. H., Lin, Y. C., Chou, L. C., Hsu, M. H., Lin, H. Y., Huang, C. H., ... & Huang, L. J. (2013). Synthesis and anticancer activity of 2, 4-disubstituted furo [3, 2-b] indole derivatives. *European journal of medicinal chemistry*, *66*, 466-479.
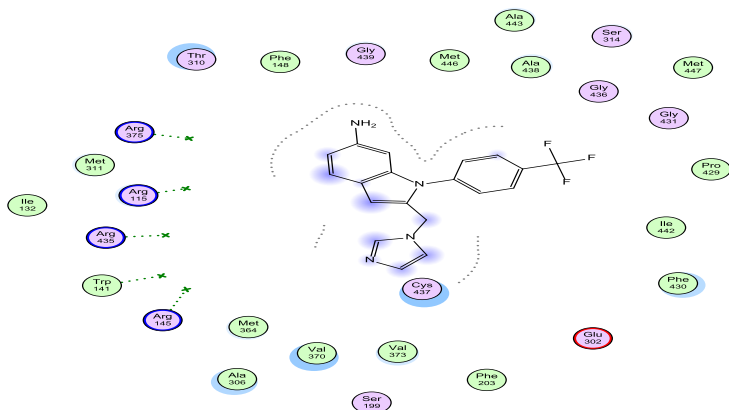
Appendix

2D binding mode of 17 new designed indole-imidazole derivatives with the active site of aromatase containing iron porphyrin (haem)
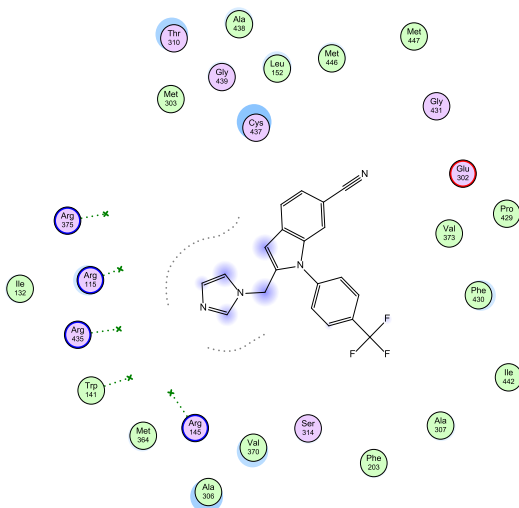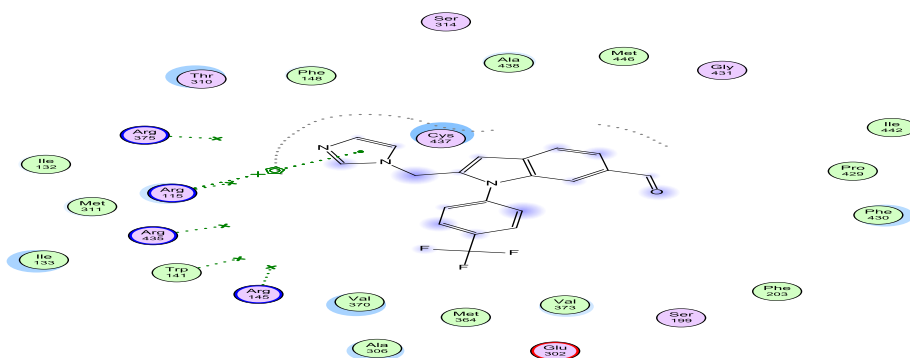

Appendix 1: interaction mode of compound I with the receptor (Aromatase with haem)
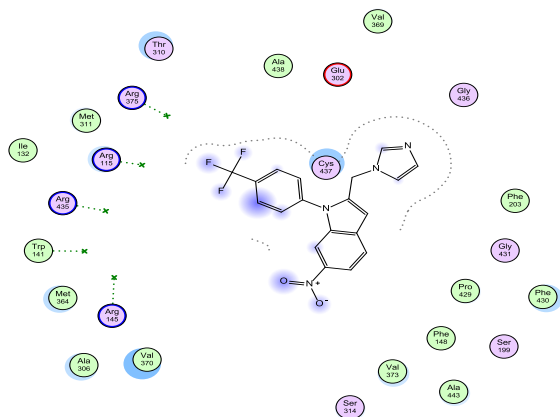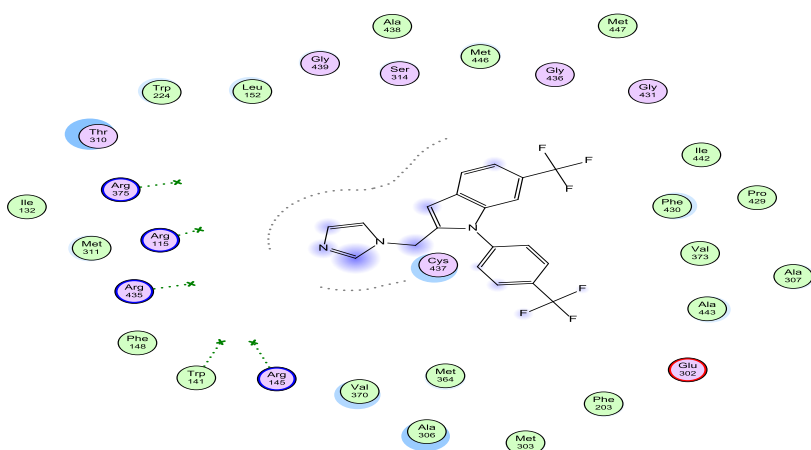

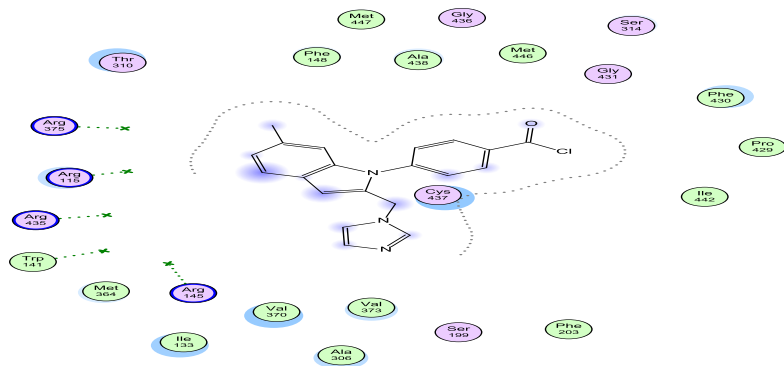Appendix 2: interaction mode of compound II with the receptor (Aromatase with haem)

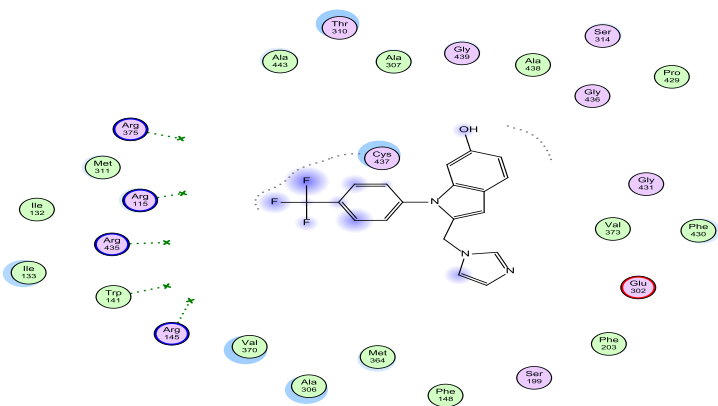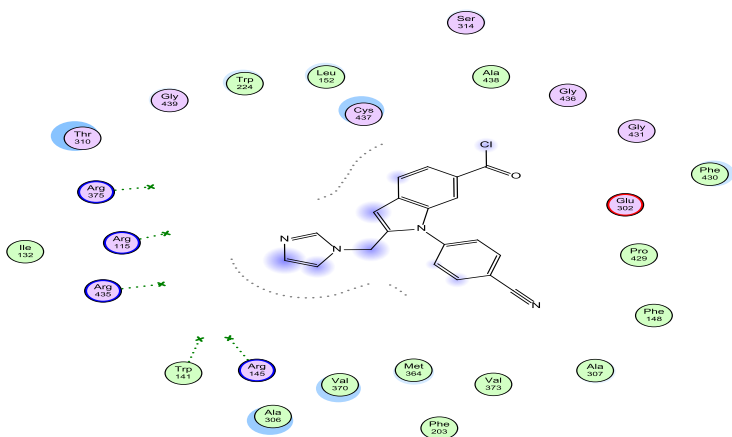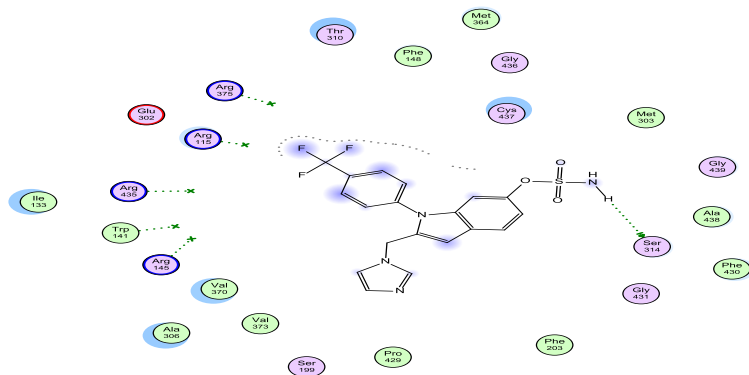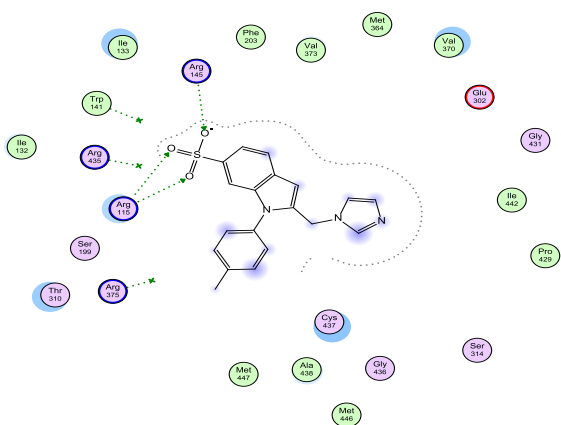Appendix 3: interaction mode of compound III with the receptor (aromatase with haem)

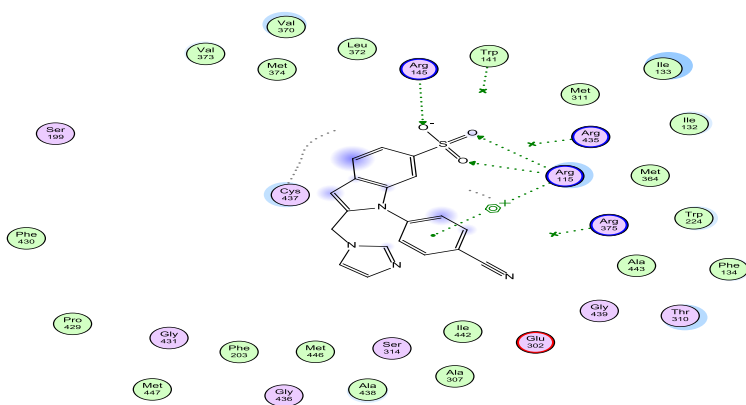Appendix 4: interaction mode of compound XI with the receptor (aromatase with haem)

Appendix 5: interaction mode of compound XII with the receptor (aromatase with haem)

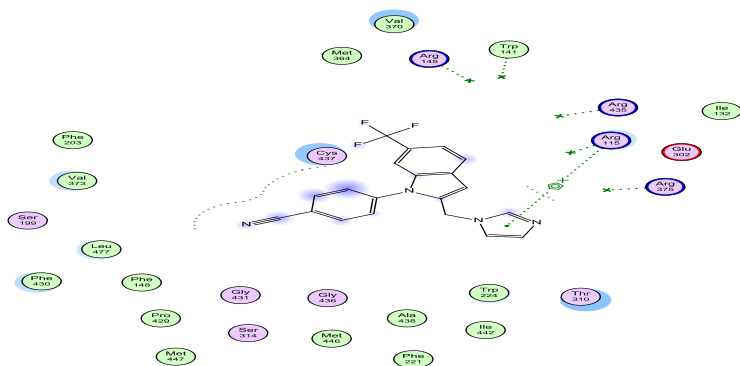Appendix 6: interaction mode of compound XXXIII with the receptor (aromatase with haem)

Appendix 7: interaction mode of compound LXV with the receptor (aromatase with haem)

Appendix 8: interaction mode of compound LXXVIII with the receptor (aromatase with haem)

Appendix 9: interaction mode of compound CVII with the receptor (aromatase with haem)

Appendix 10: interaction mode of compound XV with the receptor (aromatase with haem)

Appendix 11: interaction mode of compound LXXXIV with the receptor (aromatase with haem)

Appendix 12: interaction mode of compound CIX with the receptor (aromatase with haem)
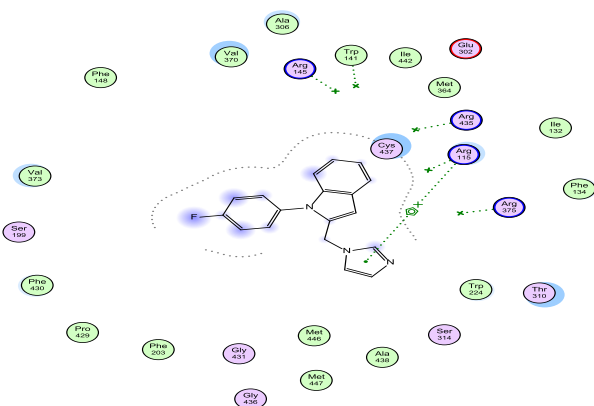

Appendix 13: interaction mode of compound CXI with the receptor (aromatase with haem)
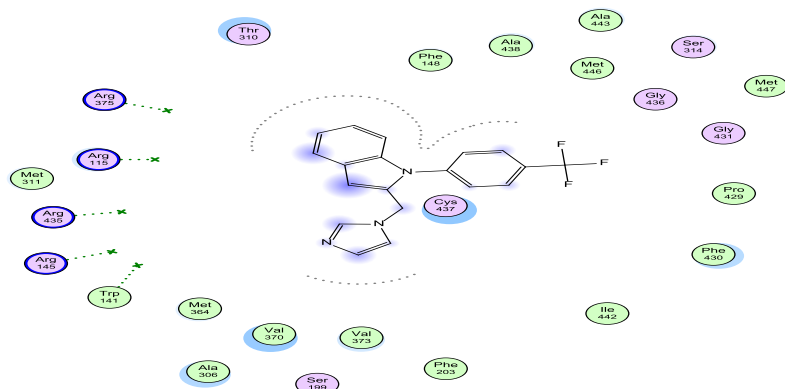

Appendix 14: interaction mode of compound CXII with the receptor (aromatase with haem)
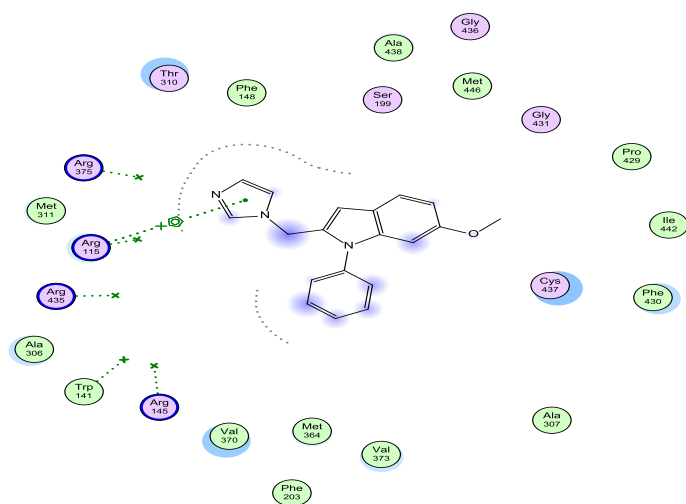
Appendix 15: interaction mode of compound LXXVII with the receptor (aromatase with haem)

Appendix 16: interaction mode of compound 10a with the receptor (aromatase with haem)

Appendix 17: interaction mode of compound 10d with the receptor (aromatase with haem)

Appendix 18: interaction mode of compound 10k with the receptor (aromatase with haem)