IN SILICO APPROACH TO UNDERSTANDING THE MECHANISM OF ACTION AND PREDICT THE CHEMICAL TOXICITY RELATED TO HUMAN AROMATASE ENZYME

Ana Yisel Caballero Alfonso

Doctoral Dissertation
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

Supervisor: Prof. Marjana Novič, National Institute of Chemistry, Ljubljana, Slovenia
Co-Supervisor: Prof. Emilio Benfenati, Mario Negri Institute for Pharmacological Research, Milano, Italy

Evaluation Board:
Prof. Dr. Eva Žerovnik, Chair, IPS and Jožef Stefan Institute, Ljubljana, Slovenia
Dr. Marjan Vračko Grobelšek, Member, National Institute of Chemistry, Ljubljana, Slovenia
Prof. Dr. Giuseppina Gini, Member, Politecnico di Milano, Milano, Italy

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

Ana Yisel Caballero Alfonso

IN SILICO APPROACH TO UNDERSTAND THE MECHANISM OF ACTION AND PREDICT THE CHEMICAL TOXICITY RELATED TO HUMAN AROMATASE ENZYME

Doctoral Dissertation

IN SILICO PRISTOP ZA RAZUMEVANJE MEHANIZMA DELOVANJA IN NAPOVEDOVANJE KEMIJSKE TOKSIČNOSTI, POVEZANE S ČLOVEŠKIM ENCIMOM AROMATAZO

Doktorska disertacija

Supervisor: Prof. Marjana Novič

Co-Supervisor: Prof. Emilio Benfenati

Ljubljana, Slovenia, August 2022

To my parents and God

Acknowledgments

Initially, I would like to express my deepest gratitude to the Marie Sklodowska-Curie Action - Innovative Training Network project in3, grant number 721975, for providing the funding for PhD research, and to the Istituto di Ricerche Farmacologiche “Mario Negri” IRCCS, Milano, Italy, for providing the computational software and other resources.

My sincere gratitude to my supervisors Prof. Emilio Benfenati and Prof. Marjana Novič for the unceasing support of my Ph.D. study and related research, for their patience, inspiration, and immense knowledge. Their guidance helped me in all the time of research.

Besides my advisors, I would like to thank Prof. Chayawan who was a colleague, mentor, and a friend during all these years. My sincere thanks for the valuable discussions, hours of sacrifice, patience, and his crucial support to the research. I could not have imagined having a better support for my Ph.D. study.

My sincere thanks also go to Prof. Mark T.D. Cronin, for his valuable suggestions, motivation, his insightful comments, encouragement, constant availability, and his constructive criticism during this investigation.

My gratitude also goes to Dr. Domenico Gadaleta for his confidence, interesting discussions and all the support throughout the project.

I thank my colleagues and professors from the in3 project: Kristijan Vuković, Dr. Nicoleta Spinu, Dr. Liadys Mora Lagares, Dr. Ivo Djidrovski (https://uk.linkedin.com/in/ivo-djidrovski-741508124) and the rest of the in3 project for the friendly advice and for all the fun we have had in the last four years. Particularly, I am grateful to my professors Dr. Paul Jennings and Dr. Maxime Culot for their support, confidence and for a really enjoyable time spent together.

My genuine gratitude goes to Prof. Yunierkis Perez Castillo for his motivation, confidence, and fresh ideas during the darkest times.

A very special gratitude goes out to the Bioinformatics Laboratory of Liverpool John Moores University and L'Oréal Research and Innovation Digital Sciences department for the unforgettable and meaningful secondments.

Finally, I owe a large amount of gratitude to my parents, my sister, my aunt, my husband, Lianet and Sonny for supporting me spiritually and economically throughout my PhD study, writing this thesis and my life in general.

Abstract

Aromatase is an enzyme member of the cytochrome P450 superfamily coded by the CYP19A1 gene. Its main biological function in humans is the conversion of androgens into estrogens, transforming androstenedione into estrone and testosterone into estradiol. This enzyme is present in several tissues, and it has a key role in maintaining the androgens and estrogens level, and thereby the endocrine regulation system. Azole compounds, which are used as agrochemicals and pharmaceuticals, can be potential endocrine disruptors and thus a reason for chemical safety and human health concerns. A toxicological evaluation of commonly used azole-based drugs and agrochemicals with respect to CYP19A1 is currently requested by the European Union - Registration, Evaluation, Authorization and Restriction of Chemicals (EU-REACH) regulations to screen hazardous chemicals for chemical safety and ecotoxicology concerns. During the last decades, the reduction of *in vivo* tests has become indispensable in terms of resources and animal testing. Much progress has been made in understanding steroid-aromatase interactions using *in silico* methods, but nonsteroidal ligands such as azoles, which are widely used, have also raised potential safety concerns that remain to be explored with respect to CYP19A1 for human health. Therefore, the overall aim of this research was to explore the chemical interactions between azoles and the human aromatase CYP19A1 enzyme by using different *in silico* approaches, which allows to predict the biological and toxicological profile of new azoles and also provides a deeper mechanistic understanding and rationality in relation to currently available evidence with potential applications in drug discovery and risk assessment. The first part of the present work describes the development of mechanistic structural alerts of azole-based chemicals as a guide for the evaluation of activity and toxicity, and their application as classification models. The second part of this thesis uses the structure-activity relationships to explore the structural factors of azoles that affect activity/toxicity on human aromatase, which may be helpful for medicinal chemistry in the design and synthesis of new compounds against breast cancer. Finally, the thesis addresses the development of a read-across workflow and chemical categories formation to support toxicity characterization of chemicals and data gap feeling. Overall, the findings contribute to understanding the structural requirements of azole chemicals affecting the activity of human aromatase CYP19A1 using different approaches, improve the knowledge of the mechanism of action of azoles on aromatase, and explain the effect of known compounds. Other benefits to the field include streamlining the design of new azole-based drugs/chemicals according to desired medicinal/chemical applications, opening new opportunities for drug screening with anticipation of potential candidates, suggesting recommendations, and establishing future directions in terms of molecular moieties that might be associated with a particular effect under investigation.

Povzetek

Aromataza je encim iz superdružine citokromov P450, ki jo kodira gen CYP19A1. Njegova glavna biološka funkcija pri ljudeh je pretvorba androgenov v estrogene, konverzija androstendiona v estron in testosterona v estradiol. Ta encim je prisoten v več tkivih in ima ključno vlogo pri vzdrževanju ravni androgenov in estrogenov v telesu ter s tem endokrinega regulacijskega sistema. Azoli, ki se uporabljajo kot agrokemikalije in farmacevtski izdelki, so potencialni endokrini motilci in zato predstavljajo nevarnost za zdravje ljudi. Trenutno predpisi Evropske unije za preverjanje nevarnih kemikalij glede kemijske varnosti in ekotoksikoloških vprašanj, EU-REACH (registracija, vrednotenje, avtorizacija in omejevanje kemikalij), zahtevajo toksikološko oceno pogosto uporabljenih zdravil na osnovi azolov in agrokemikalij glede vpliva na CYP19A1. V zadnjih desetletjih se je pokazalo, da je potrebno zmanjšati število *in vivo* preskusov zaradi omejevanja virov in potrebnega zmanjševanja poskusov na živalih. Z uporabo metod *in silico* je bil dosežen velik napredek pri razumevanju interakcij med steroidi in aromatazo, vendar so nesteroidni ligandi, kot so pogosto uporabljani azoli, prav tako sprožili morebitne varnostne pomisleke za zdravje ljudi, ki jih je treba še raziskati v zvezi z vplivom na CYP19A1. Zato je bil splošni cilj te raziskave preučiti kemične interakcije med azoli in encimom človeške aromataze CYP19A1 z uporabo različnih *in silico* pristopov, kar omogoča napovedovanje biološkega in toksikološkega profila novih azolov. Zagotavlja tudi globlje mehanično razumevanje in racionalnost glede na trenutno razpoložljive dokaze z možno uporabo pri odkrivanju zdravil in oceni tveganja. Prvi del pričujočega dela opisuje razvoj mehanističnih strukturnih opozoril kemikalij na osnovi azola kot vodila za vrednotenje aktivnosti in toksičnosti ter njihovo uporabo za klasifikacijske modele. Drugi del doktorske disertacije uporablja razmerja med kemijsko strukturo in aktivnostjo za raziskovanje strukturnih dejavnikov azolov, ki vplivajo na aktivnost/toksičnost za človeško aromatazo, kar je lahko v pomoč medicinski kemiji pri načrtovanju in sintezi novih spojin proti raku dojke. Nazadnje doktorska disertacija obravnava razvoj delovnega toka navzkrižnega branja in oblikovanje kategorij kemikalij za podporo karakterizacije toksičnosti kemikalij ter za prepoznavanje in zapolnjevanje podatkovnih vrzeli. Na splošno ugotovitve, pridobljene z uporabo različnih pristopov, prispevajo k razumevanju strukturnih zahtev azolnih kemikalij, ki vplivajo na aktivnost človeške aromataze CYP19A1, izboljšujejo poznavanje mehanizma delovanja azolov na aromatazo in pojasnjujejo učinek znanih spojin. Drugi koristni prispevki za to področje vključujejo racionalizacijo načrtovanja novih zdravil/kemikalij na osnovi azolov glede na želene medicinske/kemične aplikacije, odpiranje novih priložnosti za pregledovanje zdravil z namenom ponovne uporabe potencialnih kandidatov, predlaganje priporočil in določanje prihodnjih usmeritev na osnovi molekulskih fragmentov in podstruktur, povezanih z določenim učinkom, ki ga preiskujemo.

Contents

List of Figures

List of Tables

Abbreviations

µM	...	Micromolar
17β-HSD	...	17β-hydroxysteroid dehydrogenase
3β-HSD	...	3β-hydroxysteroid dehydrogenase/Δ5−4 isomerase
Acc	...	Accuracy
AD	...	Applicability Domain
AIs	...	Aromatase Inhibitors
AOP	...	Adverse Outcome Pathway
ASD	...	Androstenedione
CSV	...	Comma Separated Values
CYP11A1	...	Cholesterol side-chain cleavage enzyme
CYP17A1	...	Steroid 17-α-hydroxylase/17,20-lyase
CYP19A1	...	Human aromatase
CYP450	...	Cytochrome p450
DHEA	...	Dehydroepiandrosterone
ECHA	...	European Chemicals Agency
EDG	...	Electron Donating Group
ER	...	Estrogen Receptor
FN	...	False Negatives
FP	...	False Positives
HTST	...	16α-hydroxytestosterone
IC_{50}	...	Half maximal inhibitory concentration
LR	...	Likelihood Ratio
MW	...	Molecular Weight
N	...	Negatives
NSAIs	...	Non-steroidal aromatase inhibitors
P	...	Positives
PC	...	Physical Chemical
PSA	...	polar surface area
QSAR	...	Quantitative Structure Activity Relationships
SAR	...	Structure Activity Relationships
SC	...	Source Compound
Se	...	Sensitivity
SMILES	...	Simplified Molecular Input Line Entry Specification
Sp	...	Specificity

STARD1	...	Steroidogenic acute regulatory protein
T	...	Testosterone
TC	...	Target Chemical
TN	...	True Negatives
TP	...	True Positives
TPSA	...	Topological Polar Surface Area

Chapter 1

Introduction

1.1 Human Aromatase

1.1.1 Structure of human aromatase

The aromatase enzyme, such as oxidoreductase, belongs to the cytochrome p450 family and it is encoded by the CYP19A1 gene, located on chromosome 15 (15q21.21) with approximately 123 kb. Out of 123kb, 30 kb represents the coding region with nine exons and an additional larger region (nine untranslated exons I) is starting at the 50 ends. Aromatase is a complex that consists of two proteins: a ubiquitous NADPH-cytochrome P450 reductase and a specific cytochrome P450 aromatase (P450arom), which contains the heme and steroid binding pocket [1, 2]. Aromatase is a glycosylated and integral protein of the endoplasmic reticulum membrane with 58 kDa [1]. The catalytic site of aromatase contains a heme group, and a steroidal binding pocket, which intervene in the main step of the steroidogenesis process [3].

In the year 2009, Ghosh et al. proposed the first X-ray structure of the enzyme CYP19A1, elucidating a rigid core [3]. The aromatase was classified as a functional monomer containing a sequence of 503 amino acids, arranged into twelve α-helices (labelled from A to L) and ten β-strands (numbered from 1 to 10) [3, 4]. Diverse packed hydrophobic residues have been located at the active site of the enzyme, which help to stack the α-face backbone of androstenedione [5, 6]. It is not surprising that androgens fit constrainedly into the androgen-specific cleft, since the active site with less than 400$Å^3$ is small and compact [5, 6]. The catalytic site of aromatase is located at the juncture of the I and F helices, β-sheet 3, and B-C loop. The loop between helices B and C, β7 and β8, β9 and β10 is part of the enzyme active site [68]. Its binding site consists of residues Ile305, Ala306, Asp309, Thr310, Phe221, Trp224, Arg115, Ile133, Phe134, Val370, Leu372, Val373, Met374, Leu477 and Ser478 [7]. The structure of human aromatase in complex with androstenedione is presented in Figure 1.1.

Studies evaluating mutagenicity have recognized that the residues Phe221, Trp224 and Met374 are particularly significant for the binding with natural substrates, suggesting also that their lacking can reduce the catalytic activity [1, 7]. The polar and non-polar regions cast by the heme group with a single polypeptide chain enclosing 503 amino-acids residues in coordination of iron ion with androstenedione are shown in Figure 1.1 [8-10]. The polar region can be recognized where the keto oxygen atoms of androstenedione make hydrogen bonds with Asp309 and Met374, and also weakly interact with Arg115 residues of enzyme. Androstenedione forms the van der Waals interactions with amino acid residues, namely Ala306, Phe134, Phe221, Arg115, Leu477, Ile133, Thr310, Trp224, Val370, Val373 and Met374 at the active site of aromatase enzyme [8].

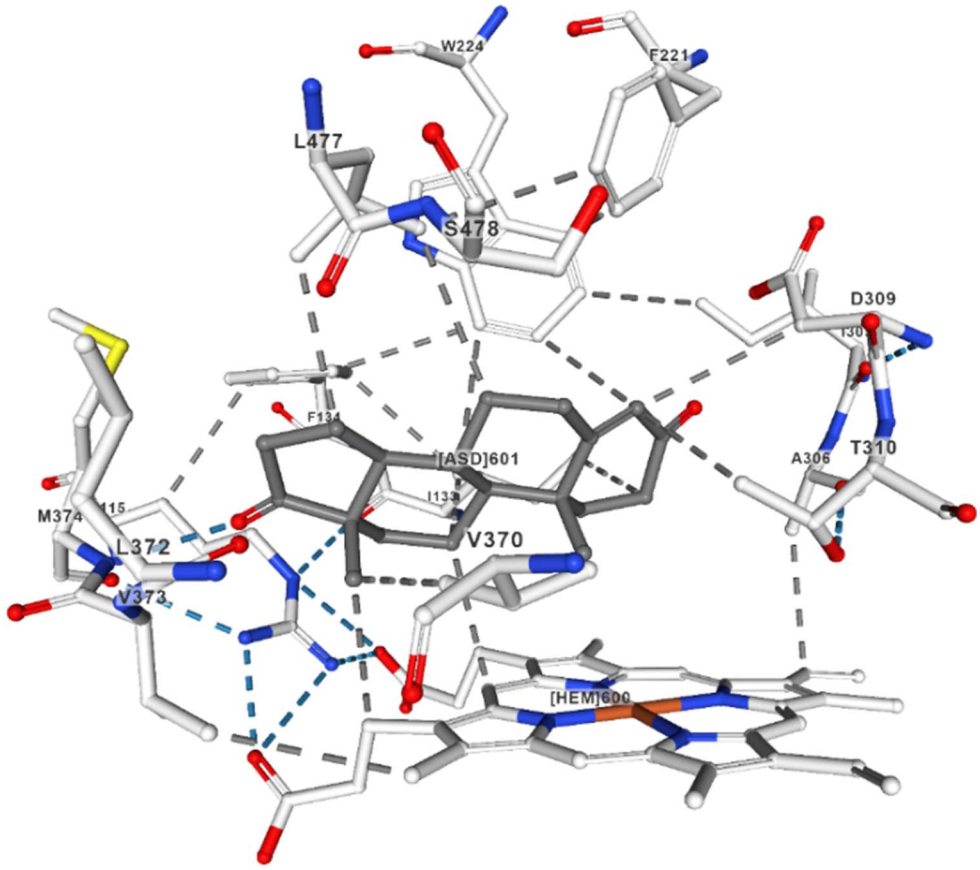


Figure 1.1: Representation of the ligand interaction of androstenedione and human placental aromatase cytochrome P450 (PDB code 5JL9) [10].

The docking experiments have concluded the role of the aromatic interactions through the F134, F221 and W224 amino acid residues, polar interactions through D309, T310, S478 and M374 amino acid residues and non-polar interactions through A306, A307, V370, L372 and L477 amino acid residues as common for all the azole aromatase inhibitors (70,71,55,). It must be noted that the inhibition may be a result of the allosteric effects or directly binding to the heme prosthetic group, however, the similar chemicals with respect to basic parent ring, shape and size can be assumed to follow the same mode of actions [3, 11]. The elucidation of the enzyme structure, the binding site and the type of interactions with the natural substrates lead to the development of new and potent aromatase inhibitors, and impulse the research in other ways to modulate the enzyme as allosteric modulation and its phosphorylation status [6, 12, 13].

1.1.2 Expression of human aromatase

The aromatase CYP19A1 enzyme is a membrane-bound protein, which is an integral part of the endoplasmic reticulum of estrogen producing cells and mainly plays an active role in the epithelial cells [14, 15]. Aromatase is widely expressed in human tissues, including adult and fetal brain, placenta, skin fibroblasts, adipose tissue, gonads, bone, blood vessels, endometrium, and breast tissue [16-19]. In women, the expression of aromatase fluctuates across the age, while during the fertile age, the aromatase is mainly expressed in the granulosa cells of ovaries, during the post-menopausal period the enzyme expression became

mainly present at peripheral tissues, as for example, adipose breast tissue. The expression of aromatase is also significant in the placenta of pregnant women [1, 20]. In breast cancer, the tumoral stromal cells and the adipose and epithelial cells surrounding the carcinoma also highly express the enzyme [21]. On the other hand, aromatase expression is regulated by tissue-specific promoters and associated enhancers and suppressors, which are controlled by particular mechanisms, therefore the regulation of oestrogens biosynthesis takes place under a complex tissue-specific process [22, 23].

1.1.3 Mechanism of action of human aromatase

The aromatase CYP19A1 enzyme is responsible for the main steps in the conversion of androstenedione into oestrogen during steroid genesis. Aromatase substrates are androstenedione, testosterone and 16α-hydroxytestosterone which by aromatization of the A-ring are converted into C18 steroids, estrone, 17β-estradiol and 16α-estriol, respectively [4, 24]. The mechanism of action of aromatase is not completely elucidated, however there is agreement regarding the overall process which involves three oxidative steps: The first two steps are hydroxylations of the C19 methyl group with a posterior dehydration of the products into aldehyde and the third step is the loss of the C-19 carbon atom in the form of a molecule of formic acid, leading to the aromatization of A-ring; however this third hydroxylation is not well understood [25, 26].

1.1.4 Estrogen biosynthesis

To properly understand the biological relevance of human aromatase and its essential function, it is necessary to emphasize the steroidogenesis process. The cholesterol acquired by diet is the main substrate for steroidogenesis [27]. This organic molecule is transformed in gonads, adrenal cortex, and adipose tissue to the 21-carbon (pregnanes, progestogens), 19-carbon (androstanes), and 18-carbon (estranes), with the higher rate of synthesis in the granulosa cells of ovaries[28].

During steroidogenesis, the cholesterol is translocated into the inner mitochondrial membrane where the steroidogenic acute regulatory protein acts as a regulator of the process and, probably, as a shuttle enzyme [29]. The StAR is the rate-limiting step of steroidogenesis in all tissues, and its expression is controlled for several biological mechanisms. The cholesterol is converted by the cholesterol side-chain cleavage enzyme to pregnolone which takes action as a precursor for all steroid hormones, and can spread between neighbouring granulosa and theca cells of the ovary [30]. The next steps in the synthesis process will involve the enzymes CYP17A1 and 3β-HSD, via dehydroepiandrosterone to derive androstenedione. Androstenedione is then converted into other androgens, including testosterone and dihydrotestosterone, when andostedione is diffusing through the basal lamine to the granulose cells; the CYP19A1 enzyme catalyzes its conversion to estrone. Estrone is then transformed to estradiol by the enzyme 17β-HSD. Aromatase also catalyzes the conversion of testosterone into estradiol and estrone in peripheral tissues, including adipose cells and bone [31].

Local synthesis of estrogens is also performed in males at the level of the reproductive tract, including Sertoli cells, Leydig cells, and mature spermatocytes. Normally estrogens are generated by the ovaries, but smaller amounts can be generated in other organs such as liver, pancreas, adrenal glands, adipose tissue, and breast, however the steroidogenesis at non-gonadal sites is unusual due to their inability to produce C19 steroids from

cholesterol. At this level, a local conversion mediated by aromatase is necessary but dependent on C19 steroids transportation from other tissues [32, 33]. During the pregnancy estrogen is also synthesized by the placenta. On the other hand, estriol is produced mostly in the liver by 16α-hydroxylation of estradiol and estrone by CYP enzymes, this occurs in non-pregnant women. At last, in menopausal women, estrone synthesis is associated to the aromatization of androstenedione at extra-glandular tissues. Other synthetic routes of estradiol implicate the conversion of estrone, by the enzyme 17β-hydroxysteroid dehydrogenase in peripheral tissues, including adipose and breast tissue, vascular endothelium, smooth muscle cells, brain tissue, and bone cells, where it is metabolized or passes into the circulation in small quantities [34]. The synthetic route of estrogens is represented within Figure 1.2.

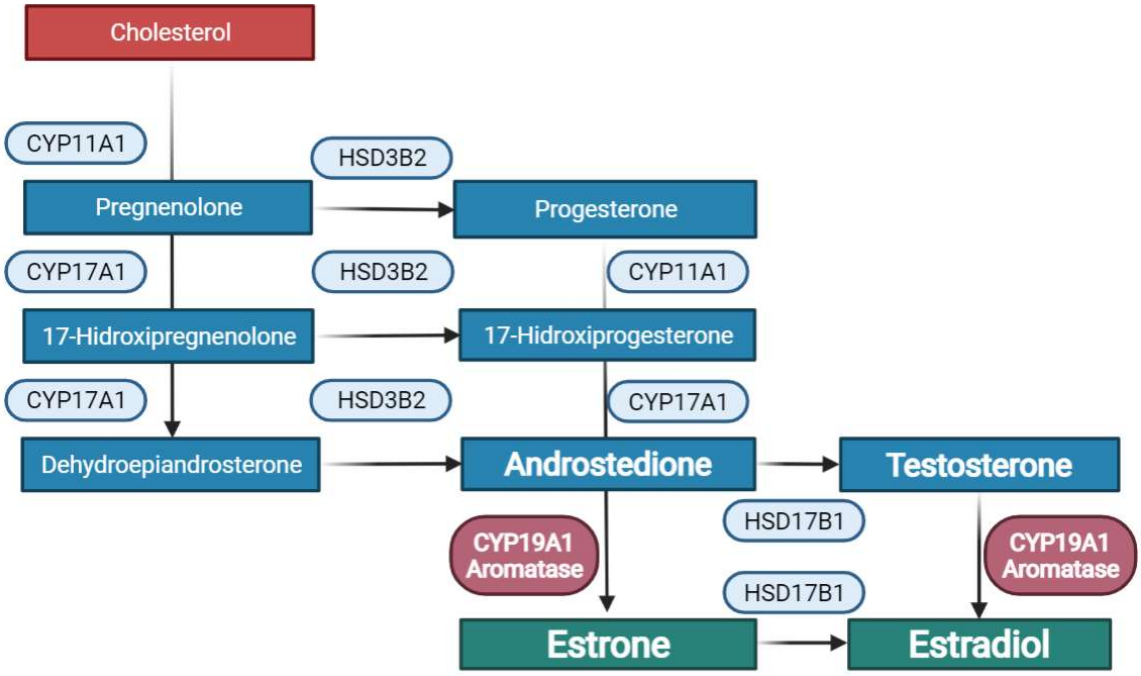


Figure 1.2: Schematic representation of estrogen biosynthesis pathway [33].

1.1.5 Biological relevance

Several physiological functions have been attributed to the estrogens, to which balance is essential for the normal functioning of the organism. These comprise the regulation of the menstrual cycle and reproduction, the bone mineral density, the brain function, cholesterol mobilization, development of breast tissue and sexual organs, and control of inflammation [35]. In females, the sexual development is estrogen dependent, determining this the sexual characteristics, while estradiol stimulates epithelial cell proliferation in the uterine endometrium and mammary glands starting in puberty [36]. In the period of pregnancy, estrogens produced by the placenta make the mammary gland ready for lactation [37]. Tendentially, estrogen relevance is associated to women, however lower levels of estrogens are essential in men for biological functions as sperm maturation, erectile function and maintenance of a healthy libido [38]. Estrogenic functions are mediated by estrogen receptors, which are of extreme importance for the process. As presented in the

previous section, aromatase is the only enzyme able to perform the conversion of androgens into estrogens in vertebrates [4], consequently, and considering the high relevance of estrogens for the organism, human aromatase became an essential enzyme to keep the biological balance and normal functioning of the organism.

1.2 Aromatase Ligands

1.2.1 Aromatase inhibitors (Antagonists)

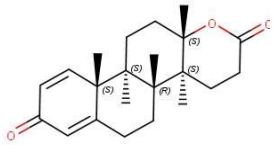
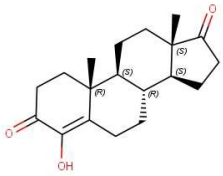
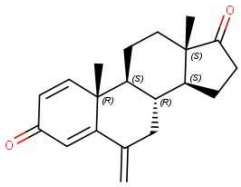
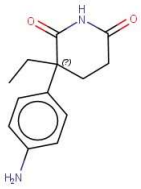
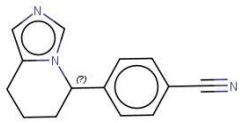
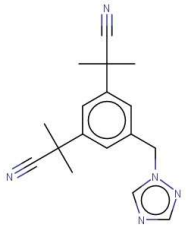
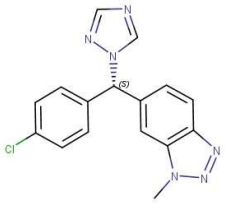
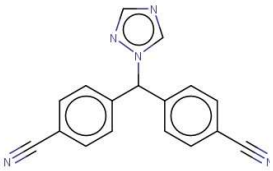
Due to its unique aromatization reaction, comprising the loss of the C-19 carbon and conversion of the steroidal-A-ring to an aromatic ring, aromatase is a topping target for inhibition to develop selective inhibitors. The aromatase inhibitors consist of two classes of compounds, generally classified as steroidal and nonsteroidal aromatase inhibitors. The steroidal compounds constitute a class of competitive, selective inhibitors that later demonstrated to be also enzyme inactivators [39]. These first aromatase inhibitors were characterized by a high degree of analogy with the natural substrate androstenedione [40]. The steroidal inhibitors interact with the steroid binding site and are subjected to the usual mechanism of action of aromatase. However, its products bind covalently the enzyme either very tightly or irreversibly causing its inactivation, therefore, the synthesis of new enzymes constitute the only alternative [41]. Formestane and later Exemestane were steroidal aromatase inhibitors approved in several countries for the endocrine treatment of breast cancer [42].

In the other site, the nonsteroidal aromatase inhibitors have been derived from antifungal drugs such as ketoconazole, and are characterized by binding the enzyme reversibly. This class of inhibitors contains a heteroatom in the heterocyclic moiety, and interferes with the steroidal hydroxylation. Unlike steroidal inhibitors, the non-steroidal inhibitors are less enzyme specific, indeed selectivity constitutes one of the main challenges in developing these types of inhibitors. For example, an interference has been observed with the enzyme 18- hydroxylase responsible for the production of aldosterone [26, 42]. In the past two decades, the non-steroidal aromatase inhibitors were found as potential inhibitors for the treatment of estrogen-related diseases as they produce fewer side effects [43].

According to their chronological order of appearance, the AIs are split into three generations. The first generation, Aminoglutethimide(non-steroidal) [44], was applied in clinical studies and marketed in 1970s, but showed to interact with other cytochrome enzymes at the doses required to inhibit aromatase, as for example 11 [3-hydroxylase], this low affinity leads to toxic effects at the level of adrenal steroid cortisol. The importance of selective inhibition then becomes crucial for effective and safe drugs [45, 46], consequently, aminoglutemide was retired. Testolactone was the other first generation AI (steroidal) but with lower potency than aminoglutemide [47]. The second generation including fadrozole and vorozole (non-steroidal) were less selective and reduced aldosterone and cortisol production besides aromatase. Both were poorly tolerated and had narrow clinical efficacy [48]. In addition, formestane (steroidal) was also a second generation AI, however this exhibited poor oral bioactivity and limited clinical efficacy [47]. Although the higher potency of those chemicals in regard to aminoglutemide ~ 700 times, the low selectivity and specificity led to several adverse effects and as a result limited their clinical efficacy [47, 49]. The third-generation AIs are highly selective for the enzyme aromatase and are fairly well tolerated. Currently, anastrozole, letrozole, (non-steroidal) and examestane (steroidal) are approved as a first line of therapy for the estrogen-dependent breast cancer

for post-menopausal women [50] [51]. Table 1.1 represents the three generations of aromatase inhibitors grouped by structural class (steroidal and non-steroidal).

Table 1.1: Chemical structure of the steroidal and non-steroidal aromatase inhibitors from the three generations.

	First generation	Second generation	Third generation
Steroidal			
	Testolactone	Formestane	Examestane
Non-steroidal			
	Aminoglutethimide	Fadrozole	Anastrozole
			
		Vorozole	Letrozole

AIs not only inhibit the ovarian synthesis of estrogens but also act at the peripheral site. In postmenopausal women, this peripheral conversion is the major source of estrogen. Aromatase activity is the highest in breast tissue with hormone-responsive breast cancer, while it is also found in adipose tissue, muscle, bone, brain, and skin [51]. Figure 1.3 summarizes the mechanism of action of aromatase inhibitors in breast cancer tissue. In general, the third-generation AIs suppress 97% to 99% of the aromatase activity and, consequently, reduce estrogens levels around 81%-94% for Anastrazole, 88–98% for Letrozole and 52–72% for Examestane [52, 53].

Some of the most used AIs are of non-steroidal nature and are azole compounds, which are used for the treatment of estrogen receptor – responsive breast cancer [1, 54, 55]. The inhibitory mode of action of these azoles toward aromatase (being antagonist) has been well examined and established in the literature [11, 56-59]. For example, docking studies

of azole chemicals with aromatase have revealed that interactions such as aromatic (π-π), polar and non-polar are very common between the protein part of the aromatase enzyme and the inhibitory azoles [11]. Such interactions that lead to the antagonist activity of azole chemicals have been reported and show that some specific part of the molecule (i.e., a fragment) is responsible for the activity rather than the entire azole molecule [56]. While a good antagonist activity toward aromatase will be used for the medicinal purposes, the undesired effects can be observed as in the case of some pesticides such as Prochloraz and propiconazole [60] [61].

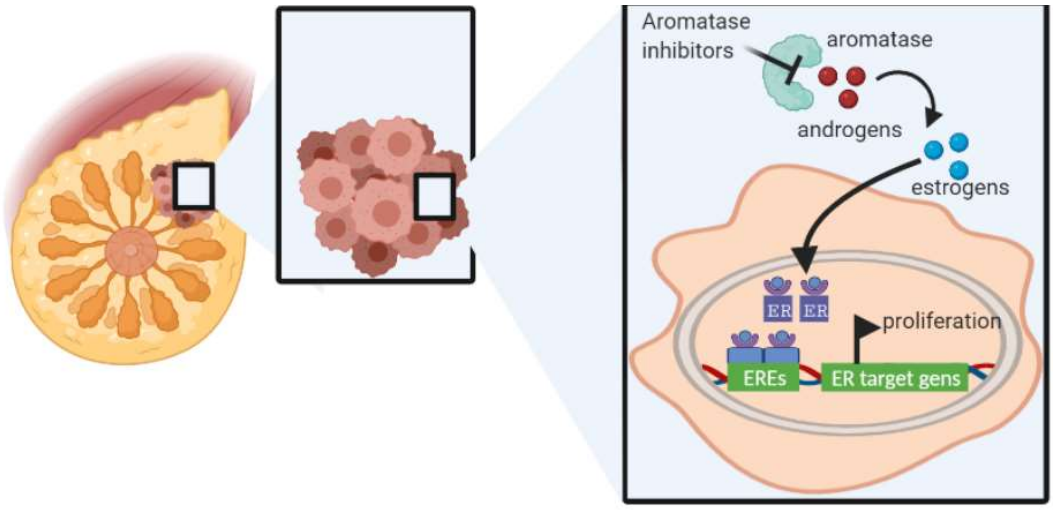


Figure 1.3: Schematic representation of the mechanism of action of aromatase inhibitors in breast cancer tissue.

1.2.2 Aromatase agonists

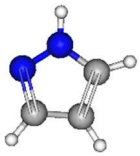
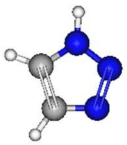
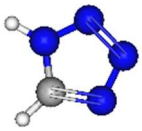
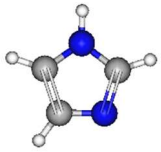
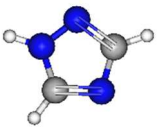
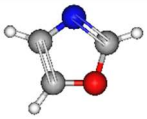
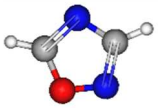
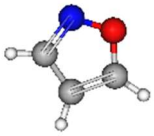
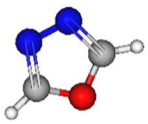
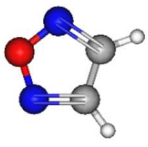
On the other hand, there are also azole-based drugs/chemicals which impart agonist activity towards aromatase CYP19A1. Though interactions of agonist azoles toward aromatase have been rarely explored, the origin of their behavior can also be expected due to involvement of structural fragments [24]. To the best of our understanding, agonist activity up to today is only considered as an undesired effect leading to hormonal unbalance, and therefore causes the endocrine disruption. Several *in vitro* studies corroborate the potential of azoles to induce endocrine disruption through aromatase activity, as for example, in the case of diverse pesticides [60, 61].

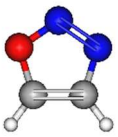
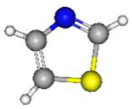
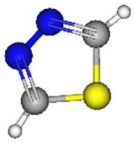
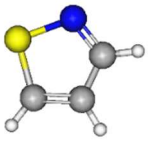
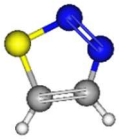
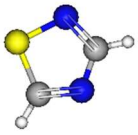
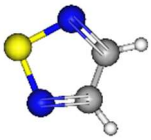
The third generation of AIs have shown to be really effective however their prolonged treatments with them, and they have been associated with side effects such as hot flashes, headache, arthralgia, mood disorders, musculoskeletal pain, cardiovascular events, sexual dysfunction, dyslipidemia and thromboembolic [48, 62]. Additionally, their use can have implications on the osseous system due to the reduction of the estrogens levels, which induce the loss of bone density and intensification of the risk of bone fractures and osteoporosis [63]. However the highest concern of AIs is related to the development of endocrine resistances, which are responsible for tumor-regrowth [64].

1.3 Azoles

Compounds containing heterocyclic fragments are an important and wide range of organic molecules, which are integrated into the common life. These fragments can be natural as well as synthetic products with applications in the areas of human health, material science, medicinal chemistry, agrochemicals, to name a few. Among the heterocyclic frameworks, azoles derivatives have shown to be a significant and interesting class of chemicals [65]. Azoles are five-membered heterocyclic compounds containing a nitrogen atom and at least one other non-carbon atom of either nitrogen, sulphur, or oxygen [65]. They can be classified as mono-, di-, tri- and tetra-azoles based on the basis of the number of nitrogen atoms (1, 2, 3 or 4 respectively) present in the ring and as thiazoles and oxazoles for rings containing N, S and N, O atoms respectively [65]. This class includes, but not only, the heterocyclic rings in Table 1.2. Table 1.2 describes the various classes of the azole rings.

Table 1.2: Compound classes of azoles grouped according to the type of substituents and the number of nitrogens.

	One nitrogen atom	Two nitrogen atoms	Three nitrogen atoms	Four nitrogen atoms
Nitrogen only		 Pyrazole	 1,2,3-Triazole	 Tetrazole
		 Imidazole	 1,2,4-Triazole	
N, O compounds	 Oxazole	 Oxadiazole		
	 Isoxazole	 1,3,4-oxadiazole		
		 Furazan		

		 1,2,3-oxadiazole		
N,S compounds	 Thiazole	 1,3,4-thiadiazole		
	 Isothiazole	 1,2,3-Thiadiazole		
		 1,2,4-thiadiazole  1,2,5-thiadiazole		

Azoles are compounds having a wide range of applications, which contribute to the development of several industrial fields as materials, energetics, catalysis, etc., constituting a versatile scaffold. Azoles have a crucial role in coordination chemistry [66, 67], structural engineering of high-mobility polymeric semiconductors [68], energetic compounds for the development of novel high-energy-density materials [69, 70], organic corrosion inhibitors [71] and components in ionic liquid chemistry [72]. Other applications, as in the case of 1,2,3-triazole-based chemosensors for the detection of metal ions, medicinal, biological, and environmental impacts seem to be also constantly growing [73].

The oxygens containing azoles have been considered due to their biological activities, possessing major agrochemical and pharmaceutical properties such as antifungals, pesticides [74-77]. Oxadiazoles, oxazoles, and isoxazoles have also been exhaustively considered for their diversified biological activities as potent antifungal agents (fungicides) since they own rewarding properties like broad spectrum of action, chemical stability, and

oral bioavailability [78-81]. Various azole derivatives have also demonstrated many other promising biological properties including antidiabetic, immunosuppressant, anti-inflammatory, antiviral, antitubercular [80, 82, 83].

Currently, one of the most attractive application of azoles is their use as aromatase inhibitors for the treatment of cancer patients. In general, the non-bonded electrons of the heteroatom(s), especially nitrogen, enable the biological applications of these compounds against the cytochrome P450enzymes. Cytochrome P450 enzymes have a heme-prosthetic group, which promotes chemical reactivity through the formation of a dative bond with non-bonded electrons of azole chemical species and hence interferes with natural biochemical reactions [84].

The problem with azoles, relating to environmental and human health concerns, is that the prolonged use of these compounds as pesticides, antibacterial and antifungals can increase their concentration in air, soil, water and living organisms. Moreover, the growing resistance of some microbes requires the discovery and development of new fungicides and antibacterial compounds [85, 86].

Additionally, azole compounds have been shown to interfere with its biological catalysis and are, as such, referred to as endocrine disruptors [43]. The specificity of azoles is compromised, therefore not only CYP51 is inhibited but also other cytochromes enzymes including aromatase, which is the case of several fungicides [87]. Endocrine disrupting chemicals can affect the organism even at low doses, particularly during organism development phases [88].

Azoles can also be classified as inactive, active agonist and active antagonist with respect to their aromatase CYP19A1 activity, where active agonists and active antagonists both contribute to endocrine disruption. New antifungal and antibacterial azoles should be inactive against human cytochrome P450 aromatase CYP19A1 enzyme, while target-specific CYP19A1 aromatase inhibitors are desired. Thus, such antagonists are required for the oestrogen receptor positive postmenopausal breast cancer patients. Modelling of the inhibitors' lone pair interactions with the iron of the heme group in the electronic and stereochemical environment of CYP19A1 is among the strategies for target-specific drug design [89].

The exhaustive effort in the field of azoles and human aromatase, covering synthesis, characterization, and *in vitro* and *in silico* studies of their numerous novel derivatives is incessantly ongoing [90-97]. Thus, the area of azoles and their derivatives, their applications and toxicity and safety assessment are of continuous high interest, and therefore this thesis will be a valuable addition to the knowledge which has been accumulated so far in this field.

1.4 Structural Alerts and Structure Activity Relationships

Structural alerts (SAs) also called toxicophores/toxic fragments are structural fragments which can cause certain adverse effects to organs when present in the molecules. A SA can be either one atom or several connected atoms [98]. The identification of such SAs can help researchers to quickly ascertain the importance of these SAs in the related toxicity of compounds and as a screening strategy [99, 100]. Additionally, the SAs can be easily interpreted and implemented, offering interesting solutions in drug design [101].

SAs are frequently used as rules defined as '*if A is B then T,*' where A is an SA, B is the value of the SA, and T is the toxicity prediction with assigned certainty level [102]. This type of rule-based models can be both, human expertise-based or computational-based [103]. Human-derived models usually are more accurate, however they are limited to the

expert knowledge and to extended literature reviews, while computational-derived models can be retrieved from large datasets in a faster manner [102, 103]. The evolution of data mining and software development tools should benefit the reduction of time and efforts needed to recognize new SAs, sometimes beyond the limits of human perception [104]. Therefore, we strongly believe that the combination of both types of modelling strategies can derive in a more complete model, which includes rules generated computationally but is assessed and modified by human experts.

During the interpretation of SAs models it is important to define the rules that drive the model, for example, if a chemical does not include SAs or does not match any toxicity rules, this does not indicate non-toxicity. Therefore, it should be important to understand how and what interpretations the model is referring to. Ideally, there should be a balance between the list of SAs and rules, their exhaustiveness, and predictive power [102, 104].

When the SAs are diverse from a structural point of view, these can be used to predict many chemicals, but this may increase false positives. On the other hand, if they are too narrow, they can be useful only for a small category of chemicals and this may increase false negatives. However, in any of these situations, especially in the second, the key aspect to be carefully defined is the applicability domain of the SAs [104]. At the end, the elucidation of SAs matches one of the most interesting approaches, the main advantage of which is the identification of compounds with the same mode of action, and that is precisely one of the main goals of investigators [98].

On the other hand, SAR are really important for toxicological researches; this method is able to link a chemical structure to a certain biological effect. The idea behind is that the structure of a chemical determines its intrinsic properties in relation to the interaction with a biological system, and thereby, establishing their biological/toxicological properties. Unlikely to the SAs, SARs not only can link a relevant chemical structure to a certain biological property, but can also understand and expose how properties are important for the effects encoded in the structure [105, 106].

In the pharmaceutical industry, SAR has long been applied, especially within drug design. Usually, many toxicological testing strategies are based on a single compound, however SARs attempt to generalize, allowing to establish a relation across toxicological data and a certain biological endpoint. Using SARs, toxicity mechanisms of action within a class of chemicals can be generalized therefore reducing the necessity of animal testing. SARs allow to develop the understanding about what could constitute a class of active or inactive chemicals, together with the possible explanation leading to a mechanism or another. Opposite to other methods as SA, which are more focused on classification, SAR also address the reason of this classification [106].

Working with SAR goes from the identification of whether a SAR actually exists in a set of structures and their related activities, to the elucidation of the aspects of one or more SARs, to later use this information to perform structural modification for activity optimization. Interestingly, many times the optimization process can involve more than one property, as for example most lead optimizations try to improve potency and reduce toxicity [106] [107-110].

SAR models may also lead to mechanistic hypotheses to guide future testing and validation, these kinds of models are not only a predictive tool, but can also be useful for lead optimization [105, 106]. Curiously, SAR and read-across have been purposed as the alternative most widely used methods for filling data gaps of substances lacking toxicological data [105].

1.5 Read-Across

1.5.1 Definition and relevance

The reduction of *in vivo* tests to be conducted has become essential in terms of resources and animal benefits. Grouping/category chemicals approaches can reduce the number of chemicals to be tested because the available information about the endpoints can be used to estimate those properties for untested substances. During the last decades read-across methodologies earned attention as important for risk assessment tools for data gap filling [111].

Read-across assesses a toxic endpoint of untested substance (target chemical) based on the results for the same endpoint for a tested substance (source chemical). In the category approach where comparisons are made, the similarity concept becomes fundamental since analogues and targets need to be "similar" in the context of structure, properties and/or activities [112].

According to the second report under Article 117(3) of the REACH Regulation by the European Chemicals Agency (on 2014), 75% of all REACH registration dossiers included read-across or category formation methodologies to fill information requirements for higher-tier toxicological studies, the majority of them applied to repeat dose toxicity, which is one of the most challenging assessments in the process of animal test replacement [112].

Significant success has been achieved in areas of computational toxicology, and therefore on the read-across techniques, but many breaches should still be filled. Aspects as endpoints, chemical space and methodologies need to be explored in a deeper way. In this sense, the EU policy and the elimination of animal models to evaluate systemic toxicity for cosmetics ingredients in 2009 led to numerous collaborative initiatives regarding in silico modelling and read-across [112, 113]. Actually, the main driver for the expansion of read-across are the legislations [113, 114].

Numerous reasons have led to growing the category formation and read-across applications. The fact that numerous chemicals have missing relevant toxicological data, which are often crucial for risk assessment, the heavy impact of legislations in non-test-methods, together with the acceptance of read-across for regulatory purposes, have directly influenced the growing. In addition, the development of tools as OECD QSAR Toolbox, and other computational methods to efficiently access the information, have facilitated the process. Read-across is a clear, simple, transparent and easily interpretable technique to *in silico* evaluate the properties of a substance, even for more complex endpoints as repeat dose toxicity and reproductive effects in humans, based on less amount of data, but with more quality and robustness [114].

Regarding read-across, the advantages seem to be higher than the disadvantages [113, 114]. Read-across approaches reduce the need to test every endpoint for every chemical, additionally, the assessment of a large number of chemicals as a category can be more efficient and accurate than the assessment of a single compound. To justify a read-across, the structural, physical-chemical and biological similarities are often used as basis. The type of similarities to be considered to read-across are not defined, which are endpoint and target dependent, however a list of more relevant similarities to be considered to form categories was proposed by Cronin et al. [112].

1.5.2 The Process of category formation and read-across

The steps to perform a read-across are not exactly defined, however several relevant contributions have been made. For example, seven key steps considering a discrete organic chemical were proposed by Grace et al. [115], including 1) Decision context, 2) Data gap analysis, 3) Overarching similarity rationale, 4) Analogue identification, 5) Analogue evaluation, 6) Data gap filling and 7) Uncertainty assessment. Another flow based in six steps has also been proposed as in the case of Escher et al. [111], including 1) Problem formulation, 2) Characterization of the target compound, 3) Identification of source compounds, 4) Evaluation of source compounds, 5) Data gap filling, and 6) Uncertainty assessment. In our opinion, the important highlight here is that the implementation of this procedure needs expertise, also due to the subjectivity associated to the approach. At the outset, the central axis is that the steps to develop a category or analogue approach should be defined on the basis of RAAF [116].

Summarizing the process, based on the different sources it is easy to distinguish various basic steps. The introductory step within all read-across approaches could be called in several manners, decision context, problem formulation, scenario definition, but in any case, the idea is the same, to define the prioritization, screening, level of hazard or risk assessment of the scenario of the read-across prediction. The scope and decision context will determine the level of uncertainty that can be tolerated [115] [111]. Then, the chemical **target must be identified as well as the effect and/or endpoint to predict**. Tendentially, it is assumed that the target will only be one chemical, even when mixtures could also be considered. This step assumes that the target has a well-defined and known structure, and that the pertinent available experimental or predicted data has been collected, from the structural and PC properties on the base to the toxicological *in vivo* studies. Usually, in this step, the molecule is considered as neutral and not a salt. As a consequence of the target identification, the endpoint to be predicted will be identified on the basis of the data availability [111, 114].

Third, the analogues or source compounds identification is essential, this step is basically to **search analogues**. These chemicals must be similar to the target; therefore, the similarity search strategy must be clearly defined case by case. For the analogues, the same data type as for the target are stored into the data matrix. The "most appropriate" analogues are ruled by the nature of the chemical, available data for read-across and the endpoint to be modelled. An important aspect to remark is that at least some of the selected analogues must have available *in vivo* data for the endpoint to be read-across [111, 114].

After the search of sources, it is necessary to evaluate their validity and pertinence. This **analogue evaluation** will become critical if the initial search was an unsupervised process [115]. Most suitable candidates should have enough data, especially for the endpoint of interest. The evaluation should consider similarities with focus on general physicochemical characteristics, metabolic profile, and reactivity. The data quality and availability are again decisive at this point [111, 114, 115]. Once this step is performed, a category is already defined, and it should be evaluated in itself for consistency, because maybe sub-categorizations could be necessary [114].

The borders between the previously discussed steps are narrow, and the links driving the process are not straight, however each step is somehow dependent and correlated to each other. For example, the identification and evaluation of source compounds are interconnected and can become an iterative process, which at the same time leads to an enrichment of the initial hypothesis.

Once a solid category has been obtained, the read-across can be performed, this step is called **data gap filling**, as it could be so easy as a simple interpolation, however in some cases other actions could be required, e.g., development of a QSAR relating activities to the properties of the molecule. Three main strategies can follow at this point, 1) a conservative strategy which will be based on a worst case, and consider the target chemical to be as toxic as the more toxic source compound; 2) a trend analysis, when feasible, a regression analysis could be performed; 3) a nearest neighborhood approach, meaning that one SC is considered as the most similar to the TC, and only this SC's endpoint data will be read-across to the TC. For qualitative read-across strategy could be as simple as a majority vote approach [111, 112, 114, 115].

The last step should include the uncertainty evaluation, just now excellent guides exist on uncertainty assessment, e.g. weight of evidence [117, 118]. The RAAF also proposes a complementary strategy to address the different sources of uncertainties [116]. Finally, when a read-across is seeking regulatory acceptance, all processes related to the prediction should be properly documented.

1.5.3 The application of computational workflows in read-across

Currently several tools have been developed to perform read-across, however other combinations of resources can be equally effective offering a more flexible and less prescriptive approach. The computational workflows are an interesting application that allows the information extracted from several sources to be integrated into a single tool. Workflows are easily adaptable and controlled by the users, additionally they can be shared and modified, allowing the development of new characteristics. In addition, the features can be included or excluded according to the data requirements. KNIME is a free and open-source data analytics, reporting and integration platform. This platform includes over 1000 nodes that can be integrated to automatize data analysis and for machine learning. For the purpose of read-across KNIME offers an interesting solution for workflow implementation, where the user can incorporate its own chemotypes and databases for the prediction of the toxicity [119].

1.6 Research Aims of This Thesis

Historically, the hazard properties of chemicals have been determined using *in vivo* toxicological tests, which indeed have been the most comparable to human toxicity. However alternative methods, especially *in silico* methods, are also a potentially acceptable and viable alternative for regulatory purposes. *In silico* approaches include a broad range of testing strategies to predict several kinds of toxicity endpoints.

The formation of chemical categories for read-across can be performed through many methods, where the mechanism of action-based method is one of the most attractive ones. The formation of categories based on the knowledge of mechanism of action may rely on the definition of structural alerts. In this sense, structural alerts are used to cover the similarity requirements for category formation. Once a certain relevant structural alert has been identified within a target chemical, it can be used to identify analogues within a database, therefore chemicals that contain the same structural alert present in the target can be grouped into a mechanism-based category [120].

Interestingly, in the opposite way, through the assessing of chemical categories, structural alerts can also be elucidated, this means that chemical categories can be obtained

based on other similarity criterions and then, from those categories structural alerts can be extracted as proposed by Hewitt et al. [121]. The assessing of those chemical categories in many cases may allow a mechanistic rationale to explain the observed toxicity [121, 122]. In fact, the application of structural alerts is used as a method of mitigating the risks of idiosyncratic drug toxicity [122].

Read-across can be referred to also as a simplistic form of Q(SAR) analysis, when enough data is existing, a simple Q(SAR) model can be obtained. A suitable example of the relationship between category formation and SARs is its application for AOPs [123]. As with any SAR, following the hypothesis that structurally similar compounds should act with the same mechanism of action, the mechanistic rationale of each category can be explored eventually, and common toxicological profiles are expected [121].

Additionally, during the analysis evaluation step in a read-across, preliminary similarity considerations can be set up by reference to existing (Q)SAR tools [113, 124, 125]. For example, QSAR tools can support the PC properties predictions and evaluation as $LogK_{ow}$, molecular weight and vapor pressure, which contribute valuable information regarding bioavailability, or for instance protein binding alerts can mimic certain toxicological/biological features [115].

We can conclude that computational methods to predict chemical toxicity can solve many predictions, however each method has its strengths, limitations, scope of application, and specificity of interpretation. Therefore**,** the overall aim of the thesis was to explore the biological actions between azoles and the human aromatase CYP19A1 enzyme applying different *in silico* approaches (see

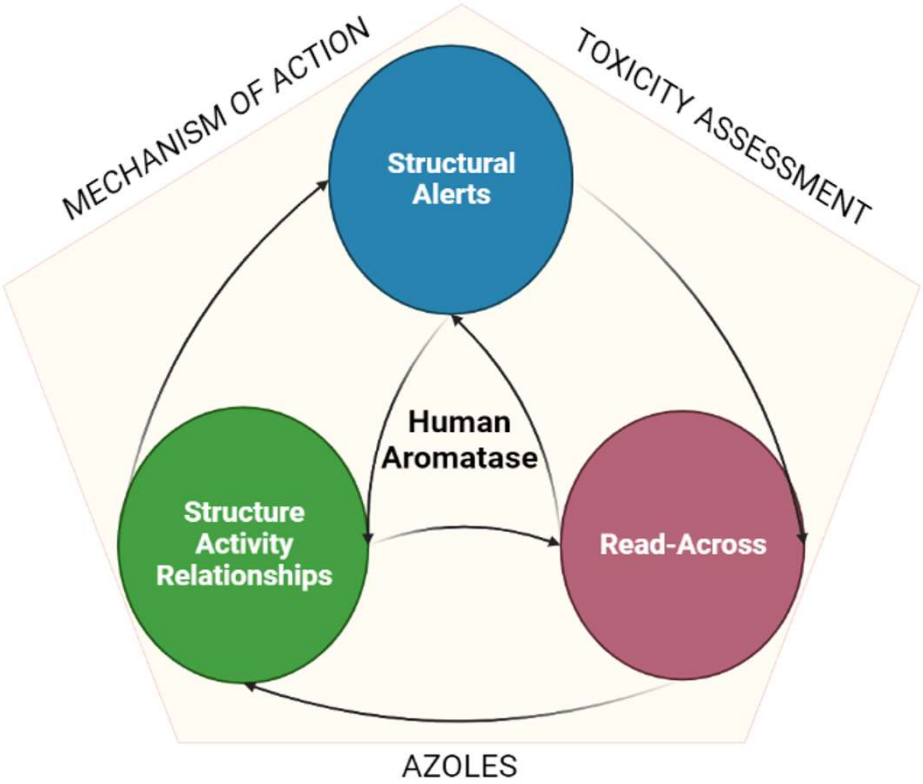


Figure 1.4), which allows to predict the biological and toxicological profile of new azoles; and provides a deeper mechanistic understanding and rationality in respect to the currently available findings with possible applications in drug discovery and risk assessment.

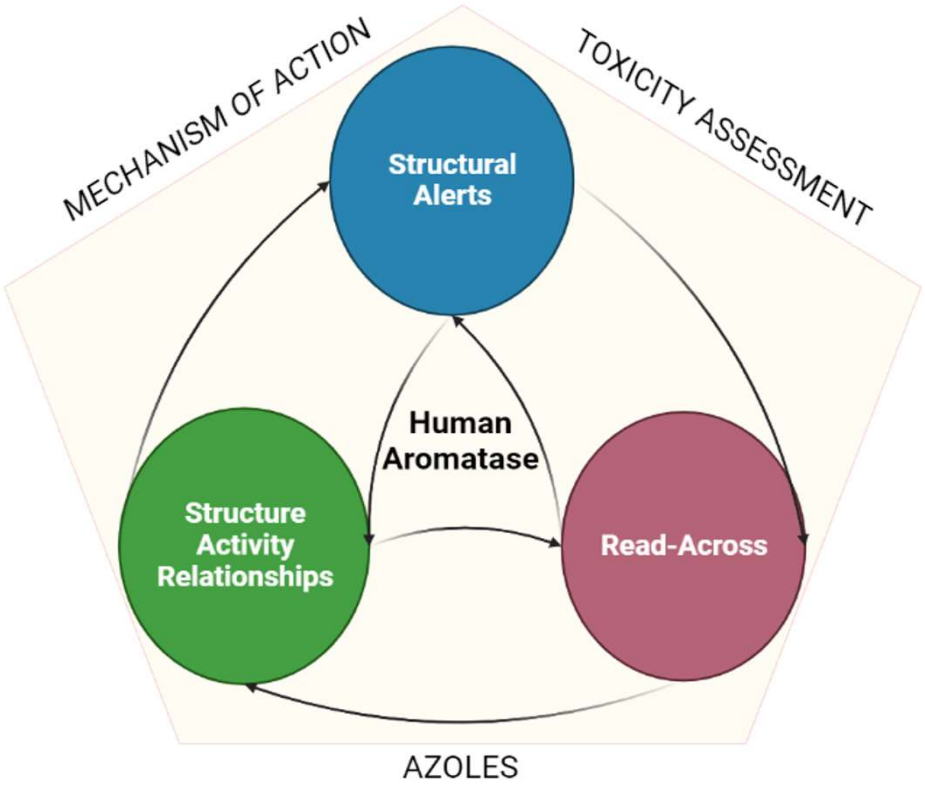


Figure 1.4: Graphical representation of the main cores addresses in this thesis.

The following is a more detailed description of specific research questions and the rationale behind them. The specific aims of the thesis were:

- Developing new structural alerts for aromatase binding leading to endocrine disruption based on different azoles classes.
- Establishing new structures-activity relationships between azoles and human aromatase.
- Identifying the principal chemicals substituents influencing the azole-CYP19A1 interaction.
- Explaining the contribution of different substituents within azoles to the chemical activity on aromatase.
- Proposing insights into the mechanism of action.
- Generating new chemical categories for a toxicological read-across to support data gap filling on aromatase.
- Developing a workflow for analogue identification for read-across on human aromatase.
- Validating the workflow and case studies.

Chapter 2

Theoretical Background of Methods

In the present work, structural alerts (SAs) (Chapter 3), structure-activity relationships (SARs) (Chapter 4) and read-across models (RAX) (Chapter 5) methods were applied to predict the human aromatase CYP19A1 activity/toxicity and to explore the toxic mechanism of action for azole chemicals. The methodology used to develop each procedure appears described within the respective chapter, however, in this chapter, the general theoretical background behind these methods has been outlined. The chapter begins by presenting the dataset compilation and cleaning process. It has been introduced how the chemicals structures were codified, and by what methods the chemical space was studied, especially with the focus on molecular descriptors. The structural alerts, SARs, and read-across methods used to explore aromatase mechanism of action, have been discussed as approaches towards solving the prediction of chemical toxicity. Finally, the roots to validation methods are discussed. The information within this chapter has been taken from several literature sources [126-134].

2.1 Dataset Compilation

The starting dataset was collected from the Tox21 library considering only Tox21_Aromatase_Inhibition (activity test). This contained 20,992 compounds encoded as SMILES, name, and CAS number [135]. The assay was performed using aromatase breast cancer cell line (MCF-7 aro) (cell-based assay) and the concentrations of testosterone (androgen and estradiol (an oestrogen)) were measured before and after exposure to azole test compound. The qualitative outcome was recorded as active agonist, active antagonist, and inactive, where quantitative agonist and antagonist activities were expressed in nanomolar units represented by AC50 in the original database [135].

2.2 Data Cleaning

The curation procedure of the data involves the retrieval of SMILE following the workflow developed by Gadaleta et al. [126]. The maximum purity is labelled “A” and only compounds with this label should be considered. The detection of inorganic compounds, organometallic compounds, mixtures, neutralization of salts, tautomeric forms, and chemotype normalization is performed using the KNIME platform [43]. The compounds with inconclusive assay outcomes are discarded and duplicate structures are classified into two cases as follows: (i) activity range lower or equal to 1:3, and (ii) activity range higher

than 1:3. In the first case, the mean of the activity is calculated, and in the second case, the structures are rejected.

2.3 Chemical Structure Codification

The ligand-based methods are useful in medicinal chemistry, these methods relate 2D structures of molecules through relatively simply and speeded approaches. For modelling, the molecules can be codified as SMILES and fingerprints [136].

SMILES are ASCII strings obtained by printing the symbol nodes (chemical elements) encountered in a depth-first tree visit of the chemical graph and fingerprints are one of the most popular techniques for structural codification, the focus of which is to support chemical similarity searching. The representations encoding the 2D structures of molecules or its features through bit representations are referred to as molecular 2D fingerprints [127, 136-139].

To analyze the chemical substructures within the database, the SMILES notations were encoded as PubChem fingerprints. The PubChem binary substructures fingerprints can be used for chemical codification and similarity searching, where a substructure is a fragment of a chemical structure, and a fingerprint is an ordered list of binary (1/0) bits. For a molecular fingerprint, each bit represents a Boolean determination for the presence or absence of certain structural properties, for example, an element count, a type of ring system, atom pairing, atom environment (nearest neighbors), etc. Figure 2.1 represents the hypothetical representation of a 10/bit fingerprint with three bits set because the substructures they represent are present in the molecule. The native format of the PubChem substructure fingerprint property is binary data with a four-byte integer prefix, where this integer prefix indicates the length of the bit list. Fingerprint computations were based on CDKit toolkit [140].

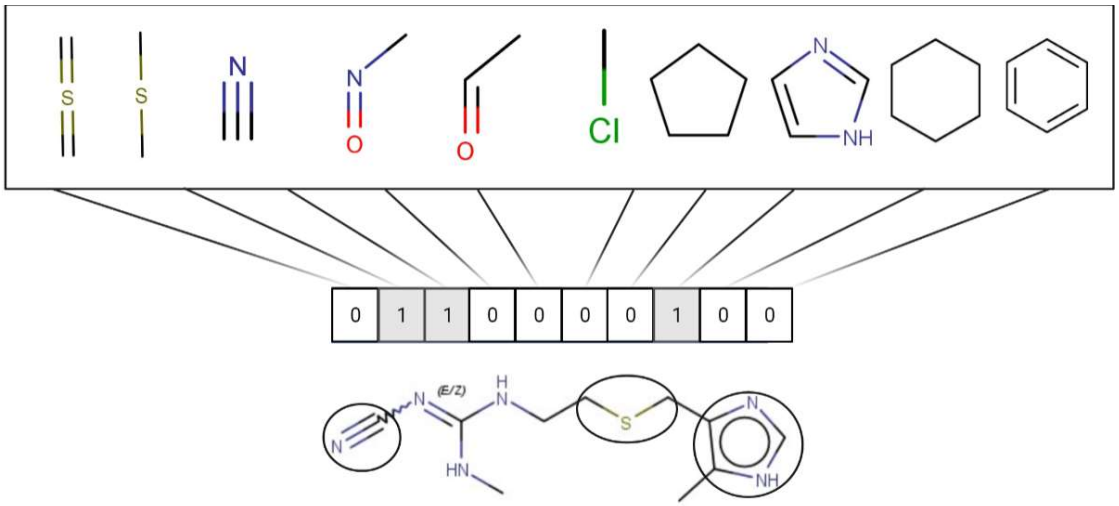


Figure 2.1: Representation of a hypothetical 10-bit substructure fingerprint with three bits set because the substructures they represent are present in the molecule (circled).

2.4 Molecular Descriptors

The chemical space can be explored in terms of molecular properties/molecular descriptors. The molecular descriptor is the final result of a logic and mathematical procedure which

transforms chemical information encoded within a symbolic representation of a molecule into a useful number and/or a result of some standardized experiment [141].

The evaluation of molecular descriptors encodes various information about the chemical nature of molecules, and its characteristics. Features such as LogP, Topological Polar Surface Area, Hydrogen bond acceptors and donors, and the Molecular Weight have been demonstrated to contribute with mechanistic knowledge regarding the pharmacokinetic process, particularly intestinal absorption [142]. At the same time, absorption is suggested as the most important principle in pharmacokinetics theory, therefore the determination of certain descriptors, as previously mentioned, is essential to characterize chemical activity/toxicity [143].

The **polar surface area** is defined as the sum of surfaces of polar atoms in a molecule. This descriptor has shown to establish good correlations between the passive molecular transport in cellular membranes and the prediction of transport features of drugs, however its determination is time-consuming. It is a descriptor that was shown to correlate well with passive molecular transport through membranes and, therefore, allows prediction of transport properties of drugs. [142, 144]. Simultaneously, the TPSA is an appropriate measure of the polar surface area that evades the necessity to calculate ligand 3D structure or the analysis of relevant biological conformations [144]. The **TPSA** can be used for the calculation of PSA, but in this case, the calculations are simply based on the summation of tabulated surface contributions of polar fragments (i.e. atoms regarding also their bonding pattern, such as oxygen, nitrogen and their attached hydrogens), measured in ($Å^2$) for the tendency for polar interactions. This method, termed topological PSA (TPSA), provides results which have showed to be practically identical with the 3D PSA, while the computation speed is 2−3 orders of magnitude faster [142] [144]. Additionally, a positive correlation with inhibitory activity data for Aromatase have been described [142] [144].

Furthermore, the LogP, the partition coefficient describes the ratio of concentrations of a compound in the two phases of a mixture of two immiscible solvents at equilibrium. Usually, these solvents are water, while the other is a hydrophobic substance, such as octanol. In this way, the partition coefficient gives an idea about the "water liking" and "lipid liking" properties of a substance, or hydrophilic and lipophilic properties respectively. The partition coefficient is an important descriptor of the physical nature of a substance, and therefore a predictor of its performance in diverse environments. For example, hydrophobic substances with high partition coefficients are preferentially distributed to hydrophobic compartments such as lipid bilayers of cells while hydrophilic chemicals (low partition coefficients) favorably are found in hydrophilic compartments such as blood and the cytoplasm [145]. The LogP value is a constant defined as shown in Eq. (2.1) and Eq. (2.2):

$$LogP = log_{10}^{(Partition\ coefficient)} \tag{2.1}$$

$$Partition\ coefficient = \frac{[organic]}{[aqueous]} \tag{2.2}$$

* [ ] indicates the concentration of solute in the organic and aqueous partition.

A negative value for logP means the compound has a higher affinity for the aqueous phase (it is more hydrophilic); when logP = 0 the compound is equally partitioned between the lipid and aqueous phases; a positive value for logP denotes a higher concentration in the lipid phase. LogP = 1 means there is a 10:1 partitioning in organic: aqueous phases.

From a biological point of view, it is recognized that IC_{50} is well correlated with logP, as for example in the case of aromatase inhibitors, where it has been deduced that high logP values are essential for good therapeutic activity, suggesting that effective aromatase inhibitors should be relatively hydrophobic [11]. Therefore, this property has shown to have implications on the binding mode of aromatase inhibitors [8, 146, 147].

The **MW** is related to the size of the molecule. As molecular size increases, solubility decreases, it also impedes passive diffusion through the bilayer membrane. The minimization of ligand molecular weight has special attention for many molecules [148].

The **hydrogen bond** represents non-covalent interactions present in many molecules. Hydrogen bonds are formed between a hydrogen atom bound to a small, highly electronegative atom and another small, highly electronegative atom with an unshared electron pair. The elements involved in hydrogen bonding are commonly nitrogen, oxygen, and fluorine. The hydrogen bonding is weaker than a covalent interaction, therefore the ligands binding enzymes through this kind of bonding tendentially are reversible [42, 149, 150].

A hydrogen atom attached to a relatively electronegative atom is a **hydrogen bond donor**. The electronegative atom attracts the electron cloud from around the hydrogen nucleus and, by decentralizing the electron cloud, leaves the hydrogen atom with a positive partial charge. Because of the small size of hydrogen relative to other atoms and molecules, the resulting charge, though only partial, is stronger. A hydrogen bond results when this strong partial positive charge attracts a lone pair of electrons on another atom, which becomes the **hydrogen bond acceptor**. An electronegative atom such as fluorine, oxygen, or nitrogen is a hydrogen bond acceptor, regardless of whether it is bonded to a hydrogen atom or not [149, 150]. Figure 2.2 represents the hydrogen bonding formation.

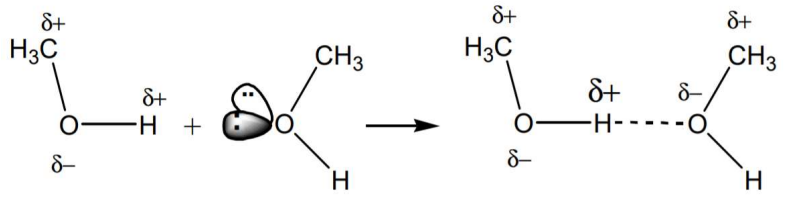


Figure 2.2: Representation of hydrogen bonding formation.

According to the Lipinski rule of five, all these descriptors have an important contribution to evaluate drug likeness, and possibility of a substance to enter the circulation when introduced into the body. It is suggested that a molecular mass less than 500 Daltons is desirable for an oral activity, as well as not more than five hydrogen bond donors, not more than ten hydrogen bond acceptors and a logP value lower than 5 [148, 151].

Other descriptors, as for example, quantum-chemical descriptors, such as HOMO, LUMO, electronegativity, electrophilicity index, harness, polarizability, dipole moment and other thermodynamic descriptors can incorporate important electronic information about molecules and could be exploited to deduce relevant mechanistic information. In addition, autocorrelations, walk index and edge matrix etc. can be helpful to deduce the models for biological targets, with important two-dimensional information. The constitutional descriptors having 1-D information such as number of aromatic rings, hetero atoms, type of bonds etc. are easily computable descriptors, which be useful for the QSARs/SARs [24].

2.5 Similarity Measure

Similarity is employed in a wide variety of areas [152-155], and therefore, quantifying the similarity of two or more chemical structures is a key concept and common task in chemoinformatic [156]. However, what exactly is meant by similarity between compounds is intrinsically subjective and consequently hard to state precisely [157].

Several methods have been proposed to quantify the similarity, but probably the most extended one is the method based on counting the common sub-structural characteristics [158]. From these methods, Tanimoto coefficient has emerged as a standard similarity measure [159].

Tanimoto coefficient is identified for several studies as one of the best similarity metrics for fragment-based similarity searching [128], and is intended as the similarity measure between two points a and b, with k dimensions. Eq. (2.3) illustrates its mathematical definition. The Tanimoto similarity is only applicable to binary variables, and coefficient ranges are from 0 to 1, where 1 represents the highest structural similarity.

$$\frac{\sum_{j=1}^{k} a_j \times b_j}{\sum_{j=1}^{k} a_j^2 + \sum_{j=1}^{k} b_j^2 - \sum_{j=1}^{k} a_j \times b_j} \tag{2.3}$$

2.6 Structural Alert Development

The structural alerts developed in this thesis were extracted using SARpy software. SARpy is a knowledge extractor software, which based on a training dataset generates the substructures in the given chemical set and mines associations between the occurrence of molecular fragments and the activity of those molecules which contain occurring fragments. To automatically construct structural alerts SARpy performs three steps, for the set of molecules encoded as SMILES string [129]. Briefly, the model training process starts from the training dataset in the form of a matrix of chemical structures in SMILES format and their experimental activity label.

The first step consists of a recursive straightforward **fragmentation algorithm**. The algorithm explores every combination of bond breaking SMILES string and allowing the determination of every substructure of the molecular set. The initially generated fragments are “rough”, however the iterative manner of fragmentation leads to more complex and refined substructures, where each substructure is the result of several iterations on the precedent ones. For example, on a general A–B–C structure, the first fragmentation step will raise the two fragments A and B–C by breaking the first bond, then the A–B and C fragments by breaking the second bond; the B fragment will be found in the next step. It is plausible to accumulate substructures until the in-depth fragmentation of the original structures is accomplished [129] as shown in Figure 2.3 which constitutes a real case.

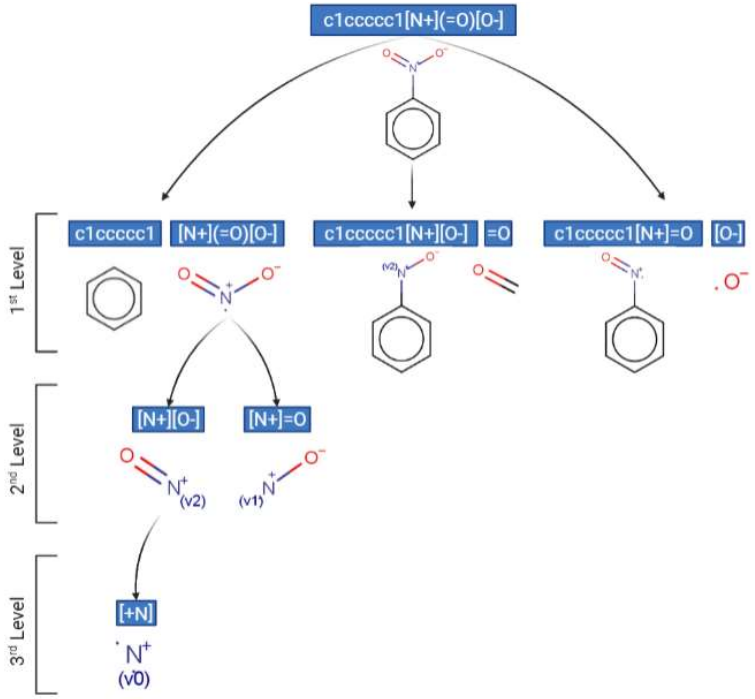


Figure 2.3: SMILES fragmentation of nitrobenzene (on top).

*2D structures were added to facilitate the visualization. Duplicates are omitted [129]

In detail, after the first iteration of fragmentation, which starts with arbitrarily non-aromatic end, the next iteration acts over every bond in the input structures and reunites the two fragment products from the hypothetical bond breakdown. Each step considers all the conceivable couples of fragments obtained from each input structure. At the end, the substructures predicted are collected and taken individually as the base for the next fragmentation step, then all the possibilities of a second bond breakage are explored, and so on, until no more new fragments can be extracted. Two important issues to outline regarding the algorithm for fragmentation are that SARpy considers rings as single entities during the fragmentation process, which means that ring bonds are not broken, and that the recommended maximum length of relevant fragments in terms of number of atoms is set as 18 [129].

The following step represents the **evaluation** phase. The evaluation consists of the validation of each substructure as a potential structural alert using the training set. This approach fully matches the substructures against the training set to evaluate the predictive capacity of each individual moiety. In this step, having the experimental labels, the binary structural alerts (positive/negative) are used to classify molecules, as for example active/inactive chemicals, respectively. With this method, the fragment matches can be split either against positive and negative structures, generating the so-called true positives, false positives, true negatives, and false negatives. Using these parameters, several indicators of the performance can be calculated.

SARpy places special attention on the calculation of the likelihood ratio, which is a measure of precision intrinsic to the test; probably the most important contribution of this parameter is that it does not depend on the occurrence of activity labels in the training set; the definition of the LR is presented in the section *Validation Methods*. The evaluation phase can identify the substructures with the best generalization capability, and with good predictivity. This step constitutes the base for the final phase of the process, which is the rule set extraction.

The **rule set extraction** step intends to obtain a reduced number of rules which should exhibit the best precision. This is an iterative procedure which initiates ordering the list of potential structural alerts using the LR, and keeping the best ranked one. The top ranked fragment will be simultaneously assigned to the final rule set and eliminated from the precedent one. The TPs and FNs matching the alert recently chosen will be removed, and consequently automatically updated through the remaining potential alerts. The likelihood ratios of potential alerts will also be updated. Instantaneously, the process is iterated from the first step. This approach minimizes interferences and maximizes efficiency. The conclusion of the rule set extraction will be aided by two diverse approaches, one that focuses on the sensitivity and another which targets the specificity. Finally the output is presented as an ordered set of rules in the form: 'IF contains THEN' [129].

The elucidation of structural alerts is a useful method in many fields to identify potential toxic compounds, and in general, to explore chemical toxicity. However, to quickly retrieve accurate structural alerts is a complex task. SARpy is a software which has been widely used for the development of structural alerts [160-163], having a good predictive performance when comparing with another software as MoSS, Bioalerts and Fingerprints [164].

2.7 R-Group and SAR Analysis

In a series of related structures, scaffold fragments can be identified, this central structure can be founded by R-group decomposition, which is a special case of substructure search. The R-group analysis not only helps in identifying scaffolds, but also recognizes ligands at certain attachment positions. The query molecule entails the scaffold and the ligand attachment points are depicted by R-groups, which are simple atoms or a combination of them. The elucidation of those R-groups facilitates the discerning of favorable substitution patterns and preferred substituents to aid in lead optimization, through SAR analysis.

The R-group analysis was performed using a workflow developed using KNIME platform, which had as central core the RDKit R-Group Decomposition node. Some of the structural alerts elucidated were taken as the scaffold for the analysis, while the R-groups were explored by a combination of automatic techniques and the expert judgment. A rigorous procedure of SAR analysis was carried out to construct detailed relationships between the substituents (R-groups) and their possible effect on the activity. Simultaneously, the scaffolds and their biological relevance were examined.

2.8 Model Development

Structural Alert Model is a knowledge-based predictive toxicology model. The approach relies on combining the extracted structural alerts into rules using SARpy [165]. While the specific methodology for model development is described within each chapter, some basic principles can be outlined.

The model building starts with the alert development procedure described in Section 2.6, complemented with the construction of rules, to allow the classification of chemicals into two different classes. Finally, a prediction is provided by the model considering diverse aspects, as for example the type of activity associated to the identified fragment and its LR value. If there is no fragment identified, the output of the model is "not_predicted"[165].

When structural alerts are implemented as *in silico* models for toxicity evaluation, it is vital to contemplate which alerts provide good predictions, and therefore to define the reliability associated for each individual alert. In this thesis, the overall structural alerts

were classified according to four reliability levels, which are assigned according to the accuracy ranges, low (Acc≤0.6), low to medium (0.6<Acc≤0.7), medium (0.7<Acc≤0.8) and high (0.9<Acc≤1).

2.9 RAX Workflow Development and Implementation

To select analogues for RAX an automated workflow was developed and implemented using KNIME platform. KNIME allows modelling workflows through nodes that process data which are transported by connectors between the nodes [130]. A standard flow commonly initiates with a reading node from some source, and then the data is stored in an internal table-based format that can be described as strings, integer, float, molecules, and many others. The following nodes modify and transform such data. The next step involves the core nodes of the algorithms, e.g., the model development nodes, followed by the visualization nodes [130].

The theoretical conception of the RAX approach developed in this thesis was inspired by a methodology proposed by Gadaleta et al. [131] and implemented using KNIME. The workflow was developed on the base of a combination of different types of similarity approaches that compute the similarity between a target and source(s) to identify potential analogues. The implemented procedure offers two main applications: i) the analogue identification and ii) the assistance during analogue evaluation steps of RAX. In both cases the principle is the same; the calculation of chemical similarities through a search performed in a dataset. The individual lists of candidates to analogues are retrieved from the dataset considering different kinds of similarities approaches, as for example, structural similarity, physical chemical similarity, a common metabolic behavior, or structural alerts. In the next step, the chemicals in the source database are ranked considering each kind of similarity, and each top ranked compound is returned as a first output. A second output is provided including only the intersection between all top ranked compounds. This intersection is then considered as the list of most suitable analogue(s), and their activity is used to predict the activity of the target chemical [134].

As a KNIME flow, one of the main advantages is that it can be easily exported and executed, and it has a simple extensibility, which means that later it can become a plugin or extension. Furthermore, considering that each individual node processes the whole input table before the output is promoted to following nodes, the stored results are constantly kept allowing to stop the execution at any point without losing information and resume it later on. Additionally, the results can be checked at any time, and new nodes can be inserted without the necessity to re-execute the precedent ones [130].

2.10 Validation Methods

To assess and compare the performance of different models and structural alerts, a consistent selection of performance statistics has been chosen and used throughout this work. These are outlined here. Any prediction can be split into four categories, as shown in the confusion matrix:

	Predicted Positive	**Predicted Negative**
Experimental Positive	"True Positive"	"False Negative"
Experimental Negative	"False Positive"	"True Negative"

Based on this confusion matrix, several statistical parameters, such as Accuracy, Sensitivity, Specificity, and others, can be calculated. The predictability and reliability can be described by the statistical parameter Acc, which was calculated as shown in Eq. (2.5). The accuracy can take values in a range of 0–1, while the values close to one were desired, and were interpreted as a better performance during a classification [133, 166].

$$Accuracy = \frac{True\ positives + True\ negatives}{Positives + Negatives} \tag{2.5}$$

TP and TN represent the number of accurate predictions irrespective of whether the predictions were active/inactive or agonist/antagonist. The sum of positives and negatives represents the total number of the predictions made.

Other statistical parameters, namely likelihood ratio, sensitivity and specificity as shown in Eqs. (2.6), (2.7) and (2.8) respectively, were also considered to assess the performance. The LR value provided a measure of accurate predictions considering the distribution between classes of compounds, e.g., active and inactive compounds, or among active agonist and active antagonist compounds and the ratio of true and wrong predictions. The ideal value of this parameter is "inf", which means that the number of wrong predictions was zero, and accordingly, the division by zero (the denominator of Eq. (2.6)) tends to infinity implying no incorrect predictions, or what is the same, a perfect classifier. Hence, the higher the LR value, the greater will be the contribution towards a single activity class. However, unlike the accuracy where the numerical range is well defined, the wide numerical range of LR values could be difficult for interpretation.

Parameters such as sensitivity Eq. (2.7) and specificity Eq. 2.8 were applicable to the classification model and RAX to calculate the measure of the proportion of true positives and true negatives with respect to the total number of positives or negatives, respectively [133]. To recognize the number of accurate predictions and total predictions, the chemicals were classified by applying the substructure filter using RDKit node of KNIME.

$$LR = \frac{True\ prediction}{Wrong\ prediction} \times \frac{Positives}{Negatives} \tag{2.6}$$

$$Sensitivity = \frac{True\ positives}{True\ positives + False\ negatives} \tag{2.7}$$

$$Specificity = \frac{True\ negatives}{True\ negatives + False\ positives} \tag{2.8}$$

2.11 Software Packages

A variety of software packages were used for the development of the analysis, see Table 2.1. All the tools proposed are usually and routinely applied by the industry and researchers. All softwares were partially/totally publicly available. The variety of software tools can be distributed into three main groups i) data analysis, ii) modelling, validation, implementation and iii) visualization. Most software packages were executed from the KNIME platform, with some exceptions as in the case of SARpy.

Table 2.1: List of software used for modelling.

Tool name	Functionality	Available from
SARpy [129]	Model generation and validation	https://www.vegahub.eu/portfolio-item/sarpy/
RDKit	Molecule codification and SAR analysis	https://www.rdkit.org/
Chemaxon (Marvin)	drawing, displaying and characterizing chemical structures	https://chemaxon.com/products/marvin/download
CDK	Descriptors calculation and Similarity evaluation	https://cdk.github.io/
Statistics	Statistical analysis	https://www.ibm.com/products/spss-statistics
JavaScript	Data visualization	https://www.javascript.com/
Erlwood nodes	Similarity evaluation and Data visualization	https://www.knime.com/community/erlwood
Microsoft Excel	Statistical analysis	https://www.microsoft.com/en-us/microsoft-365/excel
SyGMA package	Metabolite and Metabolic Pathway predictions	https://github.com/3D-e-Chem/sygma
Biorender	Figure development	https://biorender.com/
R-package [167]	Data analysis	https://r-pkgs.org/
KNIME [130]	Workflow implementation	https://www.knime.com/knime-analytics-platform

Chapter 3

Development of Structural Alerts for Aromatase Binding

3.1 Structural Alerts for Active and Inactive Chemicals

3.1.1 Introduction

The CYP19A1 enzyme mediates the hormonal balance and aromatization of androgens to form estrogens during steroidogenesis [34]. Any disturbance in the aromatase activity induces imbalance of estrogen levels in tissues which further disrupts estrogen-mediated physiological responses, leading to endocrine disruption [8, 168].

One of the most commonly used NSAIs are azole compounds which are also used for the treatment of estrogen receptor (ER)-responsive breast cancer [1, 54, 55]. The inhibitory mode of action of these azoles toward aromatase (being an antagonist) has been well examined and established in the literature [11, 56, 169]. On the other hand, there are also azole-based drugs/chemicals which impart agonist activity towards aromatase CYP19A1 [24]. Though interactions of agonist azoles toward aromatase have been rarely explored, the origin of their behavior can also be expected due to the involvement of structural fragments.

The identification of such SAs can help researchers to quickly ascertain the importance of these SAs in the related toxicity of compounds [99, 121, 129, 164]. Despite the remarkable progress, identification of SAs of azole-based drugs/chemicals that are active towards aromatase has not been given much attention [24]. In addition, the exploration of structural alerts of azole chemicals which are non-responsive to human aromatase is still awaiting investigation for desired chemical applications, environmental safety and human health, because such chemicals are being used as plant protection products, anti-fungal and anti-bacterial [170, 171].

In this section of the chapter, we attempt to rationalize and disclose unique structural alerts that can exclusively guide the activity/inactivity of azole compounds toward aromatase. To achieve this goal, all possible fragments of azole chemicals have been extracted by using SARpy software [129]. We have incisively analyzed these fragments based on preset parameters and identified key SAs that can be used to classify azoles into active (agonist or antagonist) or inactive toward aromatase CYP19A1. The potential applications of these structural alerts as a classification model have also been explored. It is further anticipated that these findings can open up new gates for medicinal chemistry

community to design azole-based drugs/chemicals as per desired applicability. Mechanistic details and toxicity attributes (agonism/antagonism) of structural alerts have also been analyzed by comparing active (agonist and/or antagonist) and inactive chemicals within the SAR analysis. Figure 3.1 summarizes the strategy implemented within this chapter.

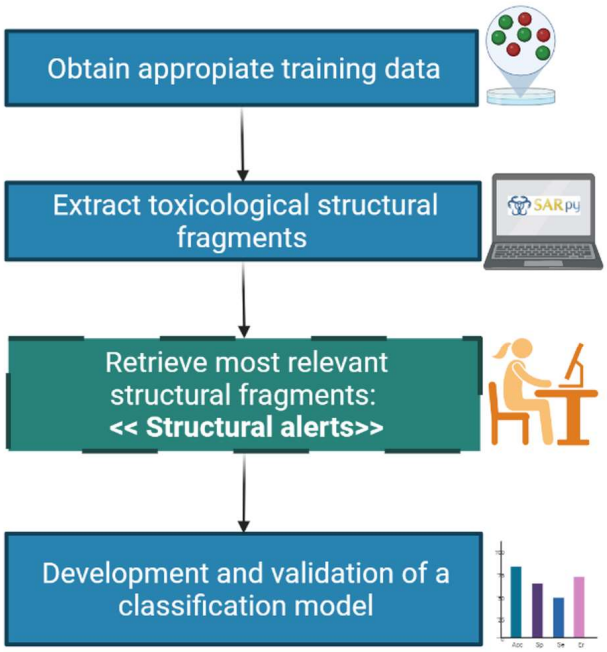


Figure 3.1: Strategy for the development of structural alerts for the prediction of aromatase binding.

3.1.2 Methods

3.1.2.1 Data composition

For the extraction of the relevant fragments and structural alerts of this section, the experimental outcome of each chemical was reported as antagonist, agonist and inactive. The final dataset comprised 326 azole compounds of;

i) 82 monoazole compounds of which 61 were inactive and 21 active (18 agonists and 3 antagonists).

ii) 198 diazoles of which 148 were inactive and 50 active (17 agonists and 33 antagonists).

iii) 46 triazoles of which 26 were inactive and 20 active (5 agonists and 15 antagonists).

To explore the diversity of the chemical space, the Tanimoto index [172] employing PubChem fingerprints was used as a similarity measure. The distance matrix was computed using KNIME [173] and can be visualized through heat maps for the final full active data set and classes of azoles. The heat maps are shown in Figure 3.2 for full active data set and classes of azoles, respectively, where green points indicate the highest similarity between two chemicals and blue points correlate to lower similarity.

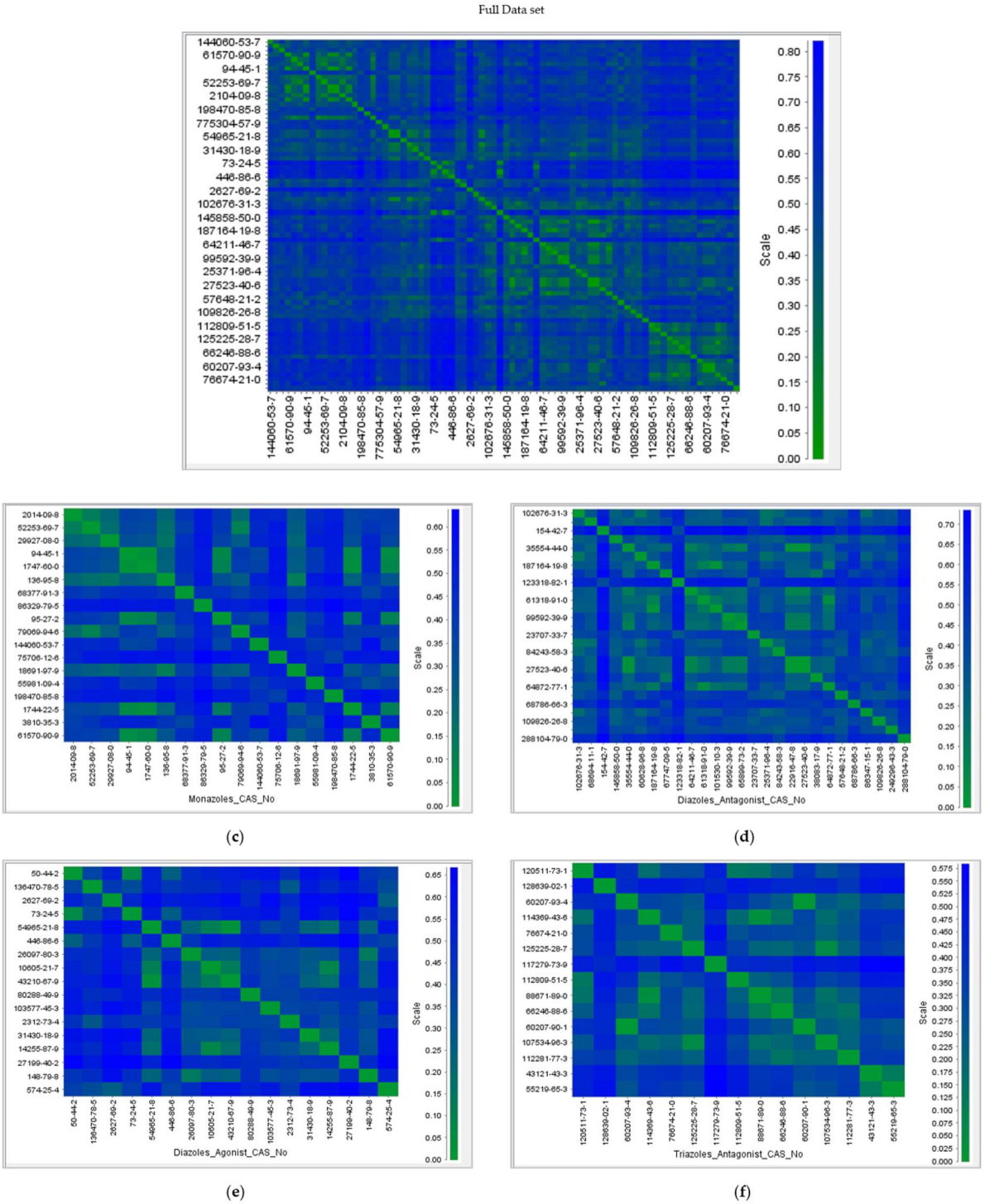


Figure 3.2: The heat maps showing the diversity of the data set. a.) Heat map for full data set. b) Heat map for the classified data set based on activity and class of the azole. On left side agonist monoazoles (Thiazole/oxazole) (c) and agonist diazoles (imidazoles and benzimidazole) (d) and antagonist triazoles (f) are shown.

3.1.2.2 Chemical structure and fragment extraction

Fragments from 326 azole compounds were extracted using the SARpy, as proposed within Section 2.6. The precision was defined as "min" to obtain results with higher sensitivity. Three training sets corresponding to monoazole, diazole and triazole subclass along with their respective activity were employed to obtain the active and inactive structural fragments towards aromatase CYP19A1.

3.1.2.3 Statistical evaluation of the fragments

The predictability and reliability of each fragment and classification model were described by the statistical parameter 'Accuracy' (Acc). In addition, other diagnostics, namely likelihood ratio (LR), sensitivity (Se) and specificity (Sp), were also calculated as proposed in Section 2.10. The reliability assessment of fragments using accuracy and LR values helped us to obtain the most relevant and meaningful fragments which were considered as the expert structural alerts. Structural alerts can work jointly to create a classification model for chemical screening. When more than one structural alert matches a chemical, activity prediction is performed based on the structural alert with higher LR value, while when no fragment is identified, no prediction is made. Applying this logic, the performance of the classification model was assessed to classify the full azole dataset into active and inactive compounds.

3.1.2.4 Cross-validation procedure

Validation of classification models comprising SARpy-extracted fragments and selected SAs was analyzed through parameters such as specificity, sensitivity, accuracy, error rate and unpredicted rate. Significance of models can be assessed by comparing their performance parameters. Such a procedure can be regarded as an internal cross-validation. It is further anticipated that screening of the full dataset using the SAs will allow the generation of structural categories that can group chemicals with respect to aromatase activity. Such categories would be useful to support a read-across assessment and are presented within Chapter 5.

3.1.3 Results and discussions

3.1.3.1 Insights to data and fragments

41 structural fragments (see

Table 3.1) were extracted; 13 for the monoazole subclass listed as SA1.1-SA1.13 (see Appendix A.1), 20 for the diazole subclass listed as SA2.1- SA2.20 (see Appendix A.2) and 8 fragments for the triazole subclass as SA3.1-SA3.8 (see Appendix A.3). These fragments were categorized into two types; i) Type-1: Fragments associated with the basic parent ring/azole moiety and ii) Type-2: Fragments representing the group(s) of atoms that can be regarded as functional groups or substituents around the basic parent ring of azole moiety. These two types of fragments along with their identity, activity, accuracy values and frequency of occurrence within subclass and full dataset are summarized in

Table 3.1. Additionally, the comparison of frequencies of occurrence of the fragments within the full dataset (FD) and within the monoazole, diazole, triazole (MDT) subclasses appears presented in Appendix A.4, the distribution of monoazole, diazole, and triazole fragments by chemical compound appears presented in Appendices A.5, A.6 and A.7, respectively.

Since most substances in all three subclasses were inactive, the number of inactive fragments was higher than active. Active fragments can be used as a screening tool to predict potential aromatase activity of new chemicals, read-across or adverse outcome pathways (AOP) development whereas inactive fragments can be regarded as an *inactive classifier*. All fragments having the accuracy of 1.00 can be referred to as *ideal classifiers*.

Table 3.1: SARpy fragments obtained from azole-based drugs/chemicals along with their identity, activity, accuracy, and frequency of occurrence within the subclass and full dataset.

Type Of azoles		Fragment identity	Activity	Accuracy within subclass	Accuracy within full dataset	Frequency within subclass	Frequency within full dataset
Monoazoles	**Type-1**	SA1.3	Active	0.75	0.75	4	4
		SA1.8	Inactive	1.00	1.0	6	6
		SA1.12	Inactive	0.9	0.91	21	22
		SA1.1	Active	1.00	1.0	5	5
		SA1.4	Active	0.73	0.73	11	11
		SA1.11	Inactive	0.83	0.83	18	18
		SA1.13	Inactive	1.00	1.00	3	3
	Type-2	SA1.2	Active	0.33	0.29	12	49
		SA1.5	Active	0.30	0.25	27	72
		SA1.6	Inactive	1.00	0.88	8	25
		SA1.7	Inactive	1.00	1.00	6	9
		SA1.9	Inactive	1.00	0.94	7	16
		SA1.10	Inactive	1.00	1.00	4	9
Diazoles	**Type-1**	SA2.1	Active	1.00	1.00	7	7
		SA2.8	Active	0.34	0.34	53	53
		SA2.10	Inactive	0.93	0.93	15	15
		SA2.12	Inactive	1.00	1.00	9	15
		SA2.17	Inactive	1.00	1.00	5	5
		SA2.11	Inactive	1.00	1.00	12	12
		SA2.19	Inactive	0.83	0.83	6	6
		SA2.9	Inactive	1.00	1.00	32	32
	Type-2	SA2.3	Active	0.79	0.71	14	21
		SA2.4	Active	0.55	0.51	31	65
		SA2.7	Active	0.39	0.44	61	105
		SA2.14	Inactive	1.00	0.96	12	26
		SA2.15	Inactive	1.00	1.00	9	10
		SA2.2	Active	0.56	0.33	9	21
		SA2.5	Active	0.55	0.50	11	14
		SA2.6	Active	0.33	0.42	18	24
		SA2.13	Inactive	1.00	0.92	15	37
		SA2.16	Inactive	1.00	0.96	10	24
		SA2.18	Inactive	0.93	0.91	43	75
		SA2.20	Inactive	0.75	0.72	97	160
Triazol es	**Typ**	SA3.3	Active	0.67	0.67	6	6
		SA3.4	Active	0.67	0.67	6	6
		SA3.5	Inactive	0.88	0.88	8	8
		SA3.1	Active	0.83	0.41	6	27

		SA3.2	Active	1.00	0.88	4	10
		SA3.7	Inactive	1.00	0.91	7	66
		SA3.8	Inactive	1.00	0.96	5	28
		SA3.6	Inactive	0.58	0.72	24	160

3.1.3.2 Fragment evaluation

Fragments listed in

Table 3.1 can form the basis of structural alerts which could be used to identify potentially toxic molecules or for chemical grouping. Structural alerts for toxicity prediction and grouping can, in many circumstances, be defined as being associated with the identification of chemicals having a common mode of action for biological or toxicological endpoints. However, since a single molecular fragment is not always responsible for the overall toxicity (or lack of toxicity) of a chemical, the analysis of the combination of fragments within a structure is crucial. Though both false negatives and false positives errors can occur while predicting the toxicity of new chemicals, false negative errors are not desirable in a precautionary approach. In this work, strict criterions for the inclusion of inactive fragments with respect to active fragments were applied to reduce false negative errors. Whilst any fragment having an accuracy greater than 0.65 can be considered as a potential classifier, there should be detailed consideration of the safety assessment of a chemical to yield the best and the most reliable outcome. Therefore, different treatments were employed to both types of active and inactive fragments with a defined applicability domain. We have assumed that an active fragment having an accuracy of ≥ 0.65 within its initial subclass and the full dataset, and with a frequency of occurrence in three or more chemicals, will have equal weight as an inactive fragment present in five or more chemicals with an accuracy of ≥ 0.90 within the full dataset.

Besides considering the accuracy and frequency of occurrence in the full dataset, Type-1 fragments that conveyed the same chemical information were also analyzed to develop the final model. Though fragments containing an azole moiety extracted from one subclass (say monoazole) have no probability of being present in other two subclasses (diazole and triazole), such fragments need to be verified for their chemical information and frequency of occurrences within the subclass dataset. For example, in monoazoles, the fragment SA1.1 (4-methoxy-2-amino-benzothiazole) was a derivative chemical of SA1.4 (2-amino-benzothiazole) having the frequency of occurrence 5 and 11, respectively (see

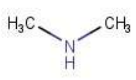
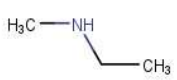
Table 3.1). Thus, broader structural applicability domain of SA1.4 with respect to SA1.1 showed to cover a larger domain of chemicals including SA1.1. Therefore, fragment SA1.4 can be considered a better fragment over SA1.1 for the development of the final model. It was also observed that in many cases the activity of a particular azole fragment can be tuned by the presence of functional groups/substituents attached to the azole ring. For example, fragment SA1.13 (2-benzoxazolinone) was an inactive fragment with the accuracy value 1.00 and mainly included the derivative chemicals having a substitution at the benzene ring of 2-benzoxazolinone. However, a derivative chemical of SA1.13, namely phosalone, was an active antagonist due to the substitution with O,O,S-trialkyl-dithiophosphate group at the nitrogen atom of the 2-benzoxazolinone. This observation indicated that the activity of phosalone was not accounted by the SA1.13, but by the functional group/substituent attached to the nitrogen of the 2-benzoxazolinone ring.

Therefore, functional groups/substituents attached to the azole moiety were also playing a pivotal role toward activity. The dependence of the activity on the functional groups/substituents was further evidenced by the inactive fragment SA1.11 (isooxazole) which had good accuracy (Acc = 0.83) and frequency of occurrence (18 chemicals). However, two active agonist chemicals, namely leflunomide and parecoxib, bearing SA1.11, showed that activity depends upon the side residues attached to azole ring rather than azole ring (isooxazole) itself. Similarly, in diazoles, the active fragment SA2.8 (imidazole) was a generic fragment and chemical imidazole itself was an inactive diazole and present in 53 chemicals but with the low accuracy (Acc = 0.34). However, in this case also, the activity of the imidazole can be investigated using structure-activity relationships by R-group decomposition around the imidazole ring. Unlike azole ring fragments, the fragments representing similar functional group/substituents can be observed in more than one subclass. Therefore, it was worth analyzing such functional group fragments before extracting them as structural alerts.

3.1.3.3 Analysis of functional group fragments

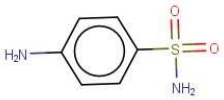
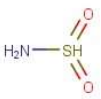
Fragments containing amine as a functional group: Inactive fragments SA1.6 (primary amine), SA2.18 (secondary amine), and SA3.7 (secondary amine) were extracted from monoazoles, diazole and triazoles with accuracy values of 1.00, 0.93 and 1.00, respectively. The secondary amine fragment SA3.7 (N-methyl,N-ethylamine) and SA2.18 (N,N-dimethyamine) differed only in number of carbon atoms in the alkyl chain. However, it was observed that fragment SA2.18 was found in total 75 molecules in full dataset along with 7 molecules that contain fragment SA3.7. Moreover, accuracy value (0.91) of fragment SA2.18 in full dataset was approximately like the accuracy (0.93) of fragment in diazole subclass. These observations showed that amine fragment SA2.18 performs better than SA3.7 as an inactive fragment for aromatase CYP19A1. On the other hand, the primary amine fragment SA1.6 was present in 8 molecules within the monoazole subclass and in 25 molecules within the full dataset. The accuracy value, which was 1.00 in the monoazole subclass, 0.88 in the full dataset, showed that the reliability of fragment SA1.6 decreases significantly when applied to the full dataset (see Table 3.2).

Table 3.2: Fragments containing amine as a functional group.

SA_ID	SA1.6	SA2.18	SA3.7
Chemical structure			
Type of amine	Primary amine	Secondary amine	Secondary amine
Chemical name	Butylamine	N,N-dimethyamine	N-methyl,N-ethylamine
Acc within the respective subset	1.00	0.93	1.00
Acc within full dataset	0.88	0.91	0.91

Functional group fragments containing sulphur (see Table 3.3)**:** There is only one active fragment SA2.2 having thiol as a functional group. Though this fragment has good frequency of occurrence 21, poor accuracy value (0.55) listed SA2.2 as a poor structural alert. Inactive fragments SA1.7 (sulphanilamide), SA2.14 (phenylsulfinyl) and SA3.8 (sulfonic amide) contained a functional group in the form of sulphur compounds and were chemically different to each other. Fragment SA1.7 and SA3.8 represented the sulphur atom in +VI oxidation state while SA2.14 represented the sulphur atom in +IV oxidation state. On the other hand, fragment SA1.7 and SA2.14 had an aromatic substituent as compared to fragment SA3.8 which was a generic form of fragment SA1.7. Screening the full dataset using SA1.7 found 9 chemicals which were a subset of 26 chemicals having SA2.14, therefore SA2.14 have a wider chemical space over SA1.7. Fragment SA3.8 had the distinguishable chemical feature of bearing +VI oxidation state of sulphur atom and represented a different subset of 28 chemicals. The accuracy values of fragment SA2.14 and SA3.8 in full dataset were 0.96 in both cases, which also accounted towards their reliabilities for the prediction purposes. In conclusion, SA2.14 and SA3.8 can encode adequate chemical information without overfitting for azoles' aromatase inactivity and hence qualified for the final classification model.

Table 3.3: Functional group fragments containing sulphur.

SA_ID	SA2.2	SA1.7	SA2.14	SA3.8
Chemical structure				
Chemical name	ethanethiol	sulphanilamide	Phenylsulfinyl	sulfonic amide
Oxidation state		+VI	+IV	+VI

Fragments containing acid as a functional group: Inactive fragments SA1.9 and SA2.13 correspond to propanoic acid and acetic acid, respectively, and have similar reliabilities (Acc = 0.94 and 0.92 when applied to the full dataset). However, fragment SA2.13 represented a bigger subset of chemicals including those from fragment SA1.9. Therefore, SA2.13 should be a better inactive classifier than SA1.9.

Fragments containing amide as a functional group: Fragments SA1.5, SA2.16 and SA2.5 represented the amide functionality (see Figure 3.3). Though SA1.5 and SA2.5 were active fragments, both have poor accuracy values within the subclass and the full data set. On the other hand, SA2.16 has high accuracy values within diazole subclass and full dataset, suggesting a good reliability of fragment SA2.16.

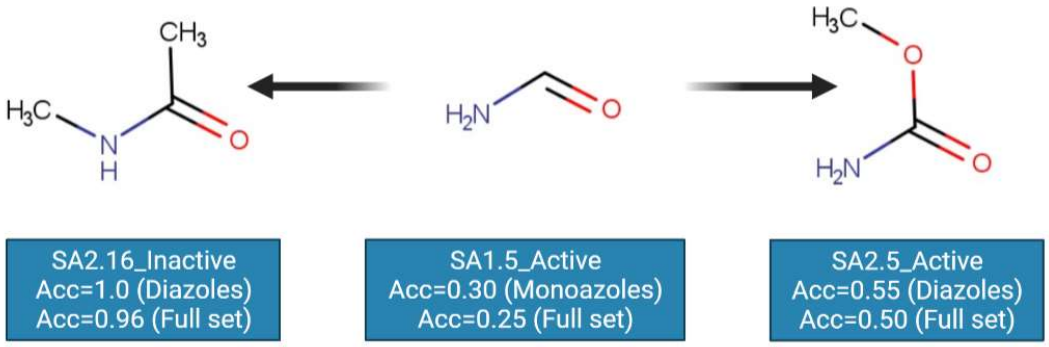


Figure 3.3: Effect of substituents on amide functionality towards CYP19A1 activity.

Guanidine and nitrile fragments functional group: Active fragment SA3.2 and inactive fragment SA1.10 were nitrile group and guanidine functional group, respectively. Both fragments qualified as expert SAs because they have a high accuracy value and required frequency of occurrence.

Fragments containing benzene ring: Active fragments SA2.7, SA2.3 and SA2.4 represented the substitution on the benzene with methyl, di-chloro and mono-chloro groups, respectively. Fragment SA2.3 had higher accuracy value (Acc=0.71) than fragments SA2.4 and SA2.7 (Acc=0.51 and 0.44, respectively) for the full dataset. Therefore, fragment SA2.3 should represent a better active classifier than SA2.4 and SA2.7. Nevertheless, all fragments were active towards the aromatase CYP19A1 irrespective of their accuracy values. Such type of behaviour can be explained by considering the electron-rich nature (π electron density) of these fragments bearing a benzene ring.

Other miscellaneous fragments: Fragment SA3.1 (ethyl propyl ether) was an active fragment with a high accuracy value (Acc=0.83) in the triazole subclass and represented six triazole chemicals. However, cross-validation of the fragment in full dataset showed that it had a low accuracy value (Acc=0.41) within the set of total 27 chemicals. Therefore, SA3.1 was specific to triazole subclass only and hence excluded from the list. Similarly, fragments SA1.2, SA2.2 and SA2.5 and SA2.6 were active but with low accuracy values for their source subclass as well as for the full dataset. Hence, these fragments should be excluded from being a member of the final model as an active classifier. However, such active fragments need to be given more attention to explore the factors which were turning them active or inactive with poor statistics. Therefore, the structure-activity relationships (SARs) were studied through R-group fragmentation around the basic parent ring of azole chemicals.

3.1.3.4 Fragment selection

Based on the above-mentioned evaluation and analysis of fragments, the following simple set of rules has been developed for the inclusion/exclusion of fragments as structural alerts and into the final classification model.

1. All fragments having the accuracy value less than 0.65 within their source subclass were excluded, irrespective of their activity class.
2. To select an active azole fragment (Type-1) and an active functional group fragment (Type-2), fragments must meet the following criteria;
   a) A minimum frequency of occurrence in three chemicals, and

b) An accuracy value greater than 0.65 for both source subclass and full dataset.

3. To select an inactive azole fragment (Type-1) and an inactive functional group fragment (Type-2), fragments must meet the following criteria;

a) A minimum frequency of occurrence in five chemicals, and

b) An accuracy value greater than 0.90 for the full dataset.

Following the implementation of above criteria, a total of 21 fragments were filtered from

Table 3.1, out of which 7 were active and 14 were inactive. In other words, we could also establish that 12 fragments were selected from the Type-1, and 9 from the Type-2. These 21 selected structural alerts (SAs) have been presented in Figure 3.4 where active and inactive SAs were highlighted with red and blue boxes, respectively.

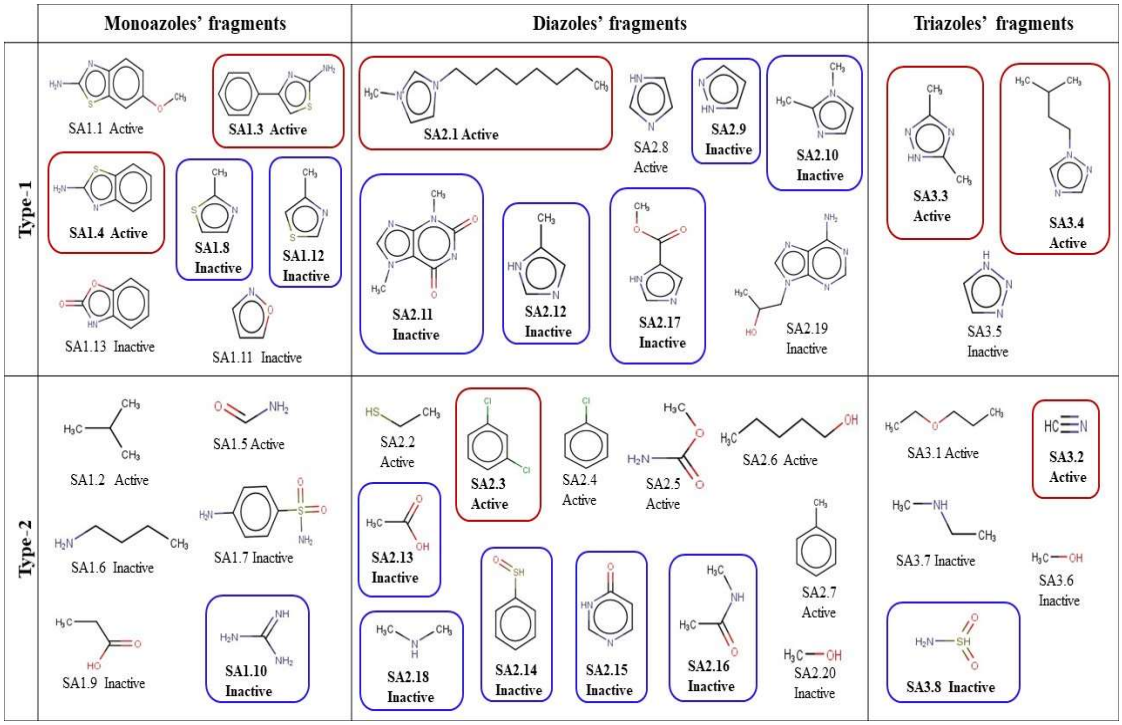


Figure 3.4: Identified 21 structural alerts (in boxes) for the final classification model.

3.1.3.5 Fragment application as classification model

Rigorous inspections of 41 fragments obtained from three initial subclasses (see

Table 3.1) have been reduced to 21 structural alerts (see Figure 3.4). These 21 SAs were used to build up a classification model. The model was evaluated through statistical parameters, sensitivity, specificity, accuracy, unpredicted rate and error rate. These parameters can be analyzed using the obtained confusion matrix which is presented within Table 3.4.

Table 3.4: Confusion matrix of the classification model composed of 21 structural alerts.

Active	Inactive	Unknown	Predicted
48	7	36	ACTIVE
9	166	60	INACTIVE

A total of 230 chemical structures, out of the initial 326 chemicals, were matched with at least one structural alert within the model of 21 SAs (see Appendix A.8). The sensitivity, specificity, accuracy, and error rate of the model were 84%, 96%, 93% and 5%, respectively, which described the correctness of the model for meaningful predictions (see Figure 3.5). The number of chemicals (96) that did not match any of the filtered SAs displayed the unpredicted rate (29%) of the purposed model. However, to interpret results more adequately, a short exercise was conducted. This exercise was executed for initially obtained 41 fragments to determine their performance rate with respect to the full dataset. The sensitivity, specificity and accuracy values for the model comprising 41 fragments were computed as 88%, 82%, and 84%, respectively. In particular, the error rate was 16%, much higher than the value of the model comprising 21 SAs (5%). It was observed that the error rate (5%) for 21 SAs was approximately three-fold lower than the error rate (16%) of the model comprising 41 fragments. The unpredicted rate of 29% by the 21 SAs model is the cost to pay in order to have more favorable error rate and accuracy. All these investigations illustrated the effectiveness and validity of the model that identified 21 SAs to precisely classify azole-based drugs/chemicals into active or inactive towards aromatase CYP19A1. Figure 3.5 below shows the statistical parameters of the filtered (21 SAs) and unfiltered (41 SAs) classification models.

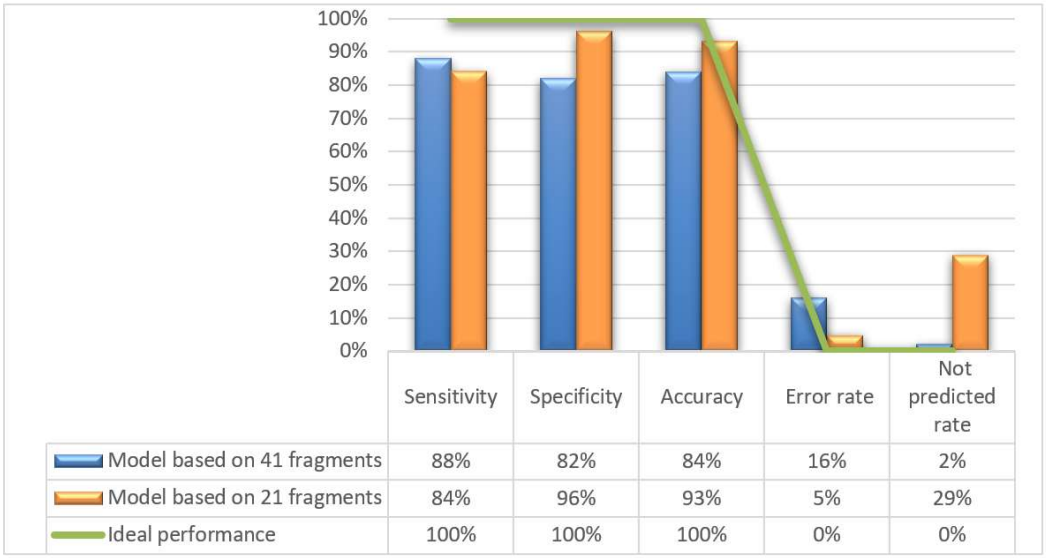


Figure 3.5: Comparison of statistical parameters for the filtered (21) and unfiltered (41) structural alerts when used as a classification model.

3.1.4 Concluding remarks

This chapter has disclosed structural requirements for azole chemicals with respect to the aromatase enzyme CYP19A1 activity. A classification model comprising 21 structural alerts has been put forward that can categorize azoles into active and inactive toward the aromatase CYP19A1. The classification model revealed that simple methylation of 1,3-thiazole, imidazole and xanthine scaffolds resulted in inactivity while methylated 1,2,4-triazoles were active. On the other hand, amination of 1,3-thiazole and benzothiazole, and arylation of 1,3-thiazole and diazole scaffolds were significant for activity. Identified structural alerts can be employed as experts for the toxicity assessment toward aromatase.

In conclusion to the classification modeling (see Figure 3.4), it was observed that the methylation of the 1,3-thiazole, imidazole, purine scaffolds and amide group leads to inactivity towards CYP19A1. Carbonyl functional groups such as ester and carboxylic acid were found to contribute to inactivity. On the other hand, methylation of 1,2,4-triazoles can lead to potentially active derivatives for CYP19A1. Like inactivity, the activity was found to be dependent on the type of substituents attached to azole rings. For example, an amine group was significant for 1,3-thiazoles and benzothiazoles while aryl substituents (benzene, chorobenzene etc.) were significant for 1,3-thiazoles and diazoles. It is important to highlight that within a well-defined applicability domain; this classification model will prioritize chemical safety.

3.2 Structural Fragments for Active Agonist and Active Antagonist Chemicals

3.2.1 Introduction

As previously mentioned, azoles active on aromatase can be classified as an active agonist and active antagonist regarding the mode of their aromatase CYP19A1 binding activity, where active agonists and active antagonists both contribute to endocrine disruption. To distinguish the two activities, the antagonist is represented by the pIC50 (logarithmic inhibition concentration) and the agonist activity is represented by the pEC50 (logarithmic effective concentration). In this section, we have extracted structural alerts for both agonism and antagonism from a subset of active azoles and we have proposed a classification model. Due to the poor agonist and antagonist data availability and the complexity associated to this data, the methods presented in Section 3.1.2 to extract structural alerts were slightly modified and adapted to the new requirements, those adjustments are specified below.

3.2.2 Methods

3.2.2.1 Data composition

For the extraction of the relevant fragments and structural alerts of this section, some extra data curation steps were performed, i) the elimination of inactive compounds, and the removal of active classes of chemicals with poor representation for modeling purposess. For these reasons 248 inactive compounds, and 3 antagonist monoazoles and 6 agonist triazoles were removed, respectively. The experimental outcome of each chemical was reported as antagonist or agonist.

The final dataset comprised 78 azole compounds of;

i) 18 active monoazole compounds (18 agonists).
ii) 45 active diazoles (18 agonists and 27 antagonists).
iii) 15 active triazoles (15 antagonists).

3.2.2.2 Fragment extraction

Fragments from 78 active azole compounds were extracted following the methodology described in Section 2.6. The precision was defined as "min" to obtain results with higher sensitivity. In this case, due the poor availability of active compounds, only one training set containing all monoazole, diazole and triazole active subclasses along with their respective agonist/antagonist activity was employed to obtain the active agonist and active antagonist structural fragments towards aromatase CYP19A1.

3.2.2.3 Classification modelling and evaluation

The accuracy calculated refers to the ability of SA to yield the correct predictions. The accuracy values close to 1 were desired, and were interpreted as a better performance during the classification [166]. All predictions by certain fragment can be true if Acc=1 and LR= infinite (inf.). However, incorrect predictions are influenced by the accuracy of the given SA, the presence of other SAs with similar accuracy value for opposite cases and by the stereo-chemical environment of the SAs in the azole compounds. Therefore, levels of accuracy (Acc), low (Acc≤0.6), low to medium (0.6<Acc≤0.7), medium (0.7<Acc≤0.8) and high (0.9<Acc≤1) were assigned to each SA. The assignation of qualitative levels of accuracy facilitated the interpretation of results and the evaluation and accuracies (acc) of SAs were also compared for their consistency with the literature. The calculated accuracies of SAs were found to be similar when crosschecked with literature studies [174-180] (see Table 3.5). The LR values calculated using Eq. (2.6), which gave a measure of the degree of accurate predictions using the distribution of the SA between the activities (agonist and antagonist compounds) and ratio of true and wrong predictions.

High LR values mean that a SA was predominantly found to contribute to one of the two activities, and the value "inf." means the alert was a perfect classifier. Largest values of LR can be interpreted as a highest relevance of the SA; however, unlike the accuracy where the numerical range is well defined, the wide numerical range of LR values could be difficult for interpretation.

Additionally to the real accuracy (accuracy as presented in Eq. (2.5)), the random accuracy and the difference between these two parameters for the classification model and for each SA were calculated using the methodology proposed by Lučić, *et al.,* 2019 [181] and Batista, *et al.*, 2016 [182], and summarized in Eq (3.1) and Eq (3.2).

$$Acc_{,rnd} = 100\ \frac{(TP + FN)(TP + FP) + (TN + FP)(TN + FN)}{N^2} \tag{3.1}$$

$$\Delta Acc = Acc - Acc_{rnd} \tag{3.2}$$

The difference of real accuracy of SA from random accuracy is called delta accuracy which can have the maximal value of 50% when the maximal value of Acc = 100 %, and the experimental data set is balanced having the same numbers of both classes' agonist/antagonist when Acc_{rnd} = 50 %. This delta accuracy is a useful parameter for validation of two-state classification model quality when challenging datasets [182]. The structural alert with a high value of delta accuracy can be regarded as privileged structural fragments.

3.2.3 Results and discussions

3.2.3.1 Classification modelling

The classification model derived for two activity classes (agonist and antagonist) consisted of 11 SAs, 4 of these were associated with agonist and 7 with antagonist activity classes. The SMARTS of the SAs, the associated activity, likelihood ratio (LR), accuracy and other relevant information are reported in Table 3.5. The fragments with high (0.9<acc<1) and medium (0.8<acc<0.9) accuracy, and high LR values were considered as privileged SAs. Additionally, the chemical structure of the 11 fragments appears represented within Figure 3.6, where the SMARTS structures were obtained from SMARTSanalyzer - Analyze Chemical Patterns https://smartsview.zbh.uni-hamburg.de/. The discontinues circles represent the aromatic atoms, and the continues circles the aliphatic atoms.

SA_ ID	Name	SMARTS1	Activity related	LR value	Accuracy (Acc) of	Statistical Reliability	Literatur e	Relevant information of selected SA and distributions
SA1	1,3-thiazoles	c2cscn2	Active agonist	inf	1.0	High 0.9<Acc≤1	Medium	-
SA2	para substituted chlorobenzenes	Clc1ccc(CC)cc1	Active antagonist	inf	1.0	High 0.9<Acc≤1	High	Antagonist Diazoles (n=9; Average pIC50=8.40) Antagonist Triazole (n=7;
SA3	1,2,4-triazoles	n1cncn1	Active antagonist	inf	1.0	High 0.9<Acc≤1	High	-
SA4	carboxylic acids	C(=O)O	Active agonist	12.83	0.92	High 0.9<Acc≤1	High	Agonist Diazoles (n=7; AveragepEC50=8.88) Agonist Monozoles (n=4; AveragepEC50=9.09;
SA5	ethyl(propyl)amin e	CCCN(CC)	Active antagonist	1.71	0.67	Low to medium 0.6<Acc≤0.7	Low	-
SA6	benzylimidazoles	c1cn(cn1)C(c1ccccc1)	Active antagonist	inf	1	High	High	Antagonist Diazoles (n=3; Average pIC50=9.79)
SA7	1-phenyl-1H- Imidazoles	c1c(cccc1)n 1ccnc1	Active antagonist	inf	1	High	High	Antagonist Diazoles (n=3; Average pIC50=7.87) -
SA8	Formamides	C(=O)N	Active agonist	4.96	0.81	Medium to high	Medium	Agonist Diazoles (n=9; Average pEC50=8.53) Agonist Monozoles (n=8 ; AveragepEC50=8.47 ;
SA9	nzenes	Clc1cccc(c1)	Active antagonist	11.57	0.93	High	High	Antagonist Diazoles (n=16; Average pIC50=8.45) Antagonist Triazoles (n=11; Average pIC50=7.72;
SA1 0	Carbon chain	CC	Active antagonist	1.67	0.66	Low to medium 0.6<Acc≤0.7	Low	-
SA1 1	purines	c1ncnc2c1nc n2	Active agonist	2.33	0.67	Low to medium	Low to medium	-

Table 3.5: Structural fragments for agonist and antagonist activity on CYP19A1 obtained from classification modelling along with

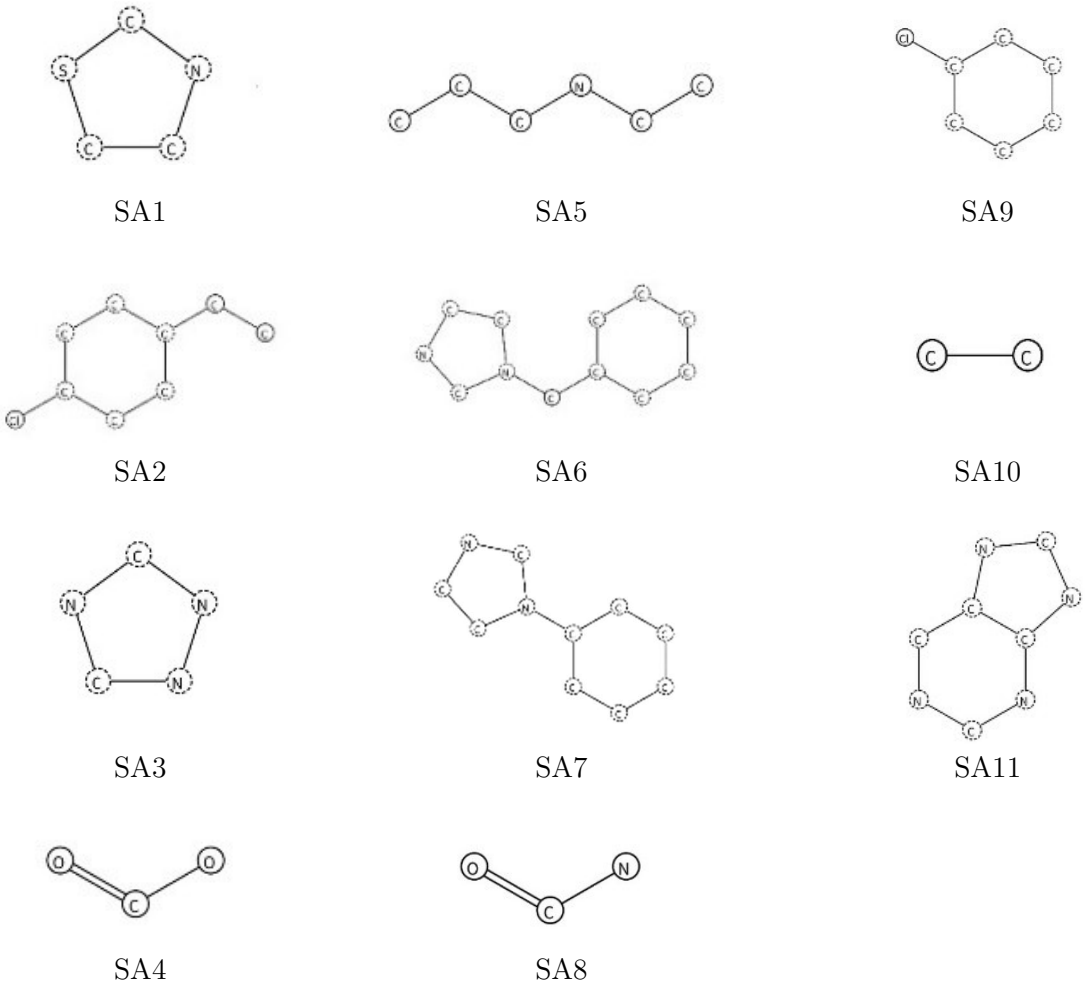


Figure 3.6: Chemical structure of structural fragments.

The real accuracy of the final classification model based on 11 SA for all 78 compounds was 92,3%, the most probable random accuracy was 50% giving the difference (delta accuracy ΔAcc) of 42% [182] (see Table 3.6). This delta accuracy (maximum value = 50%) of 42% can be considered as the real contribution of the model, indicating its good quality for the classification. The LR value was "inf." for most of the fragments; for the SA1 (monoazole) and SA3 (triazole) fragments all compounds in these classes in our data set were agonists or antagonists, respectively. This was regardless of the presence of other structural feature within monoazole and triazole classes contributing to opposing activity, even though these classes have been classified with high accuracy for respective activities in literature studies [174-180].

Table 3.6: Confusion matrix of the classification model together with the statistical parameter's accuracy, random accuracy, and delta accuracy.

Active agonist	Active antagonist	Unknown	Predicted
33	3	0	**Active agonist**
3	39	0	**Active antagonist**

Real Accuracy of the classification model (Acc)	92.30%
Random Accuracy of the classification model ($Acc_{,rnd}$)	50.30%

ΔAcc	42.01%

The diazole class in our data set was reasonably balanced for both activities (27 antagonists and 18 agonists), furthermore, the 1,3-diazole derivative antagonists SA6 (benzylimidazole) and SA7 (1-phenyl-1H-imidazole) had a LR value of "inf". The fragments SA6 and SA7 are differentiated by the chain length as one carbon atom is present between diazole ring and benzene ring in SA6, whilst in SA7, the diazole and benzene rings are directly connecting. The SA6 is more branched than SA7, however, the average activity of SA6 (average pIC50=9.79) was significantly greater than SA7 (average pIC50=7.87), thus **the branched benzyl group in SA6 imparted more inhibitory properties in diazoles to CYP19A1 than the phenyl group in SA6** (see Table 3.5). It must be noted that both SA6 and SA7 fragments were present in six diazoles (three for each) only.

The other antagonist fragments extracted were SA2 and SA9 with LR values "inf." and 11.3 respectively, and high Acc (see Table 3.5). Fragments SA2 and SA9 are likely to be unique to provide the information as these fragments were present in many antagonists, including the molecules belonging to triazole and diazoles classes (see Table 3.5). However, fragment SA9 was associated to two incorrect predictions for diazole molecules, namely selumetinib and pifexole, where the pifexole contained only SA9 fragment and selumetinib has agonist fragment SA8 along with SA9. There were only two common chemicals contained in the groups of compounds with (SA2, SA9) and (SA6, SA7), namely imazodan and liarazole, and two chemicals were common in SA2 and SA9. Chemically, the SA2 and SA9 group is chlorobenzene and p-alkyl (ethyl) substituted chlorobenzene. However, for the diazoles, their contribution to inhibitory activity was not the same as for SA6 and SA7. **The branched SA2 (p-alkyl substituted chlorobenzene) (average pIC_{50}=8.40, more branched) imparted less inhibitory properties than SA9 (chlorobenzene) (pIC_{50}=8.45, less branched) in diazoles** which is contrary to the behavior of SA6 (more branched) and SA7 (less branched). These observations showed that **the position of branching in the diazoles has a role in the inhibition of diazoles to CYP19A1.** This is in agreement with the observation by Ghodsi et al., 2016 [191] that an aromatic ring near the heterocyclic ring will cause delocalisation of HOMO electrons over the two rings and make them less available to interact with the heme group and, as a consequence, SA7 (less branched) was induced less inhibition as compared to SA6 (more branched) in diazoles. **It can be concluded that the diazole molecules favor the inhibition of CYP19A1 by including the chlorobenzene fragment and not the branched p-alkyl (ethyl) substituted chlorobenzene.**

However, SA2 and SA9 fragments present in triazoles have the opposite effects with respect to branching increasing inhibition. All 15 antagonist molecules contained either SA2 or SA9 fragments and two molecules, ipconazole and fenbuconazole, contained both SA2 and SA9 fragments. **The inhibitory properties of triazoles were more favored by the branched p-alkyl (ethyl) substituted chlorobenzene (Average pIC_{50}=7.89) than chlorobenzene (Average pIC_{50}=7.72)** which was contrary to that in diazoles (see Table 3.5). **In conclusion, the inhibition properties change with substitution of the same substituents in the different heterocyclic rings (diazole and triazole).**

For the agonist activity, only two relevant fragments, SA4 and SA8 were identified. SA4 was carboxylic functional group; the LR value was "inf." and Acc. high (0.93) (see Table 3.5). The fragment SA8 was the amide functional group having a LR value of 4.96 and its statistical reliability was medium to high (0.81). Chemically both groups offer conjugation over a short range which can cause the separation of electrostatic charges on the azole compounds. **In conclusion, for the classical classification modelling,**

fragments in azoles having aromatic resonance (e.g., in benzene) were found to be antagonists, whilst small groups which separate the charges are found to control their agonistic activity. Fragments SA5, SA10 and SA11 had low LR, accuracy and therefore poor reliability.

3.2.4 Concluding remarks

Human aromatase activity has been studied for azole classes comprising of triazoles, diazoles, thiazoles for their reversible inhibition and agonist activity. The classification modelling suggested that the chemical nature and position of substituents (chemical groups) on diazoles and triazole ring had different contributions to inhibition, while functional groups having resonating charges have a significant role for agonist activity. The understanding of the structural factors contributing to agonist/antagonist activity is crucial for drug design and safety assessment, however the poor data availability within both classes hindered the process.

Chapter 4

Structure-Activity Relationships

4.1 Introduction

Structure Activity Relationships can be used as theoretical models to predict biological/toxicological end points of substances [192, 193]. SAR analysis can establish a qualitative link between a chemical structure and a chemical containing the substructure (e.g., creation of a structural alert) to display a certain biological property or effect (e.g., aromatase activity) [194, 195]. SARs are based on the principle that similar structures may have similar physical and biological properties [196, 197]. These predictive models not only reduce the cost, time and concerns associated with toxicological assays, but can also guide the acquisition or synthesis of desirable new compounds, and the characterization of new molecules [197]. This analysis is typically assessed in the form of a table, or SAR tables, where chemical structures and substructures are displayed together with their activities. When performed well, experts can use SARs to understand mechanisms of toxic action, chemicals safety, and anticipated health risks. The confidence of the analysis depends on the database analyzed where the extracted knowledge needs to be carefully defined [192, 193, 196].

The mechanistic insights of aromatase binding of azoles have been explored by the fragmentation (R-group decomposition) of substituents in molecules around the basic parent rings and some manually generated molecular species, also considering the structural alerts developed in Chapter 3 [198]. Such fragmentation has a valuable contribution for medicinal chemistry and the synthesis of new compounds, through the definition of transparent specific structural factors contributing to the azoles activity/toxicity on human aromatase. Additionally, the analysis facilitated the decision-making process during the fragment evaluation and selection, and the explanation of the imperfections of the individual fragments which in turn imply the unreliable active fragments observed during the classification modelling process.

In this chapter, the structure activity relationships analysis was built on the structural fragments obtained from the classification model which explored active/inactive classes (see Section 3.1). Due to poor data availability, fragments retrieved for agonist/antagonist classes (see Section 3.2) were used solely to support the examination.

4.2 Results and discussions

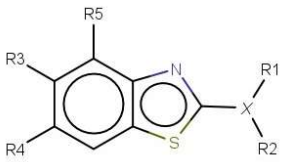
4.2.1 SAR analysis of drugs/chemicals containing monoazole fragments

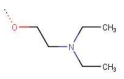
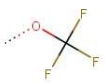
4.2.1.1 Benzothiazole scaffold

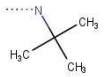
The benzothiazole scaffold was built using the structural information contained within SA1.1 (2-amino-6-methoxybenzothiazole) and SA1.4 (SA1.4 was 2-aminobenzothiazole). Both fragments are chemical derivatives of benzothiazole and were active showing that **substitution at position-6 of the 2-aminobenzothaizole with sp^3 hybridized oxygen atom can turn the compound exclusively active for aromatase CYP19A1**.

The comprehensive overview of the chemicals within this category is reported in Table 4.1. The substitution at position-2 of the benzothiazole can be hetero atom/s such as nitrogen (N) or Sulphur (S). It can be observed that substitution of sp^3 hybridized sulphur atom at position-2 of benzothiazole makes the compound antagonist while the compounds having substitution of sp^3 nitrogen atom are active agonists or inactive. A careful investigation of Table 4.1 indicated that **all compounds having electron donating group(s) at position-5 and/or 6 (R3 and/or R4) were active toward CYP19A1**.

Table 4.1: Drugs/chemicals containing benzothiazole scaffold and their activity towards aromatase.



CAS	Chemical name	R1	R2	R3	R4	R5	X	Activity (µM)	Structural Alerts
136-95-8	2 – aminobenzothiazole	H	H	H	H	H	N	Agonist (31.20)	1.4
95-27-2	Diamthazole	CH_3	CH_3	H		H	N	Agonist (22.38)	1.1, 1.4
1744-22-5	Riluzole	H	H	H		H	N	Agonist (9.44)	1.1, 1.4
1747-60-0	2-Amino-6-methoxybenzothiazole	H	H	H	$-OCH_3$	H	N	Agonist (6.60)	1.1, 1.4
18691-97-9	Methabenzthiazuron	CH_3		H	H	H	N	Agonist (2.09)	1.4
94-45-1	2-Amino-6-ethoxybenzothiazole	H	H	H	$-OC_2H_5$	H	N	Agonist (1.71)	1.1, 1.4

29927-08-0	2-Amino-5,6-dimethylbenzothiazole	H	H	CH_3	CH_3	H	N	Agonist (1.23)	1.4
61570-90-9	Tioxidazole		H	H		H	N	Agonist (0.87)	1.1, 1.4, 1.5
1477-42-5	2-Amino-4-methylbenzothiazole	H	H	H	H	CH_3	N	Inactive	1.4
19952-47-7	2-Amino-4-chlorobenzothiazole	H	H	H	H	Cl	N	Inactive	1.4
5464-79-9	2-Amino-4-methoxybenzothiazole	H	H	H	H	$-OCH_3$	N	Inactive	1.4
95-31-8	N-tert-Butyl-2-benzothiazolesulfenamide		H	H	H		S	Antagonist (42.37)	1.2

Structure-activity relationship was observed for the agonist chemicals originating from the 2-amino benzothiazole by substitutions at sites R1, R2, R3, R4 and R5, where R1, R2 were present on the nitrogen atom of amino group and R3, R4 and R5 on the benzo group of the benzothiazole. In Table 4.1, 2–Aminobenzothiazole was an active agonist with activity concentration 31.2 µM, the highest activity concentration (least active) among the series of compounds, while the variation with different groups (R1, R2, R3, R4, and R5) on 2-Aminobenzothiazole had decreased the agonist activity concentration. It can be noticed in Table 4.1 that the substitution at R5 with chlorine, methyl, and methoxy (while keeping other positions with hydrogen atom) holds the chemicals inactive. **These observations indicated that the position-R5 on the benzene ring** of 2-aminobenzothazole **was not sensitive to activity**, see compounds, namely 2-Amino-4-methylbenzothiazole, 2-Amino-4-chlorobenzothiazole and 2-Amino-4-methoxybenzothiazole. The role of the position-R4 of 2-aminobenzothiazole in compounds, namely Riluzole (9.44 µM), 2-Amino-6-methoxybenzothiazole (6.60 µM), 2-Amino-6-ethoxybenzothiazole (1.71 µM) having trifluromethoxy, methoxy, and ethoxy groups respectively, **had a greater impact on activity concentrations**. **Therefore, as the electron donating tendency of groups at position-R4 increases, the activity concentration decreases (activity increases) considerably**.

Remaining active chemicals, namely Diamthazole (22.38 µM), Methabenzthiazuron (2.09 µM), 2-Amino-5,6-dimethylbenzothiazole (1.23 µM) and Tioxidazole (0.87 µM), displayed the collective behaviour of positions, namely (R1, R2, R4), (R1, R2), (R3, R4) and (R2, R4), respectively, where R1, R2 define the substitution at nitrogen atom of the amino group and R3, R4 define the substitution at benzene ring of the 2-aminobenzothiazole. Indeed, the substitution on nitrogen atom of the amino group with the methyl group (R1) and N-methyl acetamide (R2), as in Methabenzthiazuron (2.09 µM), indicated that **the electron withdrawing carbonyl group also has a considerable impact on the decrease of the activity concentration** as compared to the parent 2-aminobenzothiazole. Similarly, the effect of substitution at R2 with carbonyl ester can be seen in Tioxidazole (0.87 µM), which is the most active compound in Table 4.1. Diamthazole (22.38 µM) having a substitution at R1, R2, R4 with methyl, methyl and 2-(N-diethyl)ethoxy, respectively, was slightly less active than the 2-aminobenzothiazole

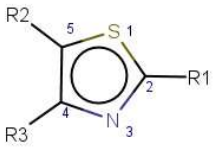
indicating that the appropriate group at one position (R1, R2, R4, R3) or two positions (R2, R4) and (R3, R4) simultaneously can be more effective in decreasing the activity concentration than the increase in number of substitution simultaneously on the 2-aminobenzothiazole. **The presence of amino group in position-2 of benzothiazole moiety seems to be crucial for the toxic effect (agonist),** since the chemicals like benzothiazole, 2-Bromobenzothiazole and 2-Methylbenzothiazole compounds were inactive. In summary, SAR analysis of Table 4.1 predicted that **the presence of electron donating group and electron withdrawing group on the benzene ring and nitrogen of the 2-aminobenzothiazole respectively were very crucial for the desired aromatase activity.**

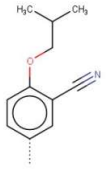
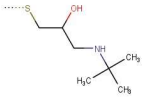
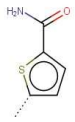
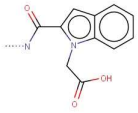
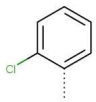
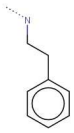
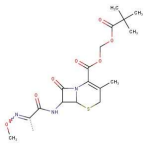
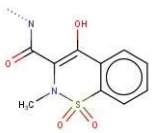
4.2.1.2 1,3-thiazole scaffold

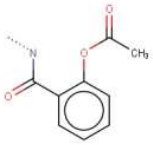
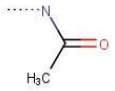
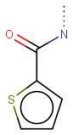
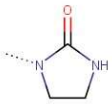
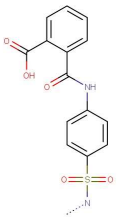
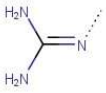
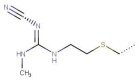
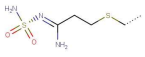
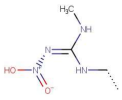
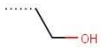
Structure alerts SA1.3, SA1.8 and SA1.12 represented substitution of different groups at diverse positions in the 1,3-thaizole scaffold. 1,3-Thiazole has three sites for substitutions at positions-2, 4 and 5 with their substituent as R1, R3, and R2, respectively. Therefore, all compounds with 1,3-thiazole ring had been presented in Table 4.2. Structural alert SA1.3 represented the substitution on position-2 and 4 with amine group and phenyl group, respectively. On the other hand, SA1.8 and SA1.12 had methyl group present at position-2 and 4, respectively. SA1.3 was an active alert with medium accuracy (Acc=0.75) while both SA1.8 and SA1.12 were inactive structural alerts. Compounds such as 2-Amino-4-(p-nitrophenyl)thiazole, 2-Amino-4-phenylthiazole, Fanetizole corresponded to SA1.3 and were an active agonist toward aromatase. These compounds had amine functionality at position-2 and aryl group at position-4. Both these groups can provide extended conjugation to the 1,3-thiazole ring and thereby show increased π-electron density over the molecule, producing a greater interaction with the aromatase. The agonist activity order of these compounds was 2-Amino-4-(p-nitrophenyl)thiazole < 2-Amino-4-phenylthiazole < Fanetizole. The high agonist activity of Fanetizole might be due to the better electron donating amine group present at position-2. Compounds such as Febuxostat, Meloxicam, Thiamine thaiozole, Cefodizime and 5-Acetyl-2,4-dimethylthiazole had SA1.8 or SA1.12 fragment as well as bearing substituents at positions-2 of the 1,3-thaizole scaffold. Febuxostat having aryl substituent at position-2 was an active agonist (0.036 µM). Meloxicam with amide as substituent at position-2 showed no activity while Thiamine thiazole without substitution at position-2 also resulted inactive. It is worth noting that Febuxostat, which requires low activity concentration/highly active (0.036 µM), has the *p*-alkoxy *m*-nitrile phenyl group at position-2, and position-4 and 5 substituents were comparable with Cefodizime and 5-Acetyl-2,4-dimethylthiazole. The high activity of Febuxostat could be dependent on the nature of the phenyl group and the substituent at position-5. It must be noted that the carboxylic group (-COOH) attached at the position-5 can withdraw the electron density from the 1,3-thiazole ring and control the flow of the negative charge from the *p*-alkoxy *m*-nitrile phenyl group in Febuxostat molecule as compared to the less active Cefodizime and inactive Fenclozic acid. **These observations had given a clear indication that an appropriate substitution at position-2 (more likely amine substitution/electron donor groups) and/or 4 (aryl group substitution) with groups that can provide extended conjugation to the 1,3-thiazole system could affect the activity toward the CYP19A1. Further, the activity can be increased to many folds by introducing a better electron donating group at position-2 of the 1,3-thiazole ring**. It must be noted that one contradictory case, namely Cefetamet pivoxil HCl, was an inactive compound in spite of

having NH_2 group at position-2. Such observations might be due to the inability of π-electron acceptance of group present at position-4 of the 1,3-thaizole ring.

Table 4.2: Drugs/chemicals containing 1,3-thiazole scaffold and their activity towards aromatase.



CAS	Name	R1 and its description	R2 and its description	R3 and its description	Activity (μM)	Structural alert
144060-53-7	Febuxostat			CH_3	Agonist (0.036)	1.2, 1.12
68377-91-3	Arotinolol hydrochloride		-		Agonist (22.38)	1.2, 1.5
136381-85-6	Lintitript		-		Inactive	1.3,1.5
2104-09-8	2-Amino-4-(p-nitrophenyl)thiazole	NH_2	-		Agonist (9.79)	1.3
52253-69-7	2-Amino-4-phenylthiazole	NH_2	-		Agonist (6.33)	1.3
79069-94-6	Fanetizole		-		Agonist (1.50)	1.3
86329-79-5	Cefodizime sodium				Agonist (29.85)	1.5, 1.12
111696-23-2	Cefetamet pivoxil HCl	NH_2	-		Inactive	1.2, 1.5, 1.9,1.12
71125-38-7	Meloxicam		CH_3	-	Inactive	1.5

55981-09-4	Nitazoxanide			-	Agonist (0.46)	1.5
140-40-9	Nithiamide			-	Inactive	1.5
3810-35-3	Tenonitrozole			-	Agonist (0.60)	1.5
61-57-4	Niridazole			-	Inactive	1.5
38205-60-6	5-Acetyl-2,4-dimethylthiazole	CH3		CH3	Inactive	-
85-73-4	Phthalylsulfathiazole		-	-	Inactive	1.5, 1.7
69014-14-8	Tiotidine		-		Inactive	1.10, 1.12
76824-35-6	Famotidine		-		Inactive	1.10, 1.12
210880-92-5	Clothianidin	Cl		-	Inactive	1.10
17969-20-9	Fenclozic acid		-		Inactive	1.12
137-00-8	Thiamine thiozole	-		CH_3	Inactive	1.12

There are few compounds (active or inactive) where nitrogen was present either in the form of amide functionality or sp2 nitrogen atom at position-2. A clear perusal of Table 4.2 reflected that **compounds (Nitazoxanide, Tenonitrozole) with aromatic amide substitution (-NCOAr) at position-2 were active agonists (EC = 0.46 and 0.60 µM), whereas those with aliphatic amide (-NCOR) or cyclic amide substitution (as in Nithiamide, Niridazole) were inactive**. In the case of thiazole having aliphatic or cyclic amide substitution, the lone pair of nitrogen might have been involved in conjugation with the carbonyl group instead of participating in the conjugation with the

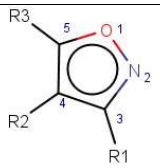
thiazole ring. On the other hand, this was not feasible when compounds had an aromatic amide at position-2 which makes the thiazole ring active toward aromatase. In such a case, the stereochemistry of the aryl group is such that it is co-planar with the carbonyl group, which facilitates the conjugation of the π-electron cloud of the aryl group with the carbonyl group. Simultaneously, the lone pair of nitrogen could therefore be available for participation in conjugation with the thiazole ring leading to its agonist activity. Compounds such as Tiotidine, Famotidine having attachment through sp2 nitrogen at position-2 could also not involve their electron pair present on nitrogen due to its higher electronegativity than compounds bearing sp3 nitrogen atom at position-2. **These observations again consolidated that a better EDG or type of the nitrogen substitution at position-2 could make the 1,3-thaizole ring an active agonist toward CYP19A1**. The requirement of planarity could also be exploited to explain the inactive compound namely, Linitript, though it was having the active structural alert SA1.3 (see Table 4.2). The ortho-chloroaryl group in the Linitript can hinder the planarity and hence delocalisation of electrons also which caused the inactiveness toward CYP19A1. In a nutshell, SAR analysis of Table 4.2 indicated that **the presence of a group (more likely EDG) that could provide π-electron density to the molecule was the deciding factor for a compound to show activity.**

Interestingly, compounds Cefodizime and Arotinolol (see Table 4.2) displayed the role of the sulphur (S) at position-2 of the 1,3-thiazole ring which made the compound active and increased the activity concentration (reducing the agonist activity). The electronegativity of sulphur (S=2.5) and carbon (C=2.5) is very similar but, interestingly, the compounds having substitution with sulphur were more active than sp^3 carbon (C) substitution. The distinguishing factor between the sulphur and carbon (sp^3) is that sulphur has a lone pair of electrons that can enable the change in the π-electron density of the 1,3-thiazole.

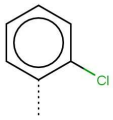
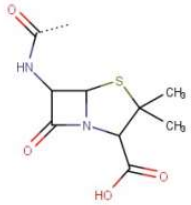
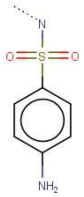
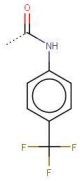
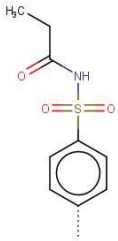
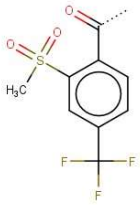
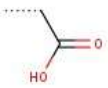
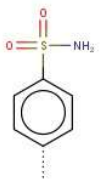
4.2.1.3 Isoxazole scaffold

The isoxazole scaffold was directly obtained from the structural alert SA1.11, most of the compounds bearing the isoxazole ring were inactive. Two special cases have been found to be active agonists, namely Leflunomide and Parecoxib (see Table 4.3) with high agonist activity (1.80µM) and low activity (26.60 µM), respectively. **The substitution on position-4 of the isoxazole could be an influencing factor for its activity**. Leflunomide has methyl (-CH_3) group at position-3 and aromatic amide (-CONAr) at position-4, which indicated the effect of the electron withdrawing amide group at position-4 of the isoxazole ring. It must be noted that the compound Isoxaflutole had electron withdrawing carbonyl aromatic substituent at position-4 and Cloxacillin had aliphatic amide (-$CONR_{Aliphatic}$) group, but both compounds were inactive which suggested the role of nitrogen atom in the aromatic amide group (-CONR).

Table 4.3: Drugs/chemicals containing isoxazole fragment and their activity towards aromatase.



CAS	Name	R1 and its description	R2 and its description	R3 and its description	Activity (µM)	Structural alert

642-78-4	Cloxacillin sodium			CH_3	Inactive	1.2, 1.5, 1.9, 1.11
127-69-5	Sulfisoxazole	CH_3	CH_3		Inactive	1.7, 1.11
75706-12-6	Leflunomide	CH_3		-	Agonist (1.80)	1.5, 1.11
198470-85-8	Parecoxib sodium			CH_3	Agonist (26.60)	1.5, 1.11
141112-29-0	Isoxaflutole	-			Inactive	1.11
78967-07-4	Mofezolac				Inactive	1.11
181695-72-7	Valdecoxib			CH_3	Inactive	1.11

4.2.2 SAR analysis of drugs/chemicals containing triazole fragments

4.2.2.1 1,2,4-triazole scaffolds

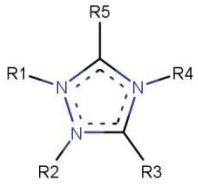
Several 1,2,4-triazole scaffolds were manually generated, those were mainly related to the presence of the structural alerts 3.3 and 3.4. The structural alert SA3.3 represented the

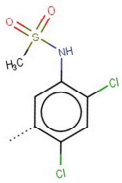
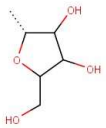
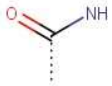
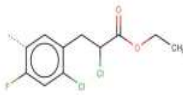
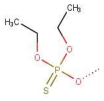
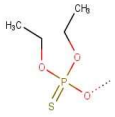
substitution on the 1,2,4-triazole ring having alkyl group at position-3 and 4. Using SA3.3, two different scaffolds were identified as shown in Table 4.5 and Table 4.4, the generic 1,2,4-traizole scaffold and the fused diazepine scaffold with 1,2,4-traizole. SA3.4 led to the N-ethyl-1,2,4-triazole scaffold identification. The following section presents the SAR analysis of each scaffold.

4.2.2.1.1 1,2,4-traizole generic scaffold

The **1,2,4-traizole** scaffold derived from the SA3.3 was substituted by the 1,2,4 triazole ring having R1, R2, R3, R4, R5 substituents as presented in Table 4.4. As there are only five examples which bear this scaffold, out of which one was an active agonist and one active antagonist, the structure activity relationship could not be established in this case. Nevertheless, it was observed that the presence of substituents like sulfonimide or phosphorothioate could orient the compounds toward inactivity irrespective of the fact whether they were present in the substituent or directly attached to the ring.

Table 4.4: Drugs/chemicals containing 1,2,4-traizole fragment and their activity toward aromatase.

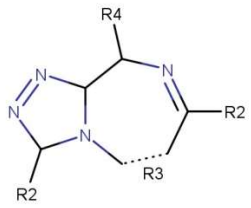


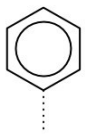
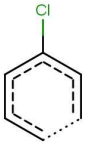
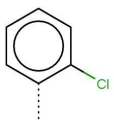
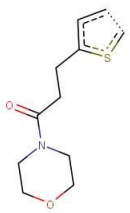
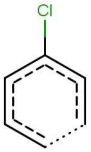
CAS	Chemical Name	R1	R2	R3	R4	R5	Activity (µM)	Structural Alert
122836-35-5	Sulfentrazone		-	CH_3		=O	Inactive	3.8
36791-04-5	Ribavirin		-		-	H	Agonist (2.56)	3.1, 3.6
128639-02-1	Carfentrazone-ethyl		-	CH_3		=O	Antagonist (54.48)	3.1, 3.6
24017-47-8	Triazophos		-		-	H	Inactive	3.6
42509-80-8	Isazofos		-		-	Cl	Inactive	3.6

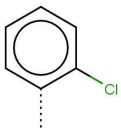
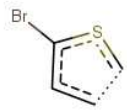
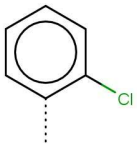
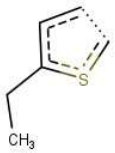
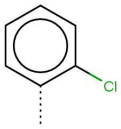
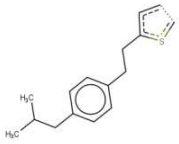
4.2.2.1.2 Fused diazepine scaffold with 1,2,4-traizole

The scaffold presented in Table 4.5 was a 1,2,4-triazole ring with a fused seven-membered azepine ring at position-3 and 4 of the triazole ring. This scaffold had four R1, R2, R3, R4 substituents, where substituent R3 represented another fused aromatic ring like thiophene, and benzene ring derivatives. The importance of the substitution at position-R1 could be observed by comparing Estazolam with Alprazolam, where Estazolam was inactive while Alprazolam was an active agonist (EC_{50} = 4.22 µM). Both compounds were similar with respect to the substituents at R2, R3 and R4, but differ in the substitution at position-R1. Replacement of the hydrogen in Estazolam by the electron donating methyl (CH_3) group (as in Alprazolam) could induce the activity. Bortizolam (IC = 11.88 µM) and Etizolam (IC = 23.71 µM) were active agonists having bromo-thiophene and ethyl-thiophene substituents respectively at position-R3. Higher agonist activity in Alprazolam as compared to the Bortizolam and Etizolam might be due to the presence of benzene ring which has more π-electron density than thiophene ring present in the later cases. Israpafant was the only compound showing antagonist activity in this category. The antagonist activity of the Israpafant could be due to the substitution at R4-position with methyl group and aromatic residue attached to the fused thiophene ring alkyl chain bridge.

Table 4.5: Drugs/chemicals containing "fused diazepine moiety with 1,2,4-traizole" fragment and their activity toward aromatase.



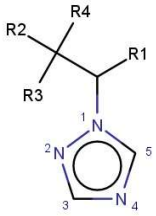
CAS	Chemical Name	R1	R2	R3	R4	Activity (µM)	Structural Alert
29975-16-4	Estazolam	H			H	Inactive	No Alert
105219-56-5	Apafant	CH_3			H	Inactive	3.3, 3.6, 3.7
28981-97-7	Alprazolam	CH_3			H	Agonist (4.22)	3.3

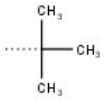
57801-81-7	Brotizolam	CH_3			H	Agonist (11.88)	3.3
40054-69-1	Etizolam	CH_3			H	Agonist (23.71)	3.3
117279-73-9	Israpafant	CH_3			CH_3	Antagonist (23.71)	3.3

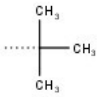
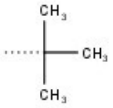
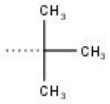
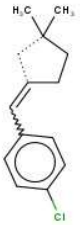
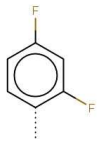
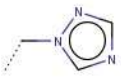
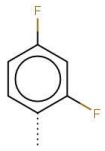
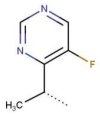
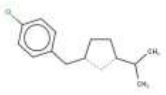
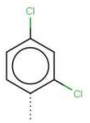
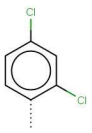
4.2.2.1.3 N-ethyl-1,2,4-triazole scaffold

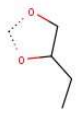
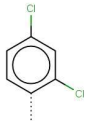
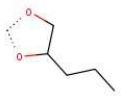
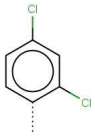
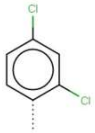
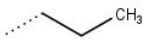
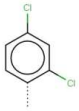
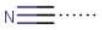
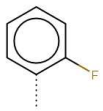
Fragment, namely SA3.4 (N-isopentyl-1,2,4-triazole), was an active fragment modified to scaffold namely (N-ethyl-1,2,4-triazole) to observe the structure activity relationships in Table 4.6. Carbons of the ethyl chain in (N-(ethyl)-1,2,4 triazole) had been labelled as α (1st carbon from nitrogen of -1,2,4-triazole) and β carbon (2nd carbon from nitrogen of -1,2,4-triazole) where R1 was situated at α carbon while R2, R3 and R4 are situated at β carbon to form the asymmetric centre. However, some chemicals such as Itraconazole, Azaconazole, Etaconazole, and Propiconazole were forming a five-membered cyclic ring on β carbon involving substituents R2 and R3 while the chemical, namely Triadimefon, had carbonyl group (=O) at the β carbon (sp^2 carbon). Most of the chemicals in Table 4.6, except stated above, had the distinguishable aliphatic substituents, aromatic substituents, and hydrogen atom at R1, R2, R3 and R4 positions.

Table 4.6: Drugs/chemicals containing N-ethyl-1,2,4-triazole fragment and their activity towards aromatase.



CAS	Chemical Name	R1	R2	R3	R4	Activity (μM)	Structural Alert
43121-43-3	Triadimefon		=O		-	Antagonist (54.43)	3.4, 3.6

55219-65-3	Triadimenol		OH		H	Antagonist (25.71)	3.4, 3.6
76738-62-0	Paclobutrazol		OH		H	Inactive	3.4, 3.6
107534-96-3	Tebuconazole	H	OH			Antagonist (33.41)	3.4, 3.6
131983-72-7	Triticonazole	H	OH		-	Inactive	3.4,3.6
86386-73-4	Fluconazole	H	OH			Inactive	3.6
137234-62-9	Voriconazole	H	OH			Inactive	3.6
125225-28-7	Ipconazole	H	OH		-	Antagonist (3.96)	
84625-61-6	Itraconazole	H		-		Inactive	3.1, 3.6, 3.7
60207-31-0	Azaconazole	H		-		Inactive	3.6

60207-93-4	Etaconazole	H		-		Antagonist (26.28)	3.1, 3.6
60207-90-1	Propiconazole	H		-		Antagonist (31.20)	3.1,3.6
112281-77-3	Tetraconazole	H	H			Antagonist (7.36)	3.1, 3.6
66246-88-6	Penconazole	H	H			Antagonist (14.22)	No Alert
114369-43-6	Fenbuconazole	H				Antagonist (33.08)	3.2
88671-89-0	Myclobutanil	H				Antagonist (5.02)	3.2
76674-21-0	Flutriafol	H	OH			Antagonist (48.56)	3.6

In Table 4.6, compound Triadmefon was an active antagonist (IC_{50} = 54.43) having para-chloro phenoxy (R1) and tertiary butyl (R3) as substituents. Interestingly, Triadmenol, the reduced form of Triadmefon, had two-fold increased activity (IC_{50}=25.71) with similar kind of substituents at R1 and R3. This observation showed **that although the role of substituents towards activity cannot be predicted, enhancement in the activity can be achieved by reducing electron withdrawing keto group (>C=O) at the β-carbon. It was interesting to note that compound Paclobutrazole, analogous to Triadmenol, was inactive towards aromatase. The only difference in the Paclobutrazole and Triadmenol was the R1 substituent present at the α-carbon. It was para-chloro alkyl benzene in the Paclobutrazole while para-chloro phenoxy in the Triadmenol. Therefore, the bonding through the C-atom at the α-carbon can turn the compound inactive. Interestingly, if para-chloro alkyl benzene was present at the β-carbon in place of α-carbon, the compound turns to an active antagonist** as in the case of Tebuconazole (another analogous compound of Triadmenol).

The rest of the compounds presented in Table 4.6 had hydrogen only at the α-carbon, indicating the role of substituents (R2, R3 and R4,) present at the β-carbon toward

activity. Substituents R2, R3 and R4, in general, offer the stereogenic position in all compounds except Triadimefon having keto group at β-carbon. Position-R2 was mainly occupied with dioxolane group, hydroxyl (-OH) and nitrile (-CN) while having dichlorophenyl group at R4-position in most of the compounds. All compounds bearing mono/di-chloro phenyl group were active antagonists except Itraconazole and Azaconazole. Compounds Itraconazole and Azaconazole had the dioxolane group also at position-R2 besides having dichlorophenyl at position-R4. Inactivity in these two compounds may be attributed due to the moderate effect of electron withdrawing inductive effect (-I effect) of the two oxygen atoms attached directly to the stereogenic β-carbon. **It was interesting to note that although the compound Azaconazole bearing dioxolane group was inactive, introducing alkyl group in the dioxolane moiety could make the compound an active antagonist** as in the case of Etaconazole and Propiconazole. **These results clearly indicated that the effect of electron withdrawing dioxolane group was moderate enough and could be compensated by introducing electron donating alkyl group to the dioxolane ring. Compounds bearing –OH and –CN group at position-R2 were associated low as well as high inhibition concentration.** Chemicals having hydroxyl (-OH) and nitrile (-CN) at R2-position with the highest activity (lowest inhibition concentration) were Ipiconazole and Myclobutanil with comparatively similar inhibition concentration (IC_{50}) as 3.96 µM and 5.02 µM, respectively. On the other hand, those requiring high inhibition concentrations were Fenbuconazole (-CN at R2) and Flutriafol (-OH at R2). **These results indicated that the inhibitor variant must include major contribution from groups present at position-R3**. For example, Myclobutanil, Penconazole and Fenbuconazole bearing n-butyl, n-propyl and ethyl derivative showed inhibition concentration values as 5.02, 14.22 and 33.08, respectively. **In general, it can be concluded that compounds bearing a better electron donating alkyl group at position-R3 showed higher activity irrespective of the presence of other groups**. Based on these findings, the trend of activity of a different kind of substituents at the β-carbon of N-ethyl-1,2,4-triazoles was suggested and is presented in Figure 4.1.

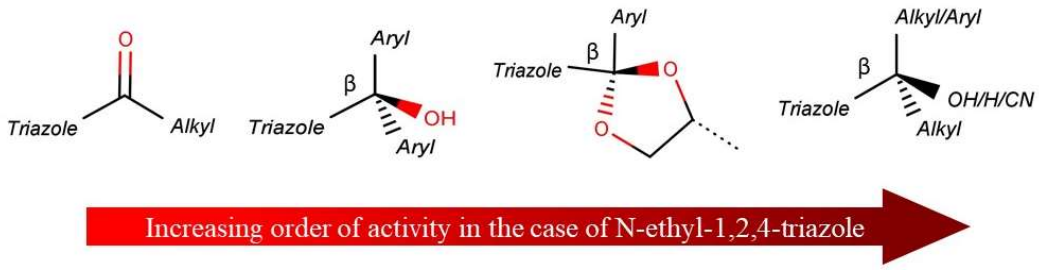


Figure 4.1: Observed trend of activity due to substituents present at the β-carbon of N-ethyl-1,2,4-triazoles.

4.2.3 SAR analysis of drugs/chemicals containing diazole fragments

Diazole structural alerts were characterized by the presence of five-member diazole rings containing diverse substitutions (see Figure 4.2). Two types of diazole fragments were present considering the position of the two nitrogen atoms in the ring; i) 1,2-diazole known as pyrazole where two nitrogen atoms are at adjacent positions, and ii) 1,3-diazole known

as imidazole having two nitrogen atoms present at alternate positions. Figure 4.2 represents the different structural alerts containing the diazole ring, which had an important contribution to the SAR analysis.

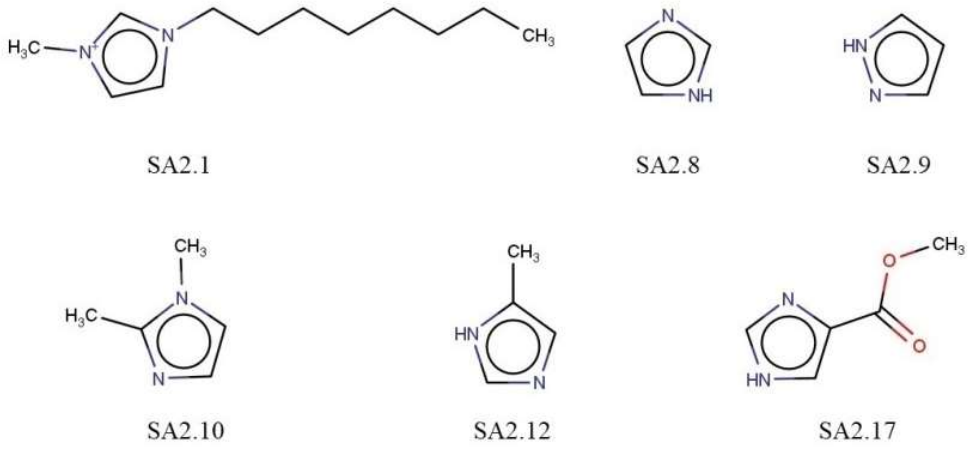


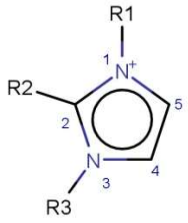
Figure 4.2: Fragments associated to the diazole ring.

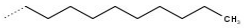
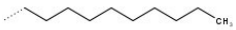
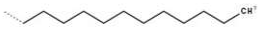
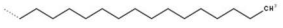
The data set was manually visualized to check the fragments in the chemicals. After careful analysis, 1,3-diazole (imidazole) and benzimidazole ring-based similar compounds were identified to compare the activity or inactivity with respect to the different stereochemical environment.

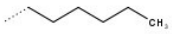
4.2.3.1 Imidazolium cation scaffolds

The imidazolium cation scaffold was obtained directly from SA2.1. Table 4.7 listed all derivatives of the imidazolium ring containing alkyl groups on position-1 and 3 of the diazole ring. Position-1 and 3 were chemically equivalent until there was no substitution, however, after substitution, these can differ due to stabilization of the positive charge on the imidazolium ring. It was observed in Table 4.7 that the substitution with benzyl group and hexyl group in 3-methylimidazolium at position-1 made the imidazolium compounds inactive. This observation indicated that **imidazolium salt chemicals having a chain length less than eight carbon atoms were not enough in size to be active.** It must be noted that the 1-hexyl-3-methylimidazolium and 1-benzyl-3-methylimidazolium were not identified within the chemical sub-group of fragments SA2.1, but were manually identified and used to support the analysis.

Table 4.7: Drugs/chemicals having the imidazolium cation fragment and their activity toward aromatase.



CAS	Chemical Name (grouped by cation)	R1	R2	R3	Activity (µM)	Structural Alert
70862-65-6	1,3-Didecyl-2-methylimidazolium		CH_3		Antagonist (3.58)	2.1, 2.10
171058-18-7, 362043-46-7, 412009-62-2	1-Decyl-3-methylimidazolium	CH_3	H		Antagonist (13.37)	2.1
219947-93-0, 244193-59-7, 404001-52-1	1-Dodecyl-3-methylimidazolium	CH_3	H		Antagonist (3.56)	2.1
61546-01-8, 244193-64-4	1-Hexadecyl-3-methylimidazolium	CH_3	H		Antagonist (7.01)	2.1
219947-96-3	1-Methyl-3-octadecylimidazolium	CH_3	H		Antagonist (3.16)	2.1
64697-40-1, 304680-36-2, 244193-52-0, 403842-84-2	1-Methyl-3-octylimidazolium	CH_3	H		Antagonist (42.26)	2.1

171058-21-2	1-Methyl-3-tetradecylimidazolium	CH_3	H		Antagonist (4.45)	2.1
36443-80-8	1-Benzyl-3-methylimidazolium	CH_3	H		Inactive	2.7
171058-17-6, 304680-35-1, 244193-50-8, 460345-16-8	1-Hexyl-3-methylimidazolium	CH_3	H		Inactive	No Alert

From Table 4.7 it was detected that the counter ion present in such compounds has a significant influence on the activity, e.g. the *active* lowest sized cation 1-methyl-3-octylimidazolium had different activities with different anions, namely chloride (Cl-), hexafluorophosphate, tetrafluoroborate and trifluoromethanesulfonate were displaying 54.09 µM, 55.03 µM, 34.34 µM and 27.52 µM inhibition concentration, respectively. In contrast, the activity of the large size imidazolium cations was found to be equally active irrespective of the nature of the counter anion attached, e.g. the largest sized cation 1-methyl-3-dodecylimidazolium had the following activities 3.06 µM, 3.11 µM and 3.85 µM for hexafluorophosphate, triflurophosphate, and tetraflurophosphate anions, respectively. **The inhibition activity of the alkyl imidazolium cations was observed to increase (decrease in inhibition concentration) with an increase in alkyl chain.**

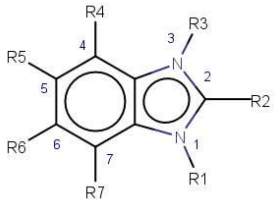
4.2.3.2 Benzimidazole scaffolds

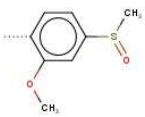
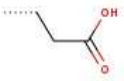
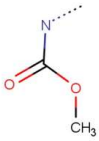
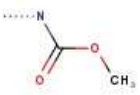
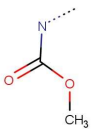
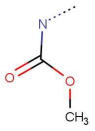
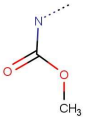
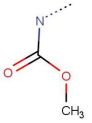
Benzimidazole scaffolds are the result of chemical modifications to the structural alert SA2.8. These scaffolds were recognized after the observation of different chemicals classes among the 54 chemicals that contained the SA2.8. It also included chemicals from the sub-group benzimidazole and xanthine derivatives.

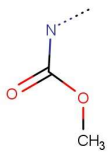
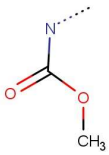
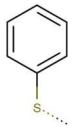
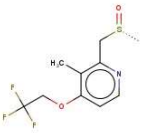
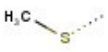
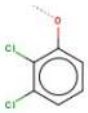
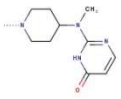
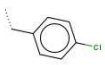
4.2.3.2.1 Benzimidazole generic scaffold

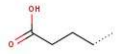
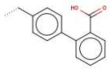
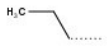
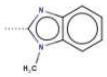
As shown in Table 4.8, benzimidazole derivatives had seven substitution positions, namely 1,2,3,4,5,6,7. The basic skeleton, where all positions were occupied by hydrogen (H), e.g., benzimidazole itself, was inactive toward aromatase. In other compounds, position-3, 4, 7 were occupied with hydrogen (H), with at least one substituent at any position (1,2,5,6) other than H, which indicated that activity towards aromatase as inactive, agonist or antagonist depends on these positions. The position-1 of the benzimidazole derivatives was mainly occupied by the hydrogen (H), benzene derivatives and methyl as substituent (see Table 4.8). **All compounds where position-R1 was occupied by an atom other than hydrogen were inactive chemicals. However, if position-R1 is H, compound could be inactive, agonist or antagonist. This observation indicated that the activity toward aromatase explicitly depended on substituents present at position-R2, R5 and R6.**

Table 4.8: Drugs/chemicals having the benzimidazole fragment and their activity toward aromatase.



CAS	Chemical Name	R1	R2	R3	R4	R5	R6	R7	Activity (μM)	SA
73384-60-8	Sulmazole	H		H	H	H	H	H	Inactive	2.8, 2.14, 2.20
23249-97-0	Procodazole	H		H	H	H	H	H	Inactive	2.8, 2.13, 2.20
53716-50-0	Oxfendazole	H		H	H		H	H	Inactive	2.5, 2.8, 2.14, 2.20
10605-21-7	Carbendazim	H		H	H	H	H	H	Agonist (43.28)	2.5, 2.8, 2.20
94-52-0	6-Nitrobenzimidazole	H	H	H	H		H	H	Inactive	2.8
31431-39-7	Mebendazole	H		H	H		H	H	Inactive	2.5, 2.7, 2.8, 2.20
31430-18-9	Nocodazole	H		H	H		H	H	Agonist (2.51)	2.5, 2.7, 2.8, 2.20
20559-55-1	Oxibendazole	H		H	H		H	H	Inactive	2.5, 2.8, 2.20
14255-87-9	Parbendazole	H		H	H		H	H	Agonist (2.18)	2.5, 2.7, 2.8, 2.20

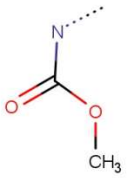
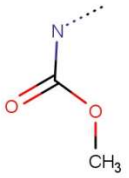
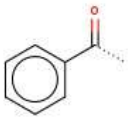
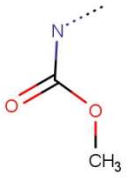
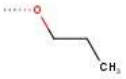
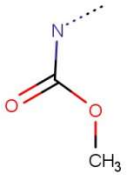
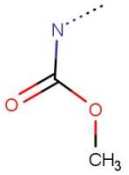
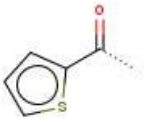
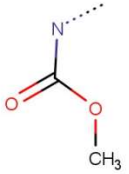
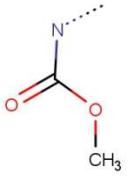
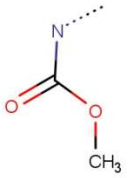
54965-21-8	Albendazole	H		H	H		H	H	Agonist (1.18)	2.2, 2.5, 2.8, 2.20
43210-67-9	Fenbendazole	H		H	H		H	H	Agonist (1.05)	2.5, 2.8, 2.20
26097-80-3	Cambendazole	H		H	H		H	H	Agonist (3.48)	2.5, 2.8, 2.20
148-79-8	Thiabendazole	H		H	H	H	H	H	Agonist (5.01)	2.8
103577-45-3	Lansoprazole	H		H	H	H	H	H	Agonist (12.35)	2.8, 2.20
145858-50-0	Liarozole	H	H	H	H	H		H	Antagonist (0.51)	2.4, 2.7, 2.8
68786-66-3	Triclabendazole	H		H	H	Cl		H	Antagonist (23.71)	2.4, 2.8
2818-69-1	5-Chloro-2-methyl-1H-Benzimidazole	H	CH_3	H	H	H	Cl	H	Inactive	2.4, 2.8
934-32-7	2-Aminobenzimidazole	H	NH_2	H	H	H	H	H	Inactive	2.8
108612-45-9	Mizolastine			H	H	H	H	H	Inactive	2.7, 2.15, 2.18
1163-36-6	Clemizole			H	H	H	H	H	Inactive	2.4, 2.7, 2.10, 2.18
3689-76-7	Chlormidazole		CH_3	H	H	H	H	H	Inactive	2.4, 2.7, 2.10

16506-27-7	Bendamustine	CH_3		H	H		H	H	Inactive	2.10, 2.13, 2.18, 2.20
144701-48-4	Telmisartan			H	H	H		H	Inactive	2.7, 2.10, 2.20
51-17-2	Benzimidazole	H	H	H	H	H	H	H	Inactive	2.8

In most of the active benzimidazole compounds, position-R2 was occupied by methyl carbamate (-NHCOOCH$_3$) which was also found as an active fragment (SA2.5) from the diazole subclass. There was also a compound having isopropyl carbamate at position-R5 of the benzimidazole ring (see Table 4.8). It was interesting to observe that benzimidazole is inactive but its derivative containing methyl carbamate at position-R2 of benzimidazole (Carbendazim, IC = 43.20µM) was an active agonist compound, suggesting the crucial role of the methyl carbamate in the benzimidazole scaffold for aromatase activity. However, substitution of other different groups at position-R5 of the benzimidazole containing methyl carbamate at position-R2 could turn the activity of the resultant compounds in any way. Compounds Oxfendazole and Mebendazole, for example, having sulfinyl benzene (-SOPh) and benzoyl (-COPh) group attached at position-R5, were inactive despite carrying methyl carbamate at position-R2. **This observation concluded that the presence of electron withdrawing/releasing group at position-R5 could affect the activity.** This was supported by the fact that compound 6-nitrobenzimidazole having electron withdrawing nitro group (-NO_2) at position-R5 was also inactive. Although Nocodazole, analogous compound of Oxfendazole and Mebendazole, also had electron withdrawing thiophene-2-carbonyl group, its activity towards aromatase suggested the crucial role of the thiophene ring in making the Nocodazole active agonist. Compounds such as Parbendazole and Albendazole were active agonists with n-butyl (n-Bu) and thio-propyl (-SPr) group respectively, suggesting the importance of positive inductive effect towards activity. However, their analogous compound, namely Oxibendazole, bearing propoxy (-OPr) group was inactive. In this case, the oxygen, being an electronegative atom, might not have allowed its electron pair to easily participate in conjugation with the π-electron density of benzene ring of the benzimidazole. Role of position-R5 of the benzimidazole had been summarized in Table 4.9. These observations depicted the role of substituents in position-R5 when position-R2 is occupied with methyl carbamate in benzimidazole. Negative inductive electron withdrawing groups (-IEWG) moved toward inactivity as in Mebendazole and Oxibendazole, while the positive inductive electron releasing groups (+IERG) contributed toward the agonist activity as in Nocodazole, Parbendazole, Albendazole, and Fenbendazole.

Table 4.9: Role of position-R5 in the benzimidazole towards aromatase activity.

Compound	Position-R2	Position-R5	Activity	Remark

Oxfendazole			Inactive	Negative inductive effect
Mebendazole			Inactive	Negative inductive effect
Oxibendazole			Inactive	Oxygen electronegative atom
Carbendazim		H	Agonist (43.28μM)	Least active
Nocodazole			Agonist (2.51μM)	Role of thiophene ring
Parbendazole			Agonist (2.18μM)	Positive inductive effect
Albendazole			Agonist (1.18μM)	Positive inductive effect
Fenbendazole			Agonist (1.05μM)	Positive inductive effect

The linkage with sulphur atom at position-R5 could also tune the activity by lowering the effective concentration (EC_{50}) values. For example, Albendazole (1.18 μM), and Fenbendazole (1.05μM) having linkage through S-atom showed higher activity than Nocodazole (2.51μM) and Parbendazole (2.18μM) that were joined through C-atom.

Another compound, namely Cambendazole (3.48µM), showed high activity as compared to Thiabendazole (5.01 µM). Both these compounds have thiazole moiety at position-R2. The only difference between these two was that Cambendazole also had isopropyl carbamate at position-5 along with thiazole at position-R2. **This observation indicated that the carbamate group tends to contribute toward the agonist activity irrespective of the fact whether it has occupied position-2 or 5 of the benzimidazole. It was also observed that if carbamate group was present at position-2, then position-5 should be occupied by the appropriate chemical group** (probably EDG). There was another compound, namely Lansoprazole, which is an active agonist (12.43µM); it had the substituent 4-(2,2,2-trifluoroethoxy)-3-methyl-2-(methylsulfinyl) pyridine with linkage through sulfinyl sulphur atom at position-R2 of the benzimidazole and was more active than the methyl carbamate substituent in Carbendazim (43.28µM).

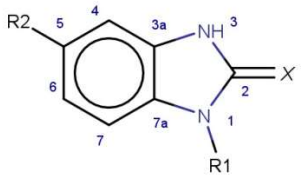
It was interesting to see that there were also two compounds with antagonist activity. Liarazole and Tricalbendazole were active antagonists having IC_{50} values 0.51 and 23.71 µM, respectively. **The antagonist activity in these compounds reflected the role of substitution at the position-R6 of the benzimidazole.** In Liarazole, position-R6 was occupied by 1-(3-chlorophenyl)-1-(1,3-diazole)methyl whereas in **Tricalbendazole**, it was occupied by 2,3-dichlorophenoxy beside methyl sulphide (-SCH_3) at the position-R2 of the benzimidazole scaffold. **This observation indicated that occupation of position-R6 of the benzimidazole could turn the activity of the benzimidazole scaffold to antagonist activity.** It is worth mentioning that Liarozole (0.51µM) had two different azole rings, and showed higher antagonist activity than Tricalbendazole (23.71 µM) having one azole ring. It must be noted that the substitution at position-R6 could not be the sole criterion of the activeness, as Telmisartan was inactive in spite of having substitution at position-6 (see Table 4.8). To exhibit any kind of activity, the chemical should have also been accompanied by appropriate substitutions at other positions of benzimidazole. Nevertheless, as earlier mentioned that all chemicals having any substitution at position-1 were inactive, so did Telmisartan. It could be due to the loss of planarity and flexibility around the nitrogen atom to provide the electronic conjugation in the benzimidazole skeleton. Chemical groups such as methyl (-CH_3), amine (-NH_2) and their different derivative at position-R2 of the benzimidazole listed in Table 4.8 were not effective to interact with the aromatase CYP19A1.

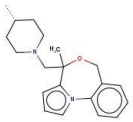
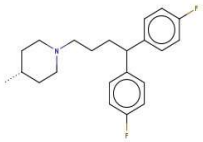
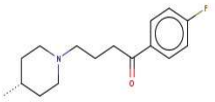
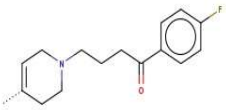
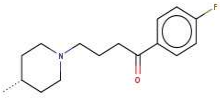
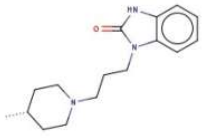
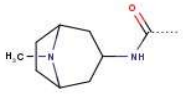
4.2.3.2.2 2-substituted benzimidazole scaffold

Table 4.10 represented another scaffold of the benzimidazole having doubly bonded oxygen or sulphur atoms at position-2. Position-1 was occupied by various kinds of the substituents. Most of the compounds of this category were inactive except Zeldaride (23.71µM) and Timiperone (18.12) which were active antagonists. Timiperone had the sulphur atom (=S) at position-2 and position-1 was occupied by 1-(4-fluorophenyl)-4-(4'-ylpiperidin-1-yl)butan-1-one. Its analogous compound Benperidol was inactive having the same substituent at position-1 but differing in substitution at position-2, which was occupied by the oxygen atom. This observation highlighted the role of the sulphur atom at position-2 in Timiperone for the antagonist activity. The only active antagonist compound which had oxygen atom at position-2 was Zeldaride which could be due to the planar substituent at position-1. The substituent at position-1 of the Zeldaride had two aromatic rings (pyrrole residue linked with the benzene residue) connected by the single bond and the puckering of these rings was provided by the seven-member aliphatic cyclic ring which could be suitable enough to provide the specific surface area and greater electron density to interact with the aromatase enzyme. Interestingly, Pimozide, analogous

compound of the Zeldaride, was inactive in spite of having two aromatic (benzene rings) residues at position-1 but the chemical structure was lacking planarity.

Table 4.10: Drugs/chemicals having the 2-substituted benzimidazole fragment and their activity toward aromatase.



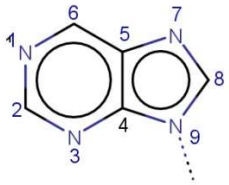
CAS	Chemical Name	1	2	3	4	5	6	7	Activity (µM)	SA
		R1	X			R2				
109826-26-8	Zaldaride		O	H	H	H	H	H	Antagonist (29.85)	2.7, 2.8, 2.18, 2.20
2062-78-4	Pimozide		O	H	H	H	H	H	Inactive	2.7, 2.8, 2.18
2062-84-2	Benperidol		O	H	H	H	H	H	Inactive	2.7, 2.8, 2.18
548-73-2	Droperidol		O	H	H	H	H	H	Inactive	2.7, 2.8, 2.18
57648-21-2	Timiperone		S	H	H	H	H	H	Antagonist (18.12)	2.7, 2.8, 2.18
57808-66-9	Domperidone		O	H	H	Cl	H	H	Inactive	2.4, 2.8, 2.18
123258-84-4	Itasetron		O	H	H	H	H	H	Inactive	2.8, 2.18

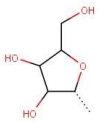
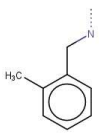
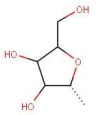
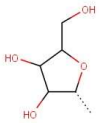
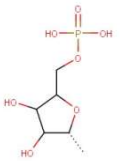
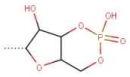
4.2.3.3 Purine scaffold

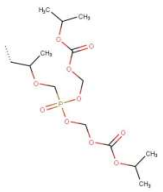
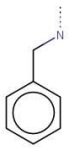
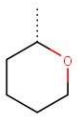
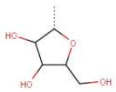
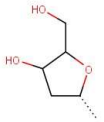
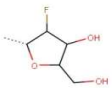
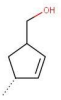
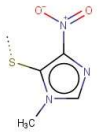
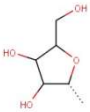
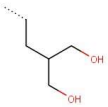
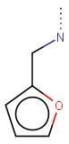
The fragment SA2.19 was modified to 9*H*-purine by the substitution at different positions as listed in Table 4.11 to obtain the purine scaffold and to investigate the influence on aromatase CYP19A1 activity. Alhough total substitution positions in the 9*H*-purine

scaffold were nine, the inactive agonists and antagonist chemicals were derivatives of the substitutions at position-2, 6 and 9 in the data set. Position-2 had substitutions with small atoms or a group of atoms such as fluorine, chlorine, and amine. Compound with fluorine (Fludarabine) is inactive, compounds bearing chlorine, namely Cladribine and Clofarabine, were inactive and active antagonist respectively, and compounds having the amine group, namely 6-O-Methylguanine, Abacavir, Penciclovir, and 6-Thioguanine were inactive, active agonist, inactive and active antagonist, respectively. These results stimulated us to investigate and compare the role of substitutions at position-6 and 9 of the purine derivatives. Position-9 of purine derivatives was mainly occupied by the furanose. A clear perusal of the position-9 in Table 4.11 indicated that the presence of furanose had no or little contribution towards the activity. Therefore, **we concluded that substituents at position-6 were solely affecting the activity of purine derivatives towards aromatase.**

Table 4.11: Drugs/chemicals having the purine fragment and their activity toward aromatase.



CAS	Chemical Name	1	2	3	6	7	8	9	Activity (µM)	Structural Alert
20535-83-5	6-O-Methylguanine	-	NH_2	-		-	H	H	Inactive	2.8, 2.20
58-63-9	Inosine	-	H	-	OH	-	H		Inactive	2.6, 2.20
73-24-5	Adenine	-	H	-	NH_2	-	H	H	Agonist (5.52)	2.8,
23707-33-7	Metrifudil	-	H	-		-	H		Antagonist (5.96)	2.6, 2.7, 2.19, 2.20
58-61-7	Adenosine	-	H	-	NH_2	-	H		Inactive	2.6, 2.19, 2.20
61-19-8	Adenosine-5-phosphate	-	H	-	NH_2	-	H		Inactive	2.6, 2.19, 2.20
60-92-4	cAMP	-	H	-	NH_2	-	H		Inactive	2.6, 2.19, 2.20

202138-50-9	Tenofovir disoproxil fumarate	-	H	-	NH_2	-	H		Inactive	2.19, 2.20
2312-73-4	N-Benzyl-9-(tetrahydro-2H-pyran-2-yl)adenine	-	H	-		-	H		Agonist (2.73)	2.6, 2.7, 2.20
21679-14-1	Fludarabine	-	F	-	NH_2	-	H		Inactive	2.6, 2.19,2.20
4291-63-8	Cladribine	-	Cl	-	NH_2	-	H		Inactive	2.6, 2.20
123318-82-1	Clofarabine	-	Cl	-	NH_2	-	H		Antagonist (1.58)	2.6, 2.20
4097-22-7	2',3'-Dideoxyadenosine	-	H	-	NH_2	-	H		Inactive	2.6, 2.20
136470-78-5	Abacavir	-	NH_2	-		-	H		Agonist (2.66)	2.6, 2.20
446-86-6	Azathioprine	-	H	-		-	H	H	Agonist (10.42)	2.8
574-25-4	Thioinosine	-	H	H	SH	-	H		Agonist (5.73)	2.6, 2.20
39809-25-1	Penciclovir	H	NH_2	-	=O	-	H		Inactive	2.20
50-44-2	6-Mercaptopurine	-	H	H	=S	H	H	-	Agonist (6.26)	2.8
154-42-7	6-Thioguanine	-	NH_2	H	=S	H	H	-	Antagonist (0.47)	2.8
525-79-1	Kinetin	-	H	-		H	H	-	Inactive	No alert

It can be noticed in Table 4.11 that position-6 was mainly connected to the oxygen, nitrogen and sulphur bearing substituents. The critical evaluation of these substitutions leads to some interesting consideration such as: **i) all compounds with oxygen**

substitution at position-6 were inactive toward aromatase as in 6-O-Methylguanine, Inosine and Penciclovir, ii) all outcomes such as inactive, active agonist and active antagonist were possible for the chemicals having connection through nitrogen, iii) both active outcomes as agonist (Thioinosine and 6-Mercaptopurine) and antagonist (6-Thioguanine) were possible for the chemicals that were connected through the sulphur (S) atom at position-6. These observations clearly demonstrated that some electronic factors were participating in contributing activity or inactivity towards aromatase. As the position-6 belonged to the pyrimidine ring of the purine derivative, loan pair electrons of O, N, S can involve in conjugation with π-electron density of the purine derivatives. Oxygen being the most electronegative atom as compared to N and S (order of electronegativity O > N > S), did not easily allow its electrons to participate in the conjugation. This could be the reason that all compounds bearing substitution through oxygen at position-6 were inactive towards aromatase activity. Sulphur being the least electronegative could share its electrons with the purine scaffold thereby leading to the agonist and antagonist activity. On the other hand, substitution through nitrogen showed activity in any direction depending upon the type of substituent bearing N.

Adenine, the first derivative of purine, had the amino group at position-6 and was an active agonist (5.52µM). However, Adenosine, Adenosine-5-phosphate, Fludarabine, cAMP, Tenofovirdisoproxilfumarate, Cladribine, 2,3-dibeooxyadenosine, with the amino group at position-6 and furanose at position-9, were inactive compounds. This observation indicated that the presence of furanose at position-9 might be contributing toward the inactivity of above-mentioned compounds. It is interesting to note that there were compounds such as Metrifudil (antagonist), Clofrabine (antagonist) and Abacavir (agonist) which showed activity toward aromatase despite having furanose at position-9. In the case of Metrifudil and Abacavir, position-6 was occupied by 2-methyl benzylamine and cyclopropyl amine, respectively, in place of -NH_2. Both these substituents are activating groups and can facilitate the participation of the electron pair of nitrogen in conjugation with the ring which might have resulted in activity. In the case of Clofarabine, although there was a NH_2 group at position-6 and furanose at position-9, the presence of fluorine in the 2-deoxyribose was turning the inactive compound to active antagonist compound. Hence, fluorine was playing a crucial role in providing the activity to the compound as most of compounds having furanose were inactive as described above. Inactivity of Cladribine and Fludarabine which did not have fluorine containing furanose further highlighted the role of fluorine towards activity.

Interestingly, N-Benzyl-9(tetrahydro-2H-pyran-2-yl)adenine showed agonist activity (2.73 µM) having the benzyl amine at position-6 and tetrahydropyran ring at position-9. Nevertheless, its analogous compound, namely Metrifudil, showed antagonist activity. The different activities of these two compounds could be realised by comparing substituents present at position-9 since these two compounds were almost similar in substitution at position-6 but differed in substitution at position-9. Metrifudil had tetrahydrofuran derivative at position-9 whereas N-Benzyl-9(tetrahydro-2H-pyran-2-yl)adenine contained tetrahydropyran ring at position-9. This observation pointed out that compound with the furan ring at position-9 directed antagonist activity while the presence of pyran ring at the same position made the compound an active agonist. This observation was also supported by the antagonist activity of Clofarabine having furan derivative at position-9. **In a nutshell, appropriate selection of the substituent at position-9 could make the compound an active agonist or antagonist keeping the position-6 reserve for benzyl amine derivatives. If position-6 was kept reserved for the –NH2 group,**

activity could be induced by the introduction of fluorine atom in the THF derivative at position-9.

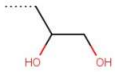
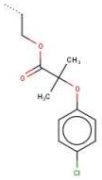
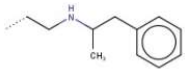
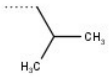
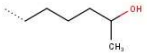
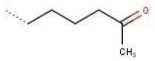
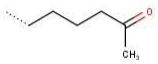
All compounds, namely Azathioprine, Thioinosine, 6-Mercaptopurine and 6-Thioguanine, having connection through the sulphur atom at position-6 were active. The role of the sulphur linkage could be understood by comparing activity of Adenosine and Thioinosine. Both had similar substitution at position-9 but differ at position-6. Adenosine, with -NH2 group at position-6, was inactive, while Thioinosine, having –SH linkage at position-6, was an active agonist. This observation indicated that the less electronegative S-atom can shape the properties of the 9*H*-purine ring in making the compounds active. **It was interesting to note that all compounds containing sulphur bonding at position-6 showed agonist activity excluding 6-Thioguanine which was an antagonist compound.** This observation pointed out that agonist and antagonist behaviour of compounds containing S-atom could depend on substituents present at position-2. The antagonist activity of compound 6-Thioguanine (0.47µM) might be due to the presence of an electron donating group (-NH_2) at position-2 besides having sulphur linkage at position-6. **Therefore, chemicals having carbon-sulphur bond (C-S) at position-6 could become an active antagonist if appropriate π-electron donor group was introduced at position-2.**

4.2.3.4 Xanthine scaffold

The structural alert SA2.11 was used to obtain the xanthine derivatives category (see Table 4.12). When chemicals were identified and extracted from the diazole dataset, it was found that most compounds were inactive. However, one compound, namely Furafylline, was a mild active agonist (26.84 µM). The xanthene scaffold had four substitution positions defined as R1, R2, R3 and R4, where most compounds had electron releasing groups as well as hydrogen bond donor groups inside residues at all positions. However, Furafylline was an exceptional case and had aromatic methyl furan ring at position-R2. From 14 chemicals in Table 4.12, 13 were automatically identified with SA2.11 (all inactive, and that is why Acc = 1 and LR = inf). However, the manual inspection of data and modification of SA2.11 found that Furafylline was not automatically recognized with SA2.11 because of a different substituent at the R3-position.

Table 4.12: Drugs/chemicals containing the xanthine fragment and their activity toward aromatase.

CAS	Name	R1	R2	R3	R4	Activity (µM)	SA
58-08-2, 69-22-7	Caffeine, Caffeine Citrated	CH3	CH3	CH3	H	Inactive	2.11

83-67-0	Theobromine	H	CH3	CH3	H	Inactive	2.11, 2.15
69975-86-6	Doxofylline	CH3	CH3		H	Inactive	2.11, 2.20
479-18-5	Dyphylline	CH3	CH3		H	Inactive	2.11, 2.20
519-37-9	Etofylline	CH3	CH3			Inactive	2.11, 2.20
54504-70-0	Etofylline clofibrate	CH3	CH3		H	Inactive	2.4, 2.11, 2.13, 2.20
1892-80-4	Fenethylline hydrochloride	CH3	CH3		H	Inactive	2.7, 2.11, 2.18
90162-60-0	Isbufylline	CH3	CH3		H	Inactive	2.11
6493-06-7	Lisofylline		CH3	CH3	H	Inactive	2.6, 2.11, 2.20
6493-05-6	Pentoxifylline		CH3	CH3	H	Inactive	2.11
55242-55-2	Propentofylline		CH3		H	Inactive	2.11
603-00-9	Proxyphylline	CH3	CH3		H	Inactive	2.11, 2.20
58-55-9	Theophylline	CH3	CH3	H	H	Inactive	2.8
80288-49-9	Furafylline	CH3		H	CH3	Agonist (29.84)	2.8

4.2.3.5 Imidazole scaffolds

Chemicals having imidazole structural alert SA2.8 had been categorized in three types of scaffolds; i) imidazole and its derivatives (see Table 4.13), ii) N-methyl imidazole and derivatives (see Table 4.14), and iii) N-ethyl imidazole and derivatives (see Table 4.15).

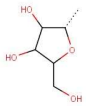
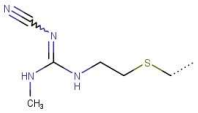
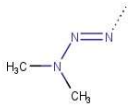
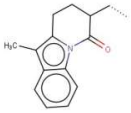
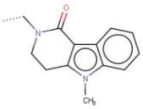
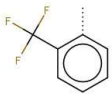
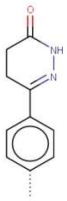
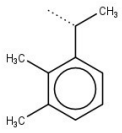
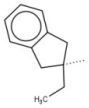
These manually extracted scaffolds presented in Table 4.13, Table 4.14, Table 4.15 were slightly different than those extracted by SARpy, therefore, few new chemicals were also recognized (e.g. 1-[2-(triflurомethyl)phenyl]-1H-Imidazole). It is worth noting that in Table 4.13, Table 4.14, Table 4.15, all compounds with EWGs like nitro, amide, etc., at positions mainly 2 or 5 of the imidazole scaffold were inactive toward the aromatase.

4.2.3.5.1 Generic imidazole scaffold

It was interesting to see that imidazole and its derivatives had hydrogen at N-1 position though they were inactive compounds; however, substitution at position-1 (N-1 position) made the imidazole scaffold active toward aromatase (see Table 4.13). For example, compounds 1-[2-(triflurомethyl)phenyl]-1H-Imidazole (IC = 6.94 µM) and Imazodan (IC = 7.22 µM) showed antagonist activity while having aryl group at position N-1. **Imidazole scaffold hence became active if hydrogen of the N-1 position is replaced by a substituent such as an aryl group. Inactivity of 4,5-diphenylimidazole beside having aryl group substituents at position-4 and 5 of the imidazole ring indicated that it was the position N-1 that required aryl group to induce activity in the imidazole scaffold.** However, Medetomidine, having alkyl aromatic substituent, was an active antagonist, but with low activity, and required more inhibition concentration (IC = 29.85 µM). **This observation showed that the appropriately 4-substituted (most likely electron donating group) imidazoles might have become active but required more inhibition concentration than those of N-substituted imidazoles.** The same trend could also be demonstrated in the N-methyl (IC = 0.87-1.38 µM, see Table 4.14) and N-ethyl (IC = 0.05-13.76 µM, see Table 4.15) substituted scaffold where all the compounds were more active than the 4-substituted Medetomidine (IC = 29.85 µM). From Table 4.13, it can be concluded that N-1-substituted imidazole derivatives contributed to the increase of the antagonist activity more than other substitutions at the imidazole ring. One compound, namely Acadesine (EC_{50} = 24.64 µM), was an active compound (agonist activity) despite having electron withdrawing amide group (-$CONH_2$) at position-4 which suggested that the activity of Acadesine could be due to the presence of other substituents. Besides having an amide group, Acadesine has also deoxyribose and amine group present at position-1 and 5, respectively. The amine group could provide electrons to increase the delocalization in the imidazole ring and deoxyribose was suitable enough to provide hydrogen bonding interactions. All other compounds without N-1 substitution were found inactive towards the aromatase.

Table 4.13: Drugs/chemicals containing the imidazole fragment and their activity toward aromatase.

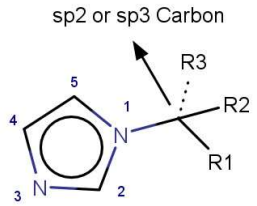
CAS	Chemical Name	1	2	3	4	5	Activity (µM)	SA

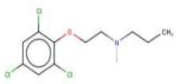
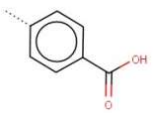
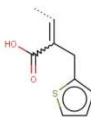
2627-69-2	Acadesine		H	-		NH_2	Agonist (24.64)	2.6, 2.20
51481-61-9	Cimetidine	H	H			CH_3	Inactive	2.2, 2.8, 2.12, 2.18
4342-03-4	Dacarbazine	H	H	-			Inactive	2.8, 2.12, 2.18
129300-27-2	Fabesetron	H	H	-		CH_3	Inactive	2.8, 2.12
122852-69-1	Alosetron	H	H	-	CH_3		Inactive	2.8, 2.12, 2.18
668-94-0	4,5-Diphenylimidazole	H	H	-			Inactive	2.8
822-36-6	4-methylimidazole	H	H	-	H	CH_3	Inactive	2.8
288-32-4	Imidazole	H	H	-	H	H	Inactive	2.8
25371-96-4	1-[2-(trifluromethyl)phenyl]-1H-Imidazole		H	-	H	H	Antagonist (6.94)	2.7
84243-58-3	Imazodan		H	-	H	H	Antagonist (7.22)	2.7
86347-15-1	Medetomidine	H	H	-		H	Antagonist (29.85)	2.7, 2.8
104054-27-5	Atipamezole	H	H	-	H		Inactive	2.7, 2.8, 2.12

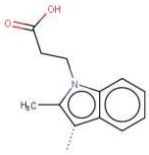
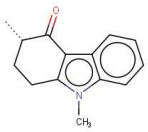
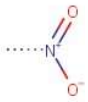
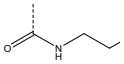
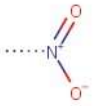
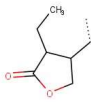
4.2.3.5.2 N-methyl imidazole

Table 4.14 represented compounds bearing N-methyl imidazole scaffold. It had three equivalent substitution positions R1, R2, and R3, on the methyl carbon and three substitution positions, namely 2, 4, 5 at the imidazole ring. The parent compound N-methyl imidazole and its derivatives did not show any kind of activity toward aromatase. Examples were 1,2-Dimethyl-5-nitroImidazole, Ipronidazole, Ronidazole and Pilocarpine. Moreover, these compounds had also EWG present at the position-5 of the imidazole ring. This observation indicated that any kind of activity in the N-methyl imidazole scaffold might be due to the substitution on the methyl carbon. It was also in agreement with the finding, as mentioned earlier, that the substitutions on imidazole ring were less relevant for the activity. There were only two compounds, namely Bifonazole (0.87 µM) and Prochloraz (1.38µM), which showed antagonist activity, while all other compounds of Table 4.14 were inactive. A meticulous inspection of Table 4.14 indicated that active antagonist compounds had two substitutions occupied out of three (R1, R2, R3). All inactive compounds in Table 4.14 had either one substituent at the methyl carbon or the EWG present in the imidazole ring. Overall, as a concluding remark for Table 4.14, **activity in the N-methyl imidazole scaffold can be induced by replacing at least two hydrogens of the methyl group with substituents like aryl group**.

Table 4.14: Drugs/chemicals having the N-methyl imidazole fragment and their activity toward aromatase.



CAS	Chemical Name	R1	R2	R3	2	3	4	5	Activity (µM)	SA
60628-96-8	Bifonazole			H	H	-	H	H	Antagonist (0.87)	2.7
67747-09-5	Prochloraz	O=		-	H	-	H	H	Antagonist (1.38)	2.3, 2.4, 2.18, 2.20
144143-96-4	Eprosartan	H		H					Inactive	2.7, 2.10, 2.13, 2.20
78712-43-3	Ozagrel	H		H	H	-	H	H	Inactive	2.7, 2.13, 2.20

76894-77-4	Dazmegrel	H		H	H	-	H	H	Inactive	2.13, 2.20
99614-02-5, 103639-04-9	Ondansetron	H		H	CH3	-	H	H	Inactive	2.10
551-92-8	1,2-Dimethyl-5-nitroimidazole	H	H	H	CH3	-	H		Inactive	2.10
14885-29-1	Ipronidazole	H	H	H			H		Inactive	2.10
7681-76-7	Ronidazole	H	H	H		-	H		Inactive	2.5, 2.10, 2.20
22668-01-5	Etandazole		H	H		-	H	H	Inactive	2.16, 2.18, 2.20
92-13-7	Pilocarpine	H	H	H	H	-	H		Inactive	2.6, 2.13, 2.20

4.2.3.5.3 N-ethyl-1,2,4 imidazole scaffold

In Table 4.15, the antagonist activity in most of the compounds showed that the introduction of ethyl chain at position N-1 of the imidazole could make interactions more effective with the aromatase. However, the parent compound Imiloxan with N-ethyl imidazole derivative itself was an inactive compound akin to N-methyl imidazole. **This observation indicated that the mere increasing of the carbon chain at the N-1 position would not make any difference in the activity.** Rather, the substituents present at the ethyl carbons were the ones accountable for the activity in the N-ethyl imidazole scaffold. It was also worth comparing the activity of compounds originating from the N-ethyl imidazole (or N-ethyl-1,3-diazole) fragment and N-ethyl-1,2,4-triazole and to see the role of the azole ring, if any. Interestingly, compounds Climbazole (from N-ethyl-1,3-diazole, Table 4.15) and Triadmefon (from N-ethyl-1,2,4-triazole, see Table 4.6) were analogous to each other with respect to the substituents present at position-R1, R2, R3 and R4, except the type of the azole ring. Climbazole with diazole ring while Triadmefon having triazole ring showed antagonist activity. Surprisingly, the activity of Climbazole

(IC_{50} = 13.33) was four times greater than the Triadmefon (IC_{50} = 54.43). Similarly, to our surprise, other compounds present in Table 4.15 showed higher activity than those present in Table 4.6. We concluded that although both N-ethyl-1,3-diazole and N-ethyl-1,2,4-triazole derivatives were active antagonists, the former was much more effective towards the aromatase. **A careful investigation of** Table 4.15 **showed that substituents R1, R2, R3 and R4 affected the activity in the same way as in the case of N-ethyl-1,2,4-triazole.**

A clear perusal of Table 4.15 showed that most the compounds were active antagonists without any substitution at position-R1 (R1 = H) of the N-ethyl imidazole scaffold. Only four compounds (out of 13), namely Climbazole, Lanoconazole, Lulizonazole and Trifumizole, were active antagonists that were having substituents at position-R1 besides other occupied positions. In addition, there were two inactive compounds such as Etomidate and Metomidate having phenyl substituent at position-R1. **These observations indicated that position-R1 has no or a minimal role in inducing the activity in the N-ethyl imidazole scaffold**. **Therefore, position-R3 and R4 were actually playing an important role towards the antagonist activity.** In nine compounds, position-R4 was mainly occupied by 2,4-dichlorophenyl group (see Table 4.15). Out of these nine compounds, eight of them were active antagonists. It should be noted that 2,4-dichlorophenyl group (SA2.3) was also found as an active fragment in the diazole rule set. At first instance, we had a quick thought that substitution with chloro-substituted phenyl derivatives at position-R4 was playing a crucial role in imparting the antagonist activity to most of the compounds. However, further investigations revealed that there were also compounds which were active (rather good active) without having chloro-substituted phenyl derivatives at position-R4. In addition to this, one compound, namely Ketoconazole, was inactive despite having dichloro phenyl substituent at position-R4. These observations led us to analyse the substituents present at position-R3 and their role towards activity. **It was observed with more lucidity that the substituents present at position-R3 were actually more effective towards aromatase activity. Though most of the active antagonist compounds had alkoxy derivatives at position-R3, the top most antagonist compound (Trifumizole, IC_{50} = 0.05) had electron donating butoxy group at the same position. This trend was quite similar to that observed in the case of N-ethyl-1,2,4-triazole derivative** (see Table 4.6). Therefore, to our belief, it was a worthwile idea indeed to compare the SAR analysis of Table 4.6 and Table 4.15 (as discussed earlier in the previous paragraph), and indeed, the trend of activity presented in Figure 4.3 was observed.

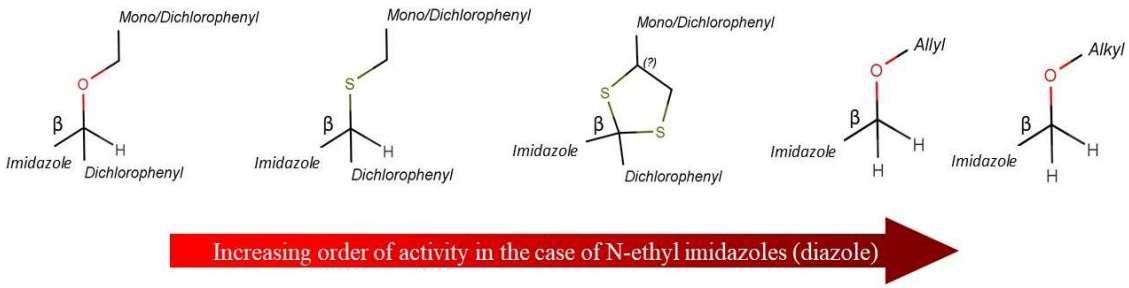
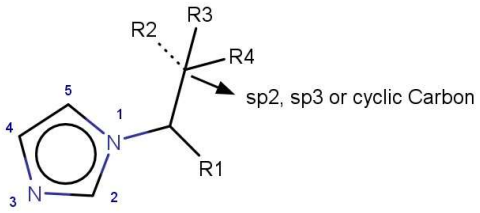
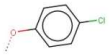
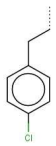
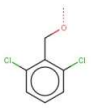
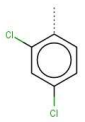
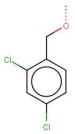
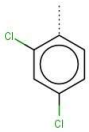
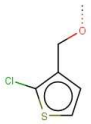
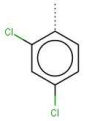
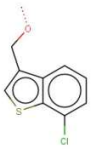
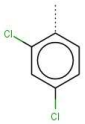
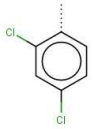


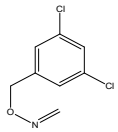
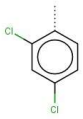
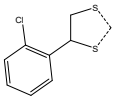
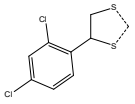
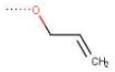
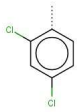
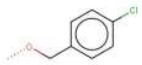
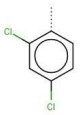
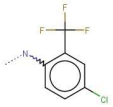
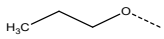
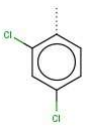
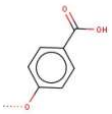
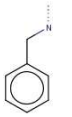
Figure 4.3: Observed trend of activity due to substituents at β-carbon of the N-ethyl imidazole derivatives.

Among inactive compounds (Benznidazole, Metronidazole, Ornidazole, Secnidazole, Etomidate and Metomidate) of Table 4.15, the presence of the electron withdrawing groups in the imidazole ring could not lead to the interaction with the aromatase, as mentioned earlier also in the description of Table 4.13 and Table 4.15. Very briefly, regarding the imidazole scaffold and the antagonist activity, Table 4.13 indicated low to moderate activity compound, Table 4.14 presented slightly high active compounds and lastly, Table 4.15 presented the fine tuning of the activity from high to moderate activity ranges.

Table 4.15: Drugs/chemicals containing the N-ethyl-1,2,4 imidazole fragment and their activity toward aromatase.



CAS	Chemical Name	R1	R 2	R3	R4	2	3	4	5	Activity (μM)	SA
38083-17-9	Climbazole		-	=O		H	-	H	H	Antagonist (13.33)	2.4, 2.20
64872-77-1	Butoconazole	H	H			H	-	H	H	Antagonist (13.76)	2.2, 2.3, 2.4, 2.7
27523-40-6	Isoconazole	H	H			H	-	H	H	Antagonist (10.12)	2.3, 2.4, 2.7, 2.20
22916-47-8	Miconazole	H	H			H	-	H	H	Antagonist (7.74)	2.3, 2.4, 2.7, 2.20
65899-73-2	Tioconazole	H	H			H	-	H	H	Antagonist (5.52)	2.3, 2.4, 2.7, 2.20
99592-39-9	Sertaconazole	H	H			H	-	H	H	Antagonist (5.07)	2.3, 2.4, 2.7, 2.20
61318-91-0	Sulconazole	H	H			H	-	H	H	Antagonist (3.08)	2.2, 2.3, 2.4, 2.7

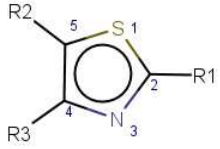
64211-46-7	Oxiconazole	H	-			H	-	H	H	Antagonist (2.66)	2.3, 2.4, 2.7, 2.20
101530-10-3	Lanoconazole		-		H	H	-	H	H	Antagonist (3.10)	2.2, 2.4, 2.7
187164-19-8	Lulizonazole		-		H	H	-	H	H	Antagonist (1.30)	2.2. 2.3, 2.4, 2.7
35554-44-0, 58594-72-2	Imazlil	H	H			H	-	H	H	Antagonist (0.68)	2.3, 2.4, 2.7, 2.20
24169-02-6	Econazole	H	H			H	-	H	H	Antagonist (0.68)	2.3, 2.4, 2.7, 2.20
68694-11-1	Trifumizole		H		H	H	-	H	H	Antagonist (0.05)	2.4, 2.7, 2.20
65277-42-1	Ketoconazole	H	-			H	-	H	H	Inactive	2.3, 2.4, 2.7, 2.16, 2.18, 2.20
78218-09-4	Dazoxiben	H	H	H		H	-	H	H	Inactive	2.7, 2.20
22994-85-0	Benznidazole	H	-	=O			-	H	H	Inactive	2.7, 2.16, 2.18
443-48-1	Metronidazole	H	H	H	OH	CH3	-	H		Inactive	2.10, 2.20
16773-42-5	Ornidazole	H	H	OH		CH3	_	H		Inactive	2.10, 2.20
3366-95-8	Secnidazole	H	H	OH	CH3	CH3	-	H		Inactive	2.10, 2.20

33125-97-2	Etomidate		H	H	H	H	-	H		Inactive	2.7, 2.17, 2.20
35944-74-2	Metomidate		H	H	H	H	-	H		Inactive	2.7, 2.17, 2.20
81167-16-0	Imiloxan	H	H	H	H		-	H	H	Inactive	2.10, 2.20

4.3 Concluding Remarks

Important information was retrieved from the **SAR analysis of monoazole-based drugs/chemicals.** In the case of benzothiazole scaffold (see Table 4.1), it was observed that the presence of amino group at position-2 of the benzothiazole seems to be crucial for the toxicity (agonist). In contrast, sulphur at position-2 led to antagonism behavior toward aromatase. Azole compounds bearing electron donating group (EDG) at position-5 and/or 6 (R3 and/or R4) were an active agonist toward CYP19A1. Conversely, inactivity was observed if EDG occupied the position-4 (R5). It must be noted that the stronger the EDG present at position-6 (R4), the higher will be the agonist activity. In the case of 1,3-thiazole scaffold (see Table 4.2), similar activity-relationships were observed akin to 2-aminobenzothiazole. However, agonist activity of 1,3-thiazole scaffold increased many-fold by introducing a better EDG at position-2 (R1) and electron withdrawing group (EWG) at position-5 (R2). It was interesting to note that the presence of aromatic amide (-NCOAr) at position-2 (R1) was associated to agonist activity, while aliphatic amide (-NCOR) or cyclic amide were linked to inactivity. Table 4.16 summarized the main outcomes from the SAR analysis of benzothiazole and 1,3-thiazole scaffolds.

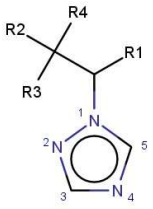
Table 4.16: Key outcomes from the SAR analysis of the monoazole scaffolds.

	**Benzothiazole**		
Agonist activity Indicator(s)	Agonist activity Performance Indicator(s)	Antagonist Indicator(s)	Inactivity Indicator(s)
• X = N • R1 to R5 = H	• R3 and/or R4 = EDG • The stronger the EDG present at R4, the better the agonist activity	X = S	• R5 = EDG • R3, R4, R5 = H
	**1,3-Thiazole**		

Agonist activity Indicator(s)	Agonist activity Performance Indicator(s)	Antagonist Indicator(s)	Inactivity Indicator(s)
• R1 = NH_2, S, -NCOAr or Ar • R3 = Ar	• R1 = -NCOAr and R2 = EWG such as -NO_2, -COOH • R1 = Better electron donating amine group	NA	• R1 = -NCOR where R is an aliphatic or cyclic amide • R1 = imine (-N=C sp2 Nitrogen)

***The SAR analysis* of the Triazole-based drugs/chemicals concluded that:** N-ethyl-1,2,4-triazole scaffold was derived from SA3.4 (N-isopentyl-1,2,4-triazole) in order to observe the SARs for triazole-based drugs/chemicals (see SAR analysis of Table 4.4**,** Table 4.5, and Table 4.6). In this case, it was observed that the role of substituents present at the triazole ring could not be predicted toward aromatase, but enhancement in the antagonist activity could be achieved by reducing the electron withdrawing keto group (>C=O) at the β-carbon. The presence of para-chloro benzyl in place of para-chloro phenoxy at the α-carbon turned the compound inactive. Conversely, the presence of para-chloro benzyl at the β-carbon resulted in activity. Although the compound bearing dioxalane moiety at the β-carbon showed inactivity toward aromatase, introducing the electron donating alkyl group in the dioxolane group could make the compounds an active antagonist. Triazoles having –OH and –CN group at position-R2 required low as well as high inhibition concentration suggesting that there is a minimal role of these groups in altering the activity. This observation indicated that inhibitor variant must include a major contribution from groups present at position-R3. In general, compounds bearing better electron donating alkyl group at position-R3 showed high activity irrespective of the presence of other groups. All these outcomes from the SAR analysis of triazole-based fragment have been summarized in Table 4.17.

Table 4.17: Key outcomes from the SAR analysis of the 1,2,4-triazole derivatives.



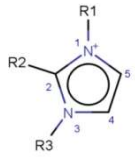
N-ethyl-1,2,4-triazole derivatives

Antagonist activity Indicator(s)	Antagonist activity Performance Indicator(s)	Inactivity Indicator(s)
• R1 = para-chloro phenoxy group • R2 = dioxolane bearing alkyl group	• R3 = alkyl or alkyloxy group	• R1 = para-chloro benzyl • R2 = dioxolane

• R3 = tertiary butyl group • R4 = mono/di-chloro phenyl group			

Finally, the detailed SAR examination of the Diazole-based drugs/chemicals has established that the imidazolium scaffold of diazole (see Table 4.7) seemed to be an important fragment and was obtained through the analysis of SA2.1. It was interesting to note that imidazolium-based ionic liquids exhibited antagonist activity toward aromatase. Imidazolium ionic liquids having alkyl chain length of less than eight carbons were not significant enough to show antagonist activity (see Table 4.18). The inhibition activity of alkyl imidazolium ionic liquids increased with the increase in the alkyl chain length.

Table 4.18: Key features describing the SAR analysis of imidazolium ionic liquids.

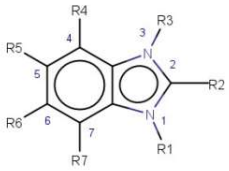


Imidazolium ionic liquids

Antagonist activity Indicator(s)	Antagonist activity Performance Indicator(s)	Agonist activity Indicator	Inactivity Indicator(s)
• R1 = Methyl /Long alkyl chain such as -$(CH_2)_n$-CH_3 where n > 6 • R2 = H/Methyl • R3 = Long alkyl chain such as -$(CH_2)_n$-CH_3 where n > 6	• R1 and R3 = long alkyl chain • R3 = -$(CH_2)_n$-CH_3 where n > 6, the greater the value of n, the greater the activity	NA	• R3 = -$(CH_2)_n$-CH_3 where n < 6 • R3 = Benzyl group

SARs of diazole-based drugs/chemicals containing benzimidazole scaffold were also established toward aromatase activity (see SAR analysis of Table 4.7). All compounds where position-R1 was occupied by an atom other than hydrogen were inactive chemicals. However, if position-R1 is H, the compound could be inactive, agonist or antagonist. This observation indicated that the activity explicitly depended on substituents present at position-R2, R5 and R6. A clear investigation of Table 4.8 indicated that the carbamate group tends to contribute toward the agonist activity irrespective of the fact whether it has occupied position-2 or 5 of the benzimidazole. It was also observed that if carbamate group was present at position-2, then position-5 should be occupied by the appropriate chemical group (probably EDG). However, substitution with sulfide group at position-2 or with benzyl imidazole at position-6 changed the activity of the benzimidazole scaffold to antagonist activity. The main outcomes from the SAR analysis of diazoles containing benzimidazole scaffold have been presented in Table 4.19.

Table 4.19: Key features describing the SAR analysis of benzimidazole derivatives.

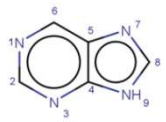


Benzimidazole derivatives

Agonist activity Indicators	Agonist activity Performance Indicators	Antagonist activity Indicators	Inactivity Indicators
• R1 = H • R2 = methyl carbamate or sulfinyl or thiazole group	• R5 = sulfide group or electron donating group	• R2 = sulfide group • R6 = benzyl imidazole	• R1 = alkyl, benzyl group • R2 = -NH_2, alkyl, phenyl • R5 = sulfinyl, benzoyl, alkoxy groups

Another important finding for diazoles was SA 2.11 in the form of purine and its derivatives, where the position-6 was mainly connected to the oxygen, nitrogen and sulphur bearing substituents Table 4.20The critical evaluation of these substitutions led to interesting conclusions such as: i) all compounds with oxygen substitution at position-6 were inactive toward aromatase, ii) all outcomes such as inactive, agonist and antagonist were possible for chemicals having nitrogen at position-6, iii) both active outcomes as agonist and antagonist were observed for chemicals that were connected through the sulphur (S) atom at position-6. All these variations toward the activity can be understood by the substituent present at position-2 and 9 as shown in Table 4.20.

Table 4.20: Key features describing the SAR analysis of purine scaffold.



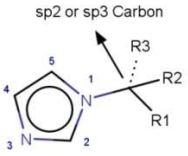
Purine scaffold

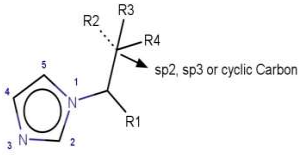
Agonist activity indicators	Antagonist activity indicators	Inactivity indicators
• Position-6 = NH2 and position-9 = H • Position-6 = benzyl amine and position-9 = pyran ring	• Position-6 = NH2 and position-9 = Fluorine containing furanose • Position-6 = benzyl amine and position-9 = furanose	• Position-6 = Oxygen linkage such as -OH, -OCH_3 or >C=O

• Position-6 = Sulhpur linkage and Position-2 = H	• Position-6 = doubly bonded sulphur and position-2 = -NH_2	

SARs of diazoles bearing the imidazole fragment and its N-substituted derivatives such as N-methyl and N-ethyl imidazoles were analysed (see Table 4.13, Table 4.14, and Table 4.15). It is worthy of note that imidazole derivatives, in general, showed antagonist activity toward aromatase. Although imidazole itself was inactive, it could become active if the hydrogen of the N-1 position is replaced by a substituent such as an aryl group (see Table 4.13). Appropriately 4-substituted (most likely EDG) imidazoles could also become active but require more inhibitory concentration than those of N-substituted imidazoles. In the case of inactive N-methyl imidazole, activity can be induced by replacing at least two hydrogens of the methyl group with substituent-like aryl group (see Table 4.14). The parent N-ethyl imidazole was also inactive like N-methyl imidazole. Its derivatives, however, showed antagonist activity. The substituents R1, R2, R3 and R4 of the N-ethyl imidazole affected the activity in the same way as in the case of N-ethyl-1,2,4-triazole (see Table 4.15). N-ethyl imidazole derivatives having substituents at position-R1 were found as both active and inactive, suggesting that position-R1 has either no or a minimal role in inducing the activity in the N-ethyl imidazole scaffold. Substituents present at position-R3 (2,4-dichlorophenyl) and R4 (alkoxy derivatives) were playing an important role toward activity. When comparing the role of position-R3 and R4 towards activity, substituents present at position-R3 were more effective. All these results have been summarized in Table 4.21.

Table 4.21: Key outcomes from the SAR analysis of imidazole and its N-substituted derivatives.

Diazole with structural alert	Antagonist activity indicator(s)	Antagonist activity Performance Indicator(s)	Inactivity indicator(s)
 **Imidazole**	• Position-1 = Aryl group in place of H	NA	• Position-1 = H • Position-2 or 5 = EWGs such as –NO_2, amide
 **N-methyl imidazole derivatives**	• Position-R2 = Aryl or 2°-amine	• R1, R2 = Aryl	• Position-R2 = aryl substituent bearing –COOH • Position 2 = EDG or EWG

			• Position-5 = EWG
 **N-ethyl imidazole derivatives**	• R1 = mono-chloro phenoxy group • R3 = mono/di-chloro benzyloxy group or mono-chloro benzyl thiol group or alkyloxy group • R4 = di-chlorophenyl group	• R3 = mono-chloro benzyloxy group or alkyloxy group or allyloxy	• Position-2 or 5 = EWG such as nitro or ester group

Chapter 5

Read-Across and Category Formation

5.1 Introduction

The reduction of *in vivo* toxicity testing has become essential in terms of the resources required and animals used. Grouping or category formation of chemicals can reduce the number of chemicals to be tested, since the available information for endpoints can be used to estimate properties of untested substances [112]. Read-across is an intriguing approach and an important tool for data gap filling [111, 112]. This chapter introduces a new workflow for the identification of analogues for RAX and exemplifies its potential applications for category formation.

5.2 Methods

5.2.1 Constructing RAX workflow

To select analogues for read-across an automated workflow was implemented using the KNIME platform, the workflow adapted and modified the scheme proposed in Gadaleta et al. [131], see

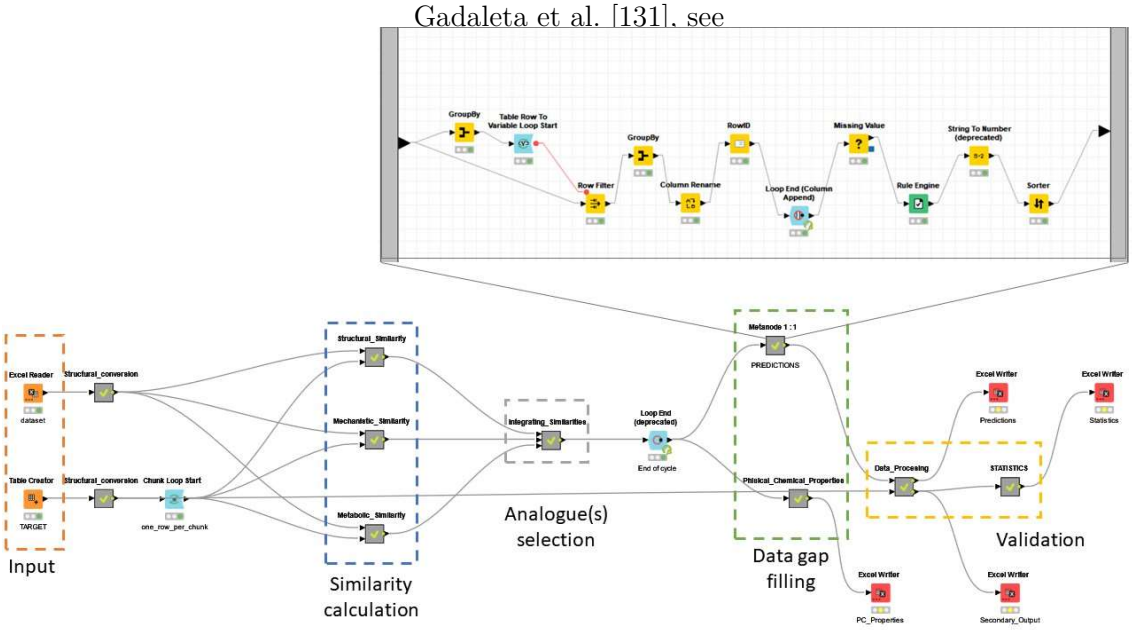


Figure 5.1. The workflow consists of the calculation of chemical, metabolic and mechanistic similarities through a search performed in a dataset. The individual lists of candidates to analogues were retrieved from the dataset considering different kinds of similarity approaches, as for example, structural similarity, a common metabolic behavior, or structural alerts. In the next step, the chemicals in the source database were ranked considering each kind of similarity. Finally, the output is provided including only the intersection between all top ranked compounds. This intersection is then considered as the list of most suitable analogue(s), and their activity is used to predict the activity of the target chemical [134].

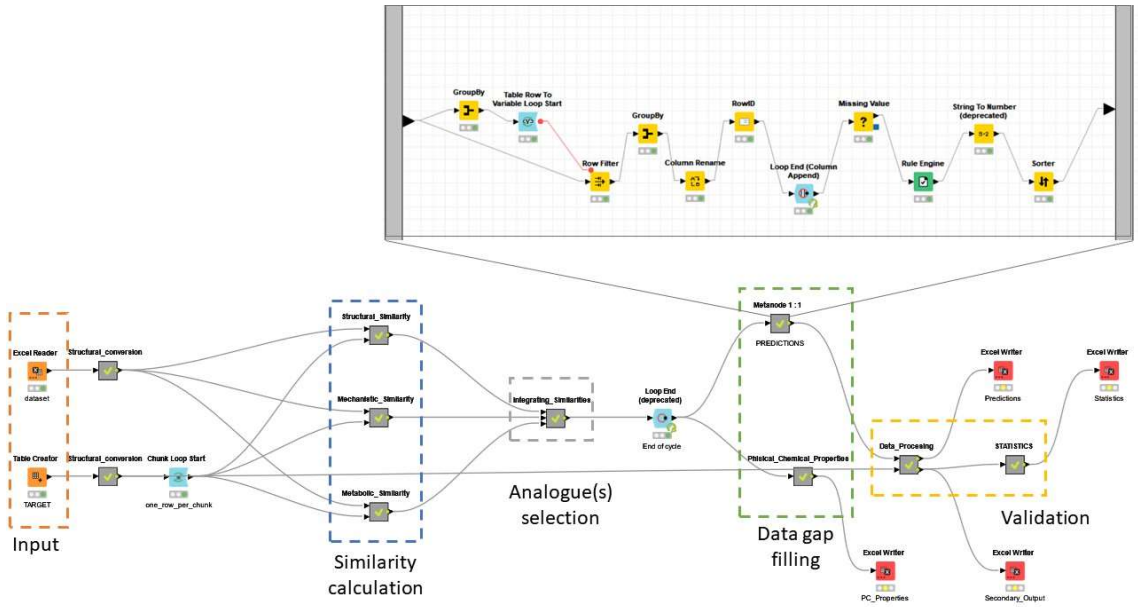


Figure 5.1: Workflow implemented in KNIME for the automatic integration of structural, metabolic and mechanistic similarities.

5.2.2 Structural Similarity (StrS)

StrS was based on PubChem fingerprints [199] that were calculated for both target and analogue(s) with the KNIME implementation of the RDKit toolset. PubChem fingerprints are usually applied to calculate structural similarity [128, 200]. This kind of codification is useful to disclose analogies for chemical features relevant for biological and toxicological profiles [131]. StrS was calculated using Tanimoto index as proposed in Eq. (2.3). A similarity index of 0.7 was taken as a threshold for the purpose of detecting structurally similar source analogues [128, 156].

5.2.3 Mechanistic similarity

Structural alerts can be used to evaluate chemical similarity since certain groups or fragments experimentally have demonstrated to be associated with specific toxic effects [121, 164]. Hence, robust mechanistic structural alerts can be applicable as a screening tool to recognize potential mechanistic similarities [121]. Compounds in the source dataset were filtered based on the presence of structural alerts in common with the target. The SAs used in this chapter were recovered from the alerts obtained in Chapter 3. The presence of 21 structural fragments towards human aromatase binding codified as SMARTS was verified for both the target and the potential analogue(s). Those SAs codify for specific mechanistic

fragments which demonstrated to be relevant to define the activity of chemicals and are available in Figure 3.4. For the purpose of the McS, any substances with at least a common SA with the target was considered as similar and therefore retrieved for the future workflow steps.

5.2.4 Metabolic similarity

Factors focused in metabolism are proposed through chemical grouping by identifying similar chemicals [112]. In this thesis, the metabolism was explored using WhichCyp package within KNIME [201]. The node predicts which cytochromes P450 isoforms (among 1A2, 2C9, 2C19, 2D6 and 3A4) a given molecule is likely to inhibit. A model output value as "1" indicates likely to bind, while "0" means unlikely to interact, using the output of the model, binary bit vectors were computed for both the target and the possible analogue(s)[201]. The generated vectors were utilized to compute the similarities using the Tanimoto index, the threshold used to include analogues was 0.7.

5.2.5 Integrating similarities

The individual lists of analogues were integrated to identify the most suitable analogues for data gap filling. The integration approach adapts and modifies the scheme proposed by Gadaleta et al. [131]. The same weight was assigned to all similarities to restrict the number of analogues to the most suitable ones. Therefore, the presence of a chemical in more than one list was interpreted as a higher level of similarity. The prediction of the activity/toxicity was made following a majority vote approach of activities of selected analogues. In the present case, all substances selected according to the similarity criteria, as described above, were used. Conversely, the approach used by Gadaleta et al. had a maximum number of similar substances to be used, according to the similarity metrics [131]. However, our approach had a low applicability domain, thus the methodology did not process most of the substances to be evaluated. For this reason, we preferred not to impose further restrictions to the number of similar compounds.

5.2.6 Validation

It is important to highlight that this workflow was developed to identify the interpretable and most suitable analogues for read-across, bearing in mind that read-across is a supervised method for data gap filling which requires the expert judgment, which may or may not achieve the same level of performance as a set of predictions from the QSAR methodologies. [131]. The presented evaluation process has been validated by predicting chemicals using source dataset by Leave-One-Out (LOO) method which led to information about the overall performance of the workflow.

Additionally, predictions were performed using the StrS to compare the contribution of multiple components to the performance of the workflow, with respect to the traditional use of structural similarity. A similarity index of 0.7 was taken as a threshold for the purposes of detecting a realistic initial number of structurally similar source analogues [202].

5.3 Results and Discussions

5.3.1 Workflow performance

Using the workflow shows in Figure 5.1, 92% (approx.) of those compounds with active properties on the enzyme were correctly classified, whereas 84% of chemicals with non-interacting (non-binding) effect were correctly identified by the workflow after integrating the three-similarity lists. So, based on the classification LOO method, we can expect that at least 88% of a set of new azoles will be correctly classified. Details of the classification performance are provided in Table 5.1.

As in Table 5.1, the difference in the classification performance obtained from StrS to Integrated Similarity (IntS) approaches were observed, where the significantly higher sensitivity for the integrated approach was observed than that of the StrS approach, while the integrated approache led more balanced predictions (Low difference between Se and Sp) and higher accuracy. These observations indicated that the true prediction (sensitivity) that is the activity (active compounds) of the azole chemicals was dependent on the metabolic similarity and mechanistic similarity along with structural similarity. The overall performance for the truthfulness of the prediction by the system was estimated by the MCC coefficient, which was 0.49 and 0.77 for the similarity approach and integrated approach respectively. The higher MCC (0.77) was significantly improved when only analogue(s) included in three similarity lists were considered. MCC has been classified as a more informative and truthful score for evaluating the binary classifications than accuracy [203], which produces a high score only if the good quality prediction is obtained in all the four confusion matrix categories.

Table 5.1: Classification matrix and classification performance metrics of the workflow using StrS and IntS approaches.

		Structural similarity		Integrated similarities	
		Predicted			
		Positive	Negative	Positive	Negative
Experimental	Positive	49	35	34	3
	Negative	23	178	5	27
Sensitivity		0.58		0.92	
Specificity		0.89		0.84	
Accuracy		0.80		0.88	
Error_rate		0.20		0.12	
Unpredicted rate		0.04		0.79	
MCC		0.49		0.77	

Overall, a good performance was reflected for most of the chemical categories identified with the IntS, however the unpredicted rate was much higher (0.79), which indicated the reduced applicability domain. The reduction in applicability domain in IntS approach is generally expected due to its inheritance of incorporating many contributing factors and approaching toward the much stringent criteria for matching the similarities.

The obtained results corroborate that the use of analogue(s) with diverse kinds of similarity enhances performance compared to the sole use of structural similarity. This is likely certain when source compounds used for RAX share more than one structural alert with the target chemical, in this regard, they can be considered as mechanistically closer to the target.

The analogues included in multiple similarity lists are more likely to match the activity of the target compound. Indeed, the combination of all three similarity lists showed the lowest ratio of source compounds having a different activity to the target, decreasing this ratio more than 3-fold with respect to the sole use of StrS. The Table 5.2 presents the number of sources compounds (non-unique) matching/not matching the target activity, along with the ratio of those not matching their target's activity.

Table 5.2: Ratio of analogues having a different activity to the target (non-unique count).

	Number of sources matching the target activity		Number of sources not matching the target activity		Ratio of sources not matching their target activity
StrS	2050		944		0.46
	active	inactive	active	inactive	
	500	1550	472	472	
IntS	211		26		0.12
	active	inactive	active	inactive	
	44	167	13	13	

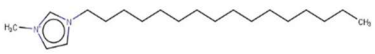
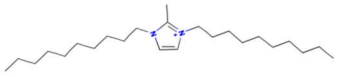
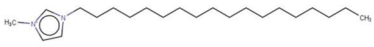
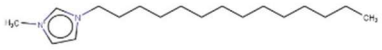
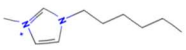
5.3.2 RAX case studies

Table 5.3 and Table 5.5 showed that IntS approach was advantageous over the StrS alone. A total of 11 analogues were identified to predict the activity of the know active compound 1-Hexadecyl-3-methylimidazolium (CAS 61546-01-8). Out of those 11 analogues, 1,3-Didecyl-2-methylimidazolium (CAS 70862-65-6), 1-Methyl-3-octadecylimidazolium (CAS 219947-96-3) and 1-Methyl-3-tetradecylimidazolium (CAS 171058-21-2) were included in all three similarity lists. These three analogues were top ranked chemicals for StrS, occupying the 4^{th}, 1^{st}, and 2^{nd} positions, respectively, however other eight analogues were also recognized for StrS out of which 3 were active compounds and 5 inactive ones (see Table 5.3).

Despite both predictions matched the real activity of the target, the metabolic and mechanistic similarities evidenced the need of extra components for identifying the most suitable analogues. Indeed, all inactive analogues from StrS were excluded from the list of IntS. Differently from the category identified using solely StrS, the three chemicals conforming the category from the intersection of three list of similarities, displayed a correlation of the metabolic pattern with the one predicted for the target with respect to the five CYP isoforms, i.e. 2D6 inhibition and absence of other interactions with isoforms 1A2, 2C9, 2C19 and 3A4. For example, in Table 5.3 it can be seen how the 1-Hexyl-3-methylimidazolium, obtained from the StrS list, differed in the metabolic pattern with respect to the target, while the 1-Methyl-3-tetradecylimidazolium showed a perfect concordance in metabolic behavior to that of target. Furthermore, among the mechanistic

similarity, the structural alert (SA2) (see Figure 3.4) (alkyl imidazolium derivatives) was also crucial to screen the analogues. For example, the structural alert SA2 had ability to discard the 1-Hexyl-3-methylimidazolium ion derivatives, even though it exhibited a high structural similarity with the target (~0.95). This finding agreed with our previous SAR analysis, where minimum lateral chain of 8 carbon atoms was necessary to observe activity on CYP19A1 enzyme (see Imidazolium cation scaffolds section).

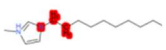
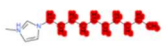
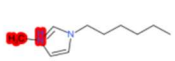
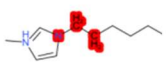
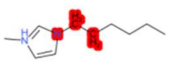
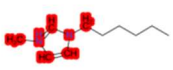
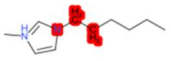
Table 5.3: RAX example from the Aromatase binding dataset.

Name	Structure	Rank	Activity	Structural Similarity	Metabolic Similarity	Common mechanistic structural alerts
1-Hexadecyl-3-methylimidazolium		target	active	-	-	-
1,3-Didecyl-2-methylimidazolium		IntS StrS	active	0.976	1	SA2
1-Methyl-3-octadecylimidazolium hexafluorophosphate		IntS StrS	active	1	1	SA2
1-Methyl-3-tetradecylimidazolium chloride		IntS StrS	active	1	1	SA2
1-Hexyl-3-methylimidazolium chloride		StrS	Inactive	0.951	-	-

6 additional analogue(s)
3 actives, 4 inactives

The Isoforms with value 1 are predicted to be inhibited. Highlighted fragments which were found to be most significant for binding/ non-binding by CYP isoform.

	1A2	2C9	2C19	2D6	3A4

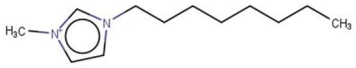
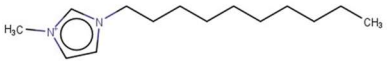
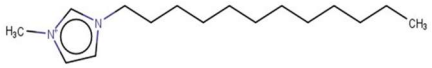
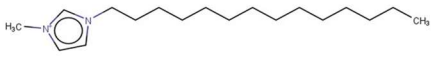
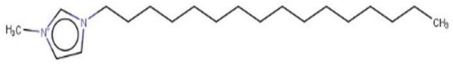
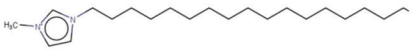
1-Hexadecyl-3-methylimidazolium chloride (TARGET)					
	0	0	0	1	0
1-Hexyl-3-methylimidazolium chloride					
	0	0	0	0	0
1-Methyl-3-tetradecylimidazolium chloride					
	0	0	0	1	0

A brief overview of the literature showed that the alkyl imidazolium compounds family tends to exhibit a common toxic profile. In fact, toxicity of ILs to the human health has been defined through cell membrane disintegration followed by increased production of reactive oxygen species and DNA damage [204]. The number of studies has concluded that imidazolium cation having a linear alkyl chain length more than five (>5) carbon atoms probably exhibits moderate-to-very high cytotoxicity. Cytotoxicity was found to exhibit a linear relationship with lipophilicity of the cation (longer alkyl chains having greater lipophilicity) [205-207]. In the literature, six N-methylimidazolium-based ILs ([Cnmim]X, n=4, 6, 8; X=Br- , Cl- , BF4 - , CF3SO3 -) exhibited lactic dehydrogenase (LDH) inhibition activity [208]. Moreover, LDH inhibition activity was also known to increase with an increase in the alkyl chain length of Ils [208]. Similarly, ILs were found to inhibit the lipase enzyme activity, where the degree of inhibition depends on the chemical nature of anions [209]. It was observed that ILs having anions such as halides (X- = Cl, Br) derived their activity through lipophilicity, while ILs having F4B- , CF3SO3- , ClO4- and N(CN)2— anions derived their inhibition activity through hydrogen bonding.

Following the examination of the achieved results, we identified a series of aromatase antagonist ionic liquids as shown in Table 5.4. In this case, it was also observed that the presence of counter anions and alkyl chain length affected the antagonist activity of ILs toward aromatase CYP19A1. A more detailed analysis regarding the chemical behavior vs the aromatase activity within this category was presented in Table 4.18.

The examined case study not only constitutes a pragmatic example about how the integration of similarities could represent an advantage for identifying analogues for RAX, but also showed a high degree of consistency between the obtained category and the literature studies reported [205-207].

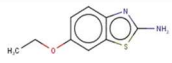
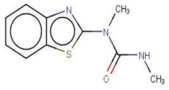
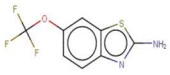
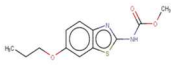
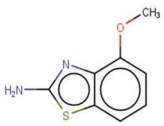
Table 5.4: Category members formed for alkyl imidazolium compounds.

Ionic Liquid containing Imidazolium Ion	Counter Anion			
	Cl-	CF3SO3-	F4B-	F6P-
 1-methyl-3-octylimidazolium	Active Antagonist (54.10 µM)	Active Antagonist (27.53 µM)	Active Antagonist (34.34 µM)	Active Antagonist (55.04 µM)
 1-methy-3-decylimidazolium	Active Antagonist (21.67 µM)	Active Antagonist (12.52 µM)	NA	Active Antagonist (4.85 µM)
 1-methyl-3-dodecylimidazolium	NA	Active Antagonist (3.12 µM)	Active Antagonist (3.85 µM)	Active Antagonist (3.06 µM)
 1-methyl-3-tetradecylimidazolium	Active Antagonist (5.39 µM)	NA	NA	NA
 1-methyl -3-hexadecyl imidazolium	Active Antagonist (4.42 µM)	NA	Active Antagonist (6.17 µM)	NA
 1-methyl-3-octadecylimidazolium	NA	NA	NA	Active Antagonist (3.83 µM)

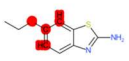
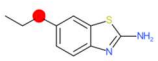
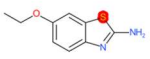
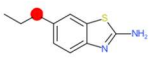
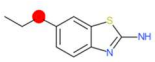
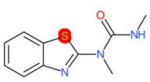
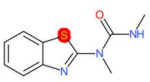
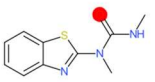
A similar observation was noticed when the target chemical was the active compound 2-Amino-6-ethoxybenzothiazole (CAS 94-45-1). Comparably, the target was correctly predicted by both RAX approaches, whereas the analogues identified by the IntS were also included within the StrS list; these were Methabenzthiazuron (CAS 18691-97-9), Riluzole (CAS 1744-22-5) and the Tioxidazole (CAS 61570-90-9).

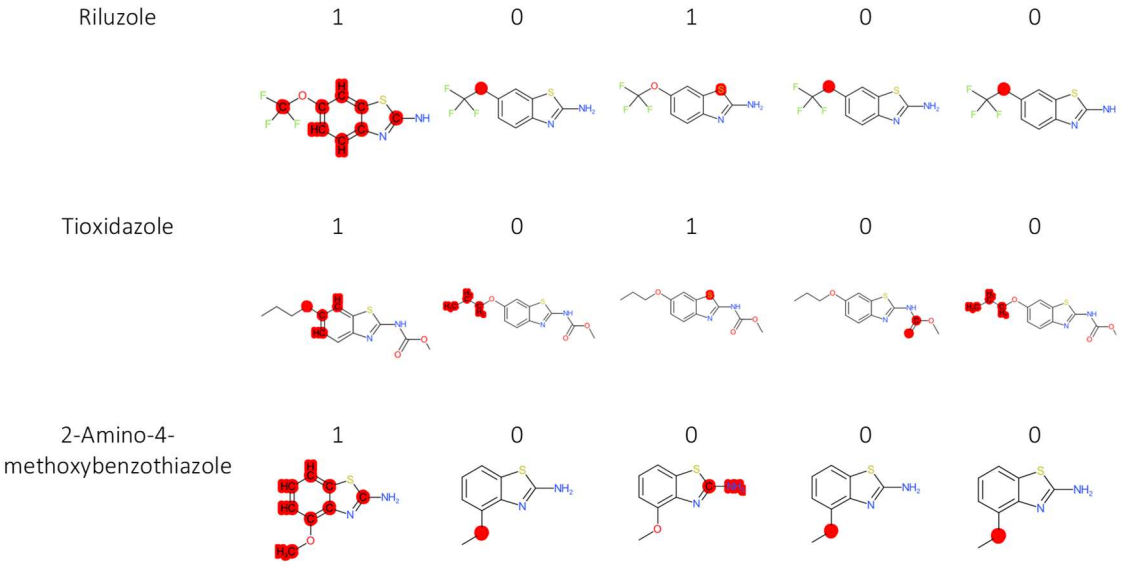
As in the previous example, the metabolic component included in the IntS was capable to reject all the inactive chemicals obtained from the StrS list, including the 2-Amino-4-methoxybenzothiazole, a chemical compound which presented a high structural similarity (0.90) with the target (see Table 5.5). These inactive chemicals removed by the IntS approach showed substitutions at position 4 of the 2-aminobenzothiazole scaffold, which according to a Structure Activity Relationship (SAR) study reported in the literature have been associated to inactive chemicals on human aromatase, see Table 4.1, which captured how the substitutions of 2-aminobenzothiazole scaffold (SA1) at position 4 leads to inactive chemicals. In this case, the prediction of the two approaches was 3 active and 0 inactive for IntS, while it was 6 actives and 3 inactives for StrS.

Table 5.5: RAX example from the Aromatase binding dataset.

Name	Structure	Rank	Activity	Structural Similarity	Metabolic Similarity	Common mechanistic structural alerts
2-Amino-6-ethoxybenzothiazole (target)		TARGET	Active	-	-	-
Methabenzthiazuron		IntS StrS	Active	0.730	1	SA1
Riluzole		IntS StrS	Active	0.936	1	SA1
Tioxidazole		IntS StrS	Active	0.944	1	SA1
2-Amino-4-methoxybenzothiazole		StrS	Inactive	0.897	-	-
5 additional analogues 2 actives, 2 inactives						

The isoforms with value 1 are predicted to be inhibited. Highlighted fragments were most significant to predict the binding/non-binding by CYP isoform.

	1A2	2C9	2C19	2D6	3A4
2-Amino-6-ethoxybenzothiazole (TARGET)	1 	0 	1 	0 	0 
Methabenzthiazuron	1 	0 	1 	0 	0 
Riluzole	1	0	1	0	0



Moreover, the integrated system demonstrated a higher performance against the use of the StrS alone in many cases. For example, when the target chemical was the inactive compound Paclobutrazol (CAS 76738-62-0), the StrS approach produced a category comprised by 13 chemical analogues, all of them with a high level of structural similarity to the target. The category included 10 active and 3 inactive chemicals, leading to a false positive prediction. Contrary, the inclusion of metabolic and mechanistic components within the IntS, led to reduce the number of analogues to a unique suitable compound, the inactive chemical Triticonazole (CAS 131983-72-7).

5.4 Concluding Remarks

This chapter describes an automated workflow that identifies analogue(s) for read-across. This approach can support a toxicologist during the analogue identification step, but also can provide an automated prediction based on most suitable analogues.

It offered an easily interpretable framework constructed upon diverse similarity considerations for read-across on human aromatase. A key component of this method involved evaluating chemical, metabolic and mechanistic similarities between the target and sources that could define the toxicological outcomes of the enzyme. As in conventional RAX, the structural similarity constituted the starting point for the approach, however the addition of new components strengthened the evidence in the analogue(s) selection.

The result of this process is a more reliable prediction, together with a list of most suitable analogue(s) which can be applied for RAX. This approach could provide a comprehensive basis for selecting appropriate analogues in RAX for other endpoints, however toxicity is evidently dependent on many variables and any new endpoint should be individually studied.

The link between the metabolism and toxicity has been understood for many compounds, as well as its relevance to evaluate similarities and differences, but the uncertainty related to metabolic predictions continued to be challenging. For similarity assessment, another metabolic consideration could be addressed, e.g., same reactivity, metabolic pathway, or bioavailability, but the introduction of new predictions can also increase the uncertainty.

In general, the methodology exhibits a good predictive performance comparable to a QSAR model but suffered from the use of a small database and a restrictive approach, causing a high number of compounds not predicted. Taking into account that this approach was designed to identify the most suitable analogues for RAX, and not to predict a large-scale of chemical toxicity, we can affirm that the main strength of this integrated RAX approach is its ability to provide reliable and simply interpretable results, joined to appropriate data to support the final evaluation of the toxicity outcome.

The application of this workflow should be performed case-by-case, for experts' decision regarding the appropriateness of the identified analogue(s) since the same weight was considered for the integration of the three similarities. The outcome of this workflow could be combined with other sources of evidence, as for example, adverse outcome pathways which have demonstrated to be an attractive and confident method of identifying toxic effects.

Chapter 6

Conclusions

This final chapter will present a summary and conclusions of the investigation described in this thesis. Conclusions are discussed in the context of the background of the project and the rationale for the research undertaken. In the following stage, an overview of the key findings from the research presented in Chapters 3, 4 and 5 will be introduced. The final part of the chapter will focus on potential upcoming work opportunities related to concerns set up throughout the work, and applications of the research with regards to alternatives to traditional toxicity testing.

6.1 Summary of Work

The work in this thesis has been focused on the development of *in silico* methods to understand and predict the mechanism of action and chemical toxicity of azoles chemicals on human aromatase enzyme. The motivation for the development of alternative methods to predict chemical activity/toxicity results from several factors; the reduction refinement and replacement philosophy, the legislation and regulation as for example REACH, which encourages the use of non-animal-based methods and provides guidelines on the application of computational methods, and the 7th amendment to the EU Cosmetics Directive which effectively bans the use of animals for studying toxicity of cosmetic ingredients.

In this sense, azoles and their chemical derivatives, their applications and safety assessment are of continuous high interest. Therefore, this thesis will be a value addition to the knowledge which has been accumulated so far in this field. Despite the progress in computational approaches for mechanism elucidation, the toxic mechanism of action for non-estrogenic compounds on human aromatase has been less studied, however such knowledge can enhance the understanding of most effective and safety compounds.

Chapter 3 of this thesis was focused on the development of new structural alerts to classify azole chemicals according to their toxic mechanism of action. Those structural alerts explored chemicals agonizing and antagonizing the enzyme, to stablish the effect of certain functional groups on the activity of the enzyme. The fragments presented within inactive chemicals were also described.

After a rigorous process of fragment elucidation and screening, the most relevant structural alerts were retrieved and interpreted. Structural alerts were translated into important pieces for chemical screening. For example, imidazolium cation was associated to chemicals antagonizing the enzyme. This kind of knowledge is certainly valuable for lead

optimization, toxicity evaluation, risk assessment and many others. On the other hand, the compilation of those alerts into a single classifier provided a new tool for the fast screening of new active and safety drug candidates. Moreover, the concordance of those results with literature findings evidenced their appropriateness.

Chapter 4 outlined the Structure Activity Relationships elucidated by comparing the antagonist and/or agonist azoles with inactive azoles. This chapter provided the core explanation regarding the mechanism of action of azoles on aromatase, and the impact of structural substituents on the activity of the enzyme. While Chapter 3 allowed the development of new structural alerts, Chapter 4 provided the chemical and mechanistic explanations of each type of interaction. The exploration of the inductive effects of substituents through the identification of EDG and EWG in the molecules regarding the scaffolds resulted in the definition of important statements to cover different types of enzymatic interactions and to complement the results obtained within Chapter 3.

A comprehensible example of these results was the amino-benzothiazole structural alert identified within Chapter 3, and later structurally modified within Chapter 4 to form the benzothiazole scaffold. In this case, robust rules were stipulated, setting that the presence of amino group at position-2 of the benzothiazole was crucial for the toxicity (agonist activity), and how a sulphur substitution at position-2 led to antagonism behaviour toward aromatase. It was also elucidated that the azole compounds which bore electron donating group (EDG) at position-5 and/or 6 (R3 and/or R4) were active while inactivity was observed if EDG occupied the position-4 (R5). Moreover, it was defined that the stronger the EDG present at position-6 (R4), the higher was the agonist activity. These types of results provide a significant improvement to the current available knowledge, and rationalize the designing of new azole-based drugs/chemicals as per the desired medical/chemical applications, which was one of the themes that motivated this thesis.

Lastly, Chapter 5 focused on the development and implementation of an automatic workflow to identify analogues for read-across on aromatase. The main novelty of this workflow was the use of not only structural similarity but also of a combination of mechanistic structural alerts and metabolic profiles of chemicals to evaluate similarities. The results indicated a better performance of the workflow with respect to the sole application of structural similarity, leading to the identification of a reduced and robust number of analogues. Moreover, the assessment of the workflow against different sceneries showed its potential applications for assisting a toxicologist during data gap filling and risk assessment, exhibiting consistent results with those reported in the literature. As a last resort, the system provided the predicted values of physical-chemical properties for analogues and the target to facilitate the data gathering step and furnish the user with a most complete chemical profile. The simplicity of this workflow allows its adaptation to a different set of configurations. Moreover, this approach could constitute a conceptual basis for the development of future workflows applied to diverse endpoints.

This study expands on the novelty of the fragment-based *in silico* methods by developing new mechanistic structural alerts and classifiers to be used in many research areas such as medicine for anti-cancer agents. Furthermore, these approaches are likely to be of primary use in regulatory and predictive toxicology with the potential to be implemented in such freely available tools as VEGA. Such implementation, together with the KNIME workflow extension and the SARs would enable the development of mechanism-based models for safety assessment to be utilized within the chemical category approach to data gap filling. Additionally, these results are likely to be applied to support the evaluation of endocrine disruption of pesticides, insecticides, and many other chemicals with agricultural applications.

6.2 Prospects for Future Work

The work carried out in this thesis highlights the development and use of novel fragment-based *in silico* methods and read-across for the human aromatase binding evaluation. These methods were shown to be capable of predicting and classifying azoles chemicals according to their toxic mechanism of action. As a result of this, it is likely that the approach will be applicable to other chemicals domains relevant to toxicology. The expansion of the fragment-based *in silico* methods to additional domains is dependent on the availability of data, therefore the inclusion of new data could have an important contribution. This is likely to be derived primarily from other toxicity data sources. It could also be feasible that additional alerts could be developed from alternative data sources offering the prospect to deal with a wider range of chemical information within each of the mechanistic classes.

Additionally, given the high degree of correlation between toxicity and the endocrine system, it is possible that the latter regulators actions could promote the development of newest data sources to be used as a surrogate for toxicity evaluation. The analysis performed in this thesis could be integrated to another endocrine disruptor-associated endpoints, such as androgen receptor, estrogen receptor and thyroid receptor disruption. The integration of those data sources into a single database could encourage the development of *in silico* classifiers based on endocrine chemical profiles that in the future could support a weight of evidence approach.

Alternatively, an investigation could be carried out into the development of docking studies. The methods utilized in this thesis have an important role in the fast screening of new chemicals, and also in the elucidation of their mechanism of action. Nevertheless, we understand that docking studies are likely to be of use for the elucidation of the interaction of small molecules and proteins. In this sense, the performance of those in the binding site and the biochemical process involved could be characterized [210]. Multiple docking studies have already been performed on chemicals antagonizing the human aromatase activity, most of them focusing on compounds with potential medical applications. However, to the best of our understanding, no docking studies have been performed on chemicals agonizing the enzyme. An example of the importance of this toxicology is that many of the currently utilized pesticides have shown to agonize the enzyme in *in vitro* assays. In addition, there are other areas of research where knowledge of the ligand conformation and the binding affinity may be of use. For example, in the development of adverse outcome pathways, where the significance of such a type of study together with the mechanistic structural alert is crucial for molecular initial events.

Appendix A

Dataset

A.1 Full Dataset of Azole Compounds with Their Name, CAS Numbers, VEGA SMILES and Experimental Endpoints

Names	CAS_NOs	VEGA SMILES	Assay outcome	AC50	Flag
2,2'-Dithiobisbenzothiazole	120-78-5	n2c1ccccc1sc2SSc4nc3ccccc3s4	inactive	NA	Monoazole
2-(2-Methylpropyl)thiazole	18640-74-9	n1ccsc1CC(C)C	inactive	NA	Monoazole
2-Acetylthiazole	24295-03-2	O=C(c1nccs1)C	inactive	NA	Monoazole
2-Amino-4-chlorobenzothiazole	19952-47-7	n1c(N)sc2cccc(c12)Cl	inactive	NA	Monoazole
2-Amino-4-methoxybenzothiazole	5464-79-9	n1c(N)sc2cccc(OC)c12	inactive	NA	Monoazole
2-Amino-4-methylbenzothiazole	1477-42-5	n1c(N)sc2cccc(c12)C	inactive	NA	Monoazole
2-Amino-4-thiazoleacetic acid	29676-71-9	O=C(O)Cc1nc(N)sc1	inactive	NA	Monoazole
2-Aminothiazole	96-50-4	n1ccsc1N	inactive	NA	Monoazole

2-Benzoxazolinone	59-49-4	O=C1Oc2ccccc2(N1)	inactive	NA	Monoazole
2-Bromobenzothiazole	2516-40-7	n2c1ccccc1sc2Br	inactive	NA	Monoazole
2-Methylbenzothiazole	120-75-2	n2c1ccccc1sc2C	inactive	NA	Monoazole
5-Acetyl-2,4-dimethylthiazole	38205-60-6	O=C(c1c(nc(C)s1)C)C	inactive	NA	Monoazole
5-Methyl-3-isoxazolamine	1072-67-9	n1oc(cc1N)C	inactive	NA	Monoazole
6-Chloro-2,3-dihydrobenzoxazol-2-one	19932-84-4	O=C1Oc2cc(ccc2(N1))Cl	inactive	NA	Monoazole
Aztreonam	78110-38-0	O=C(O)C(ON=C(C(=O)NC1C(=O)N(C1(C))S(=O)(=O)O)c2nc(N)sc2)(C)C	inactive	NA	Monoazole
Benzothiazole	95-16-9	n1csc2ccccc12	inactive	NA	Monoazole
Carumonam sodium	86832-68-0	O=C(OCC2N(C(=O)C2(NC(=O)C(=NOCC(=O)O)c1nc(N)sc1))S(=O)(=O)O)N	inactive	NA	Monoazole
Cefdinir	91832-40-5	O=C(O)C3=C(C=C)CSC2N3(C(=O)C2(NC(=O)C(=NO)c1nc(N)sc1))	inactive	NA	Monoazole
Cefepime chloride hydrochloride hydrate	123171-59-5	O=C(O)C4=C(C[N+]1(C)(CCCC1))CSC3N4(C(=O)C3(NC(=O)C(=NOC)c2nc(N)sc2))	inactive	NA	Monoazole

Cefetamet pivoxil HCl	111696-23-2	O=C(OCOC(=O)C(C)(C)C)C3=C(C)CSC2N3(C(=O)C2(NC(=O)C(=NOC)c1nc(N)sc1))	inactive	NA	Monoazole
Cefixime	79350-37-1	O=C(O)C3=C(C=C)CSC2N3(C(=O)C2(NC(=O)C(=NOCC(=O)O)c1nc(N)sc1))	inactive	NA	Monoazole
Ceftiofur	80370-57-6	O=C(O)C4=C(CSC(=O)c1occc1)CSC3N4(C(=O)C3(NC(=O)C(=NOC)c2nc(N)sc2))	inactive	NA	Monoazole
Ceftizoxime sodium	68401-82-1	O=C(O)C3=CCSC2N3(C(=O)C2(NC(=O)C(=NOC)c1nc(N)sc1))	inactive	NA	Monoazole
Chlorzoxazone	95-25-0	O=C1Oc2ccc(cc2(N1))Cl	inactive	NA	Monoazole
Clomethiazole	533-45-9	n1csc(c1C)CCCl	inactive	NA	Monoazole
Clothianidin	210880-92-5	O=[N+]([O-])N=C(NC)NCc1cnc(Cl)s1	inactive	NA	Monoazole
Cloxacillin sodium	642-78-4	O=C(O)C4N3C(=O)C(NC(=O)c1c(noc1C)c2ccccc2Cl)C3SC4(C)(C)	inactive	NA	Monoazole
Epothilone B	152044-54-7	O=C2OC(C(=Cc1nc(C)sc1)C)CC3OC3(C)(CCCC(C	inactive	NA	Monoazole

		)C(O)C(C(=O)C(C)(C)C(O)C2)C)			
Famotidine	76824-35-6	O=S(=O)(N=C(N)CCSCc1nc(N=C(N)N)sc1)N	inactive	NA	Monoazole
Fenclozic acid	17969-20-9	O=C(O)Cc2nc(c1ccc(cc1)Cl)sc2	inactive	NA	Monoazole
Hymexazol	10004-44-1	O=C1C=C(ON1)C	inactive	NA	Monoazole
Isoxaben	82558-50-7	O=C(Nc1onc(c1)C(C)(CC)CC)c2c(OC)cccc2(OC)	inactive	NA	Monoazole
Isoxaflutole	141112-29-0	O=C(c1cnoc1C2CC2)c3ccc(cc3S(=O)(=O)C)C(F)(F)F	inactive	NA	Monoazole
Isoxathion	18854-01-8	n1oc(cc1OP(OCC)(OCC)=S)c2ccccc2	inactive	NA	Monoazole
Isoxicam	34552-84-6	O=C(Nc1noc(c1)C)C3=C(O)c2ccccc2S(=O)(=O)N3C	inactive	NA	Monoazole
Lintitript	136381-85-6	O=C(O)Cn2c(cc1ccccc12)C(=O)Nc3nc(cs3)c4ccccc4Cl	inactive	NA	Monoazole
Meloxicam	71125-38-7	O=C(Nc1ncc(C)s1)C3=C(O)c2ccccc2S(=O)(=O)N3C	inactive	NA	Monoazole
Mofezolac	78967-07-4	O=C(O)Cc3onc(c1ccc(OC)cc1)c3(c2ccc(OC)cc2)	inactive	NA	Monoazole

N-Acetyl sulfamethoxazole	21312-10-7	O=C(Nc1ccc(cc1)S(=O)(=O)Nc2noc(c2)C)C	inactive	NA	Monoazole
Niridazole	61-57-4	O=C1NCCN1c2ncc([N+](=O)[O-])s2	inactive	NA	Monoazole
Nithiamide	140-40-9	O=C(Nc1ncc([N+](=O)[O-])s1)C	inactive	NA	Monoazole
Oxaprozin	21256-18-8	O=C(O)CCc1nc(c(o1)c2ccccc2)c3ccccc3	inactive	NA	Monoazole
Paliperidone	144598-75-4	O=C1C(=C(N=C2N1CCCC2(O))C)CCN5CCC(c4noc3cc(F)ccc34)CC5	inactive	NA	Monoazole
Pamicogrel	101001-34-7	O=C(OCC)Cn4cccc4(c3nc(c1ccc(OC)cc1)c(c2ccc(OC)cc2)s3)	inactive	NA	Monoazole
Phthalylsulfathiazole	85-73-4	O=C(O)c3ccccc3(C(=O)Nc1ccc(cc1)S(=O)(=O)Nc2nccs2)	inactive	NA	Monoazole
Pramipexole	104632-26-0	n1c(N)sc2c1CCC(NCCC)C2	inactive	NA	Monoazole
Risperidone	106266-06-2	O=C1C(=C(N=C2N1CCCC2)C)CCN5CCC(c4noc3cc(F)ccc34)CC5	inactive	NA	Monoazole
Ritanserin	87051-43-2	O=C2C(=C(N=C1N2(C=CS1))C)CCN5CCC(=C(c3ccc(F)cc3)c4ccc(F)cc4)CC5	inactive	NA	Monoazole

Ritonavir	155213-67-5	O=C(OCc1cncs1)NC(Cc2ccccc2)C(O)CC(NC(=O)C(NC(=O)N(C)Cc3nc(C(C)C)sc3)C(C)C)Cc4ccccc4	inactive	NA	Monoazole
Romazarit	109543-76-2	O=C(O)C(OCc1oc(nc1C)c2ccc(cc2)Cl)(C)C	inactive	NA	Monoazole
Succinylsulfathiazole	116-43-8	O=C(O)CCC(=O)Nc1ccc(cc1)S(=O)(=O)Nc2nccs2	inactive	NA	Monoazole
Sulfamethoxazole	723-46-6	O=S(=O)(Nc1noc(c1)C)c2ccc(N)cc2	inactive	NA	Monoazole
Sulfathiazole	72-14-0	O=S(=O)(Nc1nccs1)c2ccc(N)cc2	inactive	NA	Monoazole
Sulfisoxazole	127-69-5	O=S(=O)(Nc1onc(c1C)C)c2ccc(N)cc2	inactive	NA	Monoazole
Talipexole	101626-70-4	n1c(N)sc2c1CCN(CC=C)CC2	inactive	NA	Monoazole
Thiamethoxam	153719-23-4	O=[N+]([O-])N=C1N(C)COCN1Cc2cnc(Cl)s2	inactive	NA	Monoazole
Thiamine thiozole	137-00-8	OCCc1c(ncs1)C	inactive	NA	Monoazole
Tiotidine	69014-14-8	N#CN=C(NC)NCCSCc1nc(N=C(N)N)sc1	inactive	NA	Monoazole
Valdecoxib	181695-72-7	O=S(=O)(N)c1ccc(cc1)c2c(noc2C)c3ccccc3	inactive	NA	Monoazole

Zonisamide	68291-97-4	O=S(=O)(N)Cc1noc2ccccc12	inactive	NA		Monoazole
Zoxazolamine	61-80-3	n1c(oc2ccc(cc12)Cl)N	inactive	NA		Monoazole
Danazol	17230-88-5	C#CC4(O)(CCC3C5CCC2=Cc1oncc1CC2(C)C5(CCC34(C)))	active antagonist		57.71097	Monoazole
N-tert-Butyl-2-benzothiazole sulfenamide	95-31-8	n2c1ccccc1sc2SNC(C)(C)C	active antagonist		42.36881	Monoazole
Phosalone	2310-17-0	O=C2Oc1cc(ccc1N2CSP(OCC)(OCC)=S)Cl	active antagonist		51.5908	Monoazole
2-Amino-4-(p-nitrophenyl)thiazole	2104-09-8	O=[N+]([O-])c1ccc(cc1)c2nc(N)sc2	active agonist		9.787152	Monoazole
2-Amino-4-phenylthiazole hydrobromide hydrate	52253-69-7	n1c(N)scc1c2ccccc2	active agonist		6.332839	Monoazole
2-Amino-5,6-dimethylbenzothiazole	29927-08-0	n2c1cc(c(cc1sc2N)C)C	active agonist		1.230887	Monoazole
2-Amino-6-ethoxybenzothiazole	94-45-1	n2c1ccc(OCC)cc1sc2N	active agonist		1.708886	Monoazole
2-Amino-6-methoxybenzothiazole	1747-60-0	n2c1ccc(OC)cc1sc2N	active agonist		6.600727	Monoazole
2-Aminobenzothiazole	136-95-8	n1c(N)sc2ccccc12	active agonist		31.19954	Monoazole
Arotinolol hydrochloride	68377-91-3	O=C(N)c2ccc(c1nc(sc1)SCC(O)CNC(C)(C)C)s2	active agonist		22.38381	Monoazole

Cefodizime sodium	86329-79-5	O=C(O)C4=C(CSc1nc(c(CC(=O)O)s1)C)CSC3N4(C(=O)C3(NC(=O)C(=NOC)c2nc(N)sc2))	active agonist		29.84929	Monoazole
Diamthazole	95-27-2	n2c1ccc(OCCN(CC)CC)cc1sc2N(C)C	active agonist		22.38381	Monoazole
Fanetizole	79069-94-6	n1c(csc1NCCc2ccccc2)c3ccccc3	active agonist		1.496009	Monoazole
Febuxostat	144060-53-7	N#Cc1cc(ccc1(OCC(C)C))c2nc(c(C(=O)O)s2)C	active agonist		0.036163	Monoazole
Leflunomide	75706-12-6	O=C(Nc1ccc(cc1)C(F)(F)F)c2cnoc2C	active agonist		1.800247	Monoazole
Methabenzthiazuron	18691-97-9	O=C(NC)N(c2nc1ccccc1s2)C	active agonist		2.087333	Monoazole
Nitazoxanide	55981-09-4	O=C(Oc2cccc2(C(=O)Nc1ncc([N+](=O)[O-])s1))C	active agonist		0.455268	Monoazole
Parecoxib sodium	198470-85-8	O=C(NS(=O)(=O)c1ccc(cc1)c2c(noc2C)c3ccccc3)CC	active agonist		26.60321	Monoazole
Riluzole	1744-22-5	FC(F)(F)Oc1ccc2nc(N)sc2(c1)	active agonist		9.439176	Monoazole
Tenonitrozole	3810-35-3	O=C(Nc1ncc([N+](=O)[O-])s1)c2cccs2	active agonist		0.595572	Monoazole
Tioxidazole	61570-90-9	O=C(OC)Nc2nc1ccc(OCCC)cc1s2	active agonist		0.87418	Monoazole
(3,5-Dimethyl-1H-	85264-33-1	OCn1nc(cc1C)C	inactive	NA		Diazole

pyrazol-1-yl)methanol					
1,2-Dimethyl-5-nitroimidazole	551-92-8	O=[N+]([O-])c1cnc(n1C)C	inactive	NA	Diazole
1-Benzyl-3-methylimidazolium chloride	36443-80-8	c1ccc(cc1)Cn2cc[n+](c2)C	inactive	NA	Diazole
1-Hexyl-3-methylimidazolium chloride, 1-Hexyl-3-methylimidazolium hexafluorophosphate, 1-Hexyl-3-methylimidazolium tetrafluoroborate, 1-Hexyl-3-methylimidazolium trifluoromethanesulfonate	171058-17-6, 304680-35-1, 244193-50-8, 460345-16-8	c1c[n+](cn1CCCCCC)C	inactive	NA	Diazole
2',3'-Dideoxyadenosine	4097-22-7	OCC3OC(n2cnc1c(ncnc12)N)CC3	inactive	NA	Diazole
2-Aminobenzimidazole	934-32-7	n1c(N)[nH]c2ccccc12	inactive	NA	Diazole
4,5-Diphenylimidazole	668-94-0	n2c[nH]c(c1ccccc1)c2c3ccccc3	inactive	NA	Diazole
4-Dimethylaminoantipyrine	58-15-1	O=C2C(=C(N(N2(c1ccccc1))C)C)N(C)C	inactive	NA	Diazole
4-Methylimidazole	822-36-6	n1c[nH]cc1C	inactive	NA	Diazole

4-Methylpyrazole hydrochloride	56010-88-9	n1cc(c[nH]1)C	inactive	NA	Diazole
5-(5-Nitro-2-furyl)-1,3,4-oxadiazole-2-ol	2122-86-3	O=[N+]([O-])c1oc(cc1)c2nnc(O)o2	inactive	NA	Diazole
5-Chloro-2-methyl-1H-benzimidazole	2818-69-1	n1c2ccc(cc2([nH]c1C))Cl	inactive	NA	Diazole
5-Nitroindazole	5401-94-5	O=[N+]([O-])c1ccc2[nH]ncc2(c1)	inactive	NA	Diazole
6-Nitrobenzimidazole	94-52-0	O=[N+]([O-])c1ccc2[nH]cnc2(c1)	inactive	NA	Diazole
8-Bromotheophylline	10381-75-6	O=C2c1[nH]c(nc1N(C(=O)N2C)C)Br	inactive	NA	Diazole
Acetazolamide	59-66-5	O=C(Nc1nnc(s1)S(=O)(=O)N)C	inactive	NA	Diazole
Acibenzolar-S-methyl	135158-54-2	O=C(c2cccc1nnsc12)SC	inactive	NA	Diazole
Acyclovir	59277-89-3	O=C2NC(=Nc1c2(ncn1COCCO))N	inactive	NA	Diazole
Adenosine	58-61-7	OCC3OC(n2cnc1c(ncnc12)N)C(O)C3(O)	inactive	NA	Diazole
Adenosine-5-phosphate	61-19-8	O=P(O)(O)OCC3OC(n2cnc1c(ncnc12)N)C(O)C3(O)	inactive	NA	Diazole
Allopurinol	315-30-0	Oc1ncnc2[nH]ncc12	inactive	NA	Diazole
Alosetron hydrochloride	122852-69-1	O=C3c2c1cccc1n(c2CCN3Cc4[nH]cnc4C)C	inactive	NA	Diazole

Alpidem	82626-01-5	O=C(N(CCC)CCC)Cc2c(nc1ccc(cn12)Cl)c3ccc(cc3)Cl	inactive	NA	Diazole
Ampyrone	83-07-8	O=C1C(N)=C(N(N1c2ccccc2)C)C	inactive	NA	Diazole
Anagrelide hydrochloride	58579-51-4	Oc2nc3Nc1ccc(c(c1Cn3(c2))Cl)Cl	inactive	NA	Diazole
Atipamezole	104054-27-5	n1c[nH]c(c1)C3(Cc2ccccc2C3)CC	inactive	NA	Diazole
Bendamustine	16506-27-7	O=C(O)CCCc2nc1cc(ccc1n2C)N(CCCl)CCCl	inactive	NA	Diazole
Bendazac sodium	23255-99-4	O=C(O)COc1nn(c2ccccc12)Cc3ccccc3	inactive	NA	Diazole
Benperidol	2062-84-2	O=C2Nc1ccccc1N2C4CCN(CCCC(=O)c3ccc(F)cc3)CC4	inactive	NA	Diazole
Benzimidazole	51-17-2	n1c[nH]c2ccccc12	inactive	NA	Diazole
Benznidazole	22994-85-0	O=C(NCc1ccccc1)Cn2ccnc2[N+](=O)[O-]	inactive	NA	Diazole
Benzydamine hydrochloride	132-69-4	n2c(OCCCN(C)C)c1ccccc1n2Cc3ccccc3	inactive	NA	Diazole
Bretazenil	84379-13-5	O=C(OC(C)(C)C)c1ncn2c4cccc(c4(C(=O)N3CCCC3(c12)))Br	inactive	NA	Diazole
Caffeine, Caffeine, citrated	58-08-2, 69-22-7	O=C2c1c(nccn1C)N(C(=O)N2C)C	inactive	NA	Diazole

Carbimazole	22232-54-8	O=C(OCC)N1C=CN(C1=S)C	inactive	NA	Diazole
Cefoselis sulfate	122841-12-7	O=C(O)C4=C(C[n+]1ccc(N)n1CCO)CSC3N4(C(=O)C3(NC(=O)C(=NOC)c2nc(N)sc2))	inactive	NA	Diazole
Cefozopran hydrochloride	113981-44-5	O=C(O)C5=C(C[n+]2ccn1ncccc12)CSC4N5(C(=O)C4(NC(=O)C(=NOC)c3nc(N)sn3))	inactive	NA	Diazole
Chlormidazole	3689-76-7	n2c1ccccc1n(c2C)Cc3ccc(cc3)Cl	inactive	NA	Diazole
Cimetidine	51481-61-9	N#CN=C(NC)NCCSCc1nc[nH]c1C	inactive	NA	Diazole
Cladribine	4291-63-8	OCC3OC(n2cnc1c(nc(nc12)Cl)N)CC3(O)	inactive	NA	Diazole
Clemizole hydrochloride	1163-36-6	n2c1ccccc1n(c2CN3CCCC3)Cc4ccc(cc4)Cl	inactive	NA	Diazole
Dacarbazine	4342-03-4	O=C(N)c1[nH]cnc1(N=NN(C)C)	inactive	NA	Diazole
Dazmegrel	76894-77-4	O=C(O)CCn2c1ccccc1c(c2C)Cn3cncc3	inactive	NA	Diazole
Dazoxiben	78218-09-4	O=C(O)c2ccc(OCCn1cncc1)cc2	inactive	NA	Diazole
Deracoxib	169590-41-4	O=S(=O)(N)c1ccc(cc1)n3nc(cc3(c2ccc(O	inactive	NA	Diazole

		C)c(F)c2))C(F)F			
Dimetilan	644-64-4	O=C(Oc1nn(C(=O)N(C)C)c(c1)C)N(C)C	inactive	NA	Diazole
Dipyrone	68-89-3	O=C1C(=C(N(N1c2ccccc2)C)C)N(C)CS(=O)(=O)O	inactive	NA	Diazole
Domperidone	57808-66-9	O=C2Nc1cccc1N2CCCN5CCC(N4C(=O)Nc3cc(ccc34)Cl)CC5	inactive	NA	Diazole
Doramapimod	285983-48-4	O=C(Nc2ccc(OCCN1CCOCC1)c3ccccc23)Nc5cc(nn5(c4ccc(cc4)C))C(C)(C)C	inactive	NA	Diazole
Doxofylline	69975-86-6	O=C3c1c(ncn1CC2OCCO2)N(C(=O)N3C)C	inactive	NA	Diazole
Droperidol	548-73-2	O=C2Nc1cccc1N2C3=CCN(CC3)CCCC(=O)c4ccc(F)cc4	inactive	NA	Diazole
Dyphylline	479-18-5	O=C2c1c(ncn1CC(O)CO)N(C(=O)N2C)C	inactive	NA	Diazole
Enoximone	77671-31-9	O=C2NC(C(=O)c1ccc(cc1)SC)=C(N2)C	inactive	NA	Diazole
Enprofylline	41078-02-8	O=C1NC(=O)N(c2nc[nH]c12)CCC	inactive	NA	Diazole
Ensulizole	27503-81-7	O=S(=O)(O)c1ccc2[nH]c(n	inactive	NA	Diazole

		c2(c1))c3ccccc3			
Entecavir	142217-69-4	O=C3NC(=Nc1c3(ncn1C2C(=C)C(CO)C(O)C2))N	inactive	NA	Diazole
Epirizole	18694-40-1	n1c(OC)cc(nc1n2nc(cc2(OC))C)C	inactive	NA	Diazole
Eprosartan mesylate	144143-96-4	O=C(O)c1ccc(cc1)Cn2c(cnc2CCCC)C=C(C(=O)O)Cc3cccs3	inactive	NA	Diazole
Etandazole	22668-01-5	O=C(NCCO)Cn1ccnc1[N+](=O)[O-]	inactive	NA	Diazole
Etofylline	519-37-9	O=C2c1c(ncn1CCO)N(C(=O)N2C)C	inactive	NA	Diazole
Etofylline clofibrate	54504-70-0	O=C(OCCn2cnc1c2(C(=O)N(C(=O)N1C)C))C(Oc3ccc(cc3)Cl)(C)C	inactive	NA	Diazole
Etomidate	33125-97-2	O=C(OCC)c1cncn1C(c2ccccc2)C	inactive	NA	Diazole
Etridiazole	2593-15-9	n1c(nsc1OCC)C(Cl)(Cl)Cl	inactive	NA	Diazole
FD&C Yellow 5	1934-21-0	O=C(O)C=2NN(C(=O)C=2(N=Nc1ccc(cc1)S(=O)(=O)O))c3ccc(cc3)S(=O)(=O)O	inactive	NA	Diazole
FR140423	151506-44-4	O=S(c1ccc(cc1)c3cc(nn3(c2ccc(OC)cc2))C(F)F)C	inactive	NA	Diazole

Fabesetron	129300-27-2	O=C3n2c1ccccc1c(c2CCC3Cc4nc[nH]c4C)C	inactive	NA	Diazole
Famciclovir	104227-87-4	O=C(OCC(COC(=O)C)CCn2cnc1cnc(nc12)N)C	inactive	NA	Diazole
Famprofazone	22881-35-2	O=C2C(=C(N(N2(c1ccccc1))C)CN(C)C(C)Cc3ccccc3)C(C)C	inactive	NA	Diazole
Fenethylline hydrochloride	1892-80-4	O=C3c1c(ncn1CCNC(C)Cc2ccccc2)N(C(=O)N3C)C	inactive	NA	Diazole
Fludarabine	21679-14-1	Fc1nc(N)c2ncn(c2(n1))C3OC(CO)C(O)C3(O)	inactive	NA	Diazole
Flufenacet	142459-58-3	O=C(N(c1ccc(F)cc1)C(C)C)COc2nnc(C(F)(F)F)s2	inactive	NA	Diazole
Flumazenil	78755-81-4	O=C(OCC)c1ncn2c3ccc(F)cc3(C(=O)N(C)Cc12)	inactive	NA	Diazole
Ganciclovir	82410-32-0	O=C2NC(=Nc1c2(ncn1COC(CO)CO))N	inactive	NA	Diazole
Glybuzole	1492-02-0	O=S(=O)(Nc1nnc(C(C)(C)C)s1)c2ccccc2	inactive	NA	Diazole
Glysobuzole	3567-08-6	O=S(=O)(Nc1nnc(CC(C)C)s1)c2ccc(OC)cc2	inactive	NA	Diazole
Gonadorelin hydrochloride	51952-41-1	O=C(N)CNC(=O)C6N(C(	inactive	NA	Diazole

		=O)C(NC(=O)C(NC(=O)CNC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C1NC(=O)CC1)Cc2cnc[nH]2)Cc4c[nH]c3ccccc34)CO)Cc5ccc(O)cc5)CC(C)C)CCCNC(=N)N)CCC6			
Goserelin acetate	145781-92-6	O=C(N)NNC(=O)C6N(C(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C1NC(=O)CC1)Cc2nc[nH]c2)Cc4c[nH]c3ccccc34)CO)Cc5ccc(O)cc5)COC(C)(C)C)CC(C)C)CCCNC(=N)N)CCC6	inactive	NA	Diazole
Granisetron hydrochloride	107007-99-8	O=C(NC2CC1N(C)C(CCC1)C2)c3nn(c4ccccc34)C	inactive	NA	Diazole
Ibudilast	50847-11-5	O=C(c1c(nn2ccccc12)C(C)C)C(C)C	inactive	NA	Diazole
Imidazole	288-32-4	n1cc[nH]c1	inactive	NA	Diazole
Imiloxan	81167-16-0	n1ccn(c1CC2Oc3ccccc3(OC2))CC	inactive	NA	Diazole

Imiquimod	99011-02-6	n1cn(c3c1c(nc2ccccc23)N)CC(C)C	inactive	NA	Diazole
Inosine	58-63-9	Oc3ncnc1c3(ncn1C2OC(CO)C(O)C2(O))	inactive	NA	Diazole
Iodoantipyrine	129-81-7	O=C2C(=C(N(N2(c1ccccc1))C)C)I	inactive	NA	Diazole
Iomazenil	127985-21-1	O=C(OCC)c1ncn2c3cccc(c3(C(=O)N(C)Cc12))I	inactive	NA	Diazole
Ipronidazole	14885-29-1	O=[N+]([O-])c1cnc(n1C)C(C)C	inactive	NA	Diazole
Isbufylline	90162-60-0	O=C2c1c(ncn1CC(C)C)N(C(=O)N2C)C	inactive	NA	Diazole
Isradipine	75695-93-1	O=C(OC)C3=C(NC(=C(C(=O)OC(C)C)C3(c2cccc1nonc12))C)C	inactive	NA	Diazole
Itasetron	123258-84-4	O=C2Nc1ccccc1N2C(=O)NC4CC3N(C)C(CC3)C4	inactive	NA	Diazole
Ketoconazole	65277-42-1	O=C(N5CCN(c4ccc(OCC1OC(OC1)(c2ccc(cc2Cl)Cl)Cn3cncc3)cc4)CC5)C	inactive	NA	Diazole
Kinetin	525-79-1	N=1C=NC=3C=1N=CNC=3(NCc2occc2)	inactive	NA	Diazole
Leuprolide acetate salt	74381-53-6	O=C6NC(C(=O)NC(C(=O)NC(C(=O	inactive	NA	Diazole

		)NC(C(=O)NC(C(=O)NC(C(=O)NC(C(=O)NC(C(=O)N1CCCC1(C(=O)NCC))CCCNC(=N)N)CC(C)C)CC(C)C)Cc2ccc(O)cc2)CO)Cc4c[nH]c3ccccc34)Cc5cnc[nH]5)CC6			
Lisofylline	6493-06-7	O=C2c1c(ncn1C)N(C(=O)N2CCCCC(O)C)C	inactive	NA	Diazole
Lonazolac	53808-88-1	O=C(O)Cc2cn(nc2(c1ccc(cc1)Cl))c3ccccc3	inactive	NA	Diazole
Lonidamine	50264-69-2	O=C(O)c1nn(c2ccccc12)Cc3ccc(cc3Cl)Cl	inactive	NA	Diazole
Mebendazole	31431-39-7	O=C(OC)Nc1nc2cc(ccc2([nH]1))C(=O)c3ccccc3	inactive	NA	Diazole
Methazolamide	554-57-4	O=C(N=C1N(N=C(S1)S(=O)(=O)N)C)C	inactive	NA	Diazole
Methimazole	60-56-0	C1=CN(C(N1)=S)C	inactive	NA	Diazole
Metomidate hydrochloride	35944-74-2	O=C(OC)c1cncn1C(c2ccccc2)C	inactive	NA	Diazole
Metronidazole	443-48-1	O=[N+]([O-])c1cnc(n1CCO)C	inactive	NA	Diazole
Minodronic acid	180064-38-4	O=P(O)(O)C(O)(Cc2cnc	inactive	NA	Diazole

		1ccccn12)P(=O)(O)O			
Mizolastine	108612-45-9	O=C1C=CN=C(N1)N(C)C5CCN(c3nc2ccccc2n3Cc4ccc(F)cc4)CC5	inactive	NA	Diazole
Molsidomine	25717-80-0	O=C(OCC)Nc1on[n+](c1)N2CCOCC2	inactive	NA	Diazole
N-Benzyladenine	1214-39-7	N=1C=NC=3C=1N=CNC=3(NCc2ccccc2)	inactive	NA	Diazole
Nifenazone	2139-47-1	O=C(NC=2C(=O)N(c1ccccc1)N(C=2C)C)c3cnccc3	inactive	NA	Diazole
Nitrefazole	21721-92-6	O=[N+]([O-])c1nc(n(c1)c2ccc(cc2)[N+](=O)[O-])C	inactive	NA	Diazole
Nizofenone	54533-86-7	O=C(c1ccccc1Cl)c2cc(ccc2n3ccnc3CN(CC)CC)[N+](=O)[O-]	inactive	NA	Diazole
O6-Methylguanine	20535-83-5	n1c[nH]c2c1nc(nc2(OC))N	inactive	NA	Diazole
Ondansetron, Ondansetron hydrochloride dihydrate	99614-02-5, 103639-04-9	O=C3c2c1ccccc1n(c2CCC3Cn4ccnc4C)C	inactive	NA	Diazole
Ornidazole	16773-42-5	O=[N+]([O-])c1cnc(n1CC(O)CCl)C	inactive	NA	Diazole
Oxfendazole	53716-50-0	O=C(OC)Nc1nc2cc(ccc2([nH]1))S(=O)c3ccccc3	inactive	NA	Diazole

Oxibendazole	20559-55-1	O=C(OC)Nc1nc2cc(OCCC)ccc2([nH]1)	inactive	NA	Diazole
Oxypurinol	2465-59-0	O=C1NC(=O)c2c[nH]nc2(N1)	inactive	NA	Diazole
Ozagrel hydrochloride	78712-43-3	O=C(O)C=Cc1ccc(cc1)Cn2cncc2	inactive	NA	Diazole
Penciclovir	39809-25-1	O=C2N=C(N)Nc1c2(ncn1CCC(CO)CO)	inactive	NA	Diazole
Pentoxifylline	6493-05-6	O=C2c1c(ncn1C)N(C(=O)N2CCCCC(=O)C)C	inactive	NA	Diazole
Pilocarpine	92-13-7	O=C2OCC(Cc1cncn1C)C2CC	inactive	NA	Diazole
Pimozide	2062-78-4	O=C2Nc1cccccc1N2C5CCN(CCCC(c3ccc(F)cc3)c4ccc(F)cc4)CC5	inactive	NA	Diazole
Procodazole	23249-97-0	O=C(O)CCc1nc2ccccc2([nH]1)	inactive	NA	Diazole
Propentofylline	55242-55-2	O=C2c1c(ncn1CCC)N(C(=O)N2CCCCCC(=O)C)C	inactive	NA	Diazole
Propyphenazone	479-92-5	O=C2C(=C(N(N2(c1ccccc1))C)C)C(C)C	inactive	NA	Diazole
Proxazole citrate	132-35-4	n1oc(nc1C(c2ccccc2)CC)CCCN(CC)CC	inactive	NA	Diazole
Proxyphylline	603-00-9	O=C2c1c(ncn1CC(O)C)N(C(=O)N2C)C	inactive	NA	Diazole

Pyraflufen-ethyl	129630-19-9	O=C(OCC)COc1cc(c(F)cc1Cl)c2nn(c(OC(F)F)c2Cl)C	inactive	NA	Diazole
Pyrazophos	13457-18-6	O=C(OCC)c1cn2nc(OP(OCC)(OCC)=S)cc2(nc1C)	inactive	NA	Diazole
Ramifenazone	3615-24-5	O=C2C(NC(C)C)=C(N(N2(c1ccccc1))C)C	inactive	NA	Diazole
Ramosetron hydrochloride	132907-72-3	O=C(c1cn(c2ccccc12)C)C4Cc3nc[nH]c3CC4	inactive	NA	Diazole
Rifaximin	80621-81-4	O=C(OC4C(C)C(OC)C=COC1(Oc5c(C1(=O))c6c3nc2cc(ccn2c3c(NC(=O)C(=CC=CC(C)C(O)C(C)C(O)C4(C))C)c(O)c6(c(O)c5C))C)C)C	inactive	NA	Diazole
Rimonabant	168273-06-1	O=C(NN1CCCCC1)c4nn(c2ccc(cc2Cl)Cl)c(c3ccc(cc3)Cl)c4C	inactive	NA	Diazole
Ronidazole	7681-76-7	O=C(OCc1ncc(n1C)[N+](=O)[O-])N	inactive	NA	Diazole
Secnidazole	3366-95-8	O=[N+]([O-])c1cnc(n1CC(O)C)C	inactive	NA	Diazole
Sildenafil citrate	171599-83-0	O=C1NC(=Nc2c(nn(c12)C)CCC)c3cc(ccc3(OCC))S	inactive	NA	Diazole

		(=O)(=O)N4CCN(C)CC4			
Sulfaethidole	94-19-9	O=S(=O)(Nc1nnc(CC)s1)c2ccc(N)cc2	inactive	NA	Diazole
Sulfamethizole	144-82-1	O=S(=O)(Nc1nnc(C)s1)c2ccc(N)cc2	inactive	NA	Diazole
Sulfaphenazole	526-08-9	O=S(=O)(Nc2ccnn2(c1ccccc1))c3ccc(N)cc3	inactive	NA	Diazole
Sulmazole	73384-60-8	O=S(c3ccc(c1nc2ncccc2([nH]1))c(OC)c3)C	inactive	NA	Diazole
Telmisartan	144701-48-4	O=C(O)c6ccccc6(c1ccc(cc1)Cn2c3cc(cc(c3(nc2CCC))C)c5nc4ccccc4n5C)	inactive	NA	Diazole
Temozolomide	85622-93-1	O=C(N)c1ncn2C(=O)N(N=Nc12)C	inactive	NA	Diazole
Tenatoprazole	113712-98-4	O=S(c1nc2nc(OC)ccc2([nH]1))Cc3ncc(c(OC)c3C)C	inactive	NA	Diazole
Tenofovir disoproxil fumarate	202138-50-9	O=C(OCOP(=O)(OCOC(=O)OC(C)C)COC(C)Cn2cnc1c(ncnc12)N)OC(C)C	inactive	NA	Diazole
Theobromine	83-67-0	O=C1NC(=O)N(c2ncn(c12)C)C	inactive	NA	Diazole
Theophylline	58-55-9	O=C2c1[nH]cnc1N(C(=O)N2C)C	inactive	NA	Diazole

Timolol, Timolol maleate salt	26839-75-8, 26921-17-5	OC(COc1nsnc1N2CCOCC2)CNC(C)(C)C	inactive	NA		Diazole
Tinidazole	19387-91-8	O=[N+]([O-])c1cnc(n1CCS(=O)(=O)CC)C	inactive	NA		Diazole
Tizanidine	51322-75-9	N3=C(Nc1c2nsnc2(ccc1Cl))NCC3	inactive	NA		Diazole
Uric acid	69-93-2	O=C1NC(=O)C=2NC(=O)NC=2(N1)	inactive	NA		Diazole
cAMP	60-92-4	O=P3(O)(OCC4OC(n2cnc1c(ncnc12)N)C(O)C4(O3))	inactive	NA		Diazole
phenazone	60-80-0	O=C1C=C(N(N1c2ccccc2)C)C	inactive	NA		Diazole
1,3-Didecyl-2-methylimidazolium chloride	70862-65-6	c1c[n+](c(n1CCCCCCCCCC)C)CCCCCCCCCC	active antagonist		3.575745	Diazole
1-Decyl-3-methylimidazolium chloride, 1-Decyl-3-methylimidazolium hexafluorophosphate, 1-Decyl-3-methylimidazolium trifluoromethanesulfonate	171058-18-7, 362043-46-7, 412009-62-2	c1c[n+](cn1CCCCCCCCCC)C	active antagonist		13.37108	Diazole
1-Dodecyl-3-methylimidazolium hexafluorophosphate	219947-93-0, 244193-59-7, 404001-52-1	c1c[n+](cn1CCCCCCCCCCC)C	active antagonist		3.559686	Diazole

sphate, 1-Dodecyl-3-methylimidazolium tetrafluoroborate, 1-Dodecyl-3-methylimidazolium trifluoromethanesulfonate					
1-Hexadecyl-3-methylimidazolium chloride, 1-Hexadecyl-3-methylimidazolium tetrafluoroborate	61546-01-8, 244193-64-4	c1c[n+](cn1CCCCCCCCCCCCCCCC)C	active antagonist	7.014788	Diazole
1-Methyl-3-octadecylimidazolium hexafluorophosphate	219947-96-3	c1c[n+](cn1CCCCCCCCCCCCCCCCC)C	active antagonist	3.164232	Diazole
1-Methyl-3-octylimidazolium chloride, 1-Methyl-3-octylimidazolium hexafluorophosphate, 1-Methyl-3-octylimidazolium tetrafluoroborate, 1-Methyl-3-octylimidazolium trifluoromethanesulfonate	64697-40-1, 304680-36-2, 244193-52-0, 403842-84-2	c1c[n+](cn1CCCCCCCC)C	active antagonist	42.25602	Diazole

1-Methyl-3-tetradecylimidazolium chloride	171058-21-2	c1c[n+](cn1CCCCCCCCCCCCCC)C	active antagonist	4.446764	Diazole
1-[2-(Trifluoromethyl)phenyl]-1H-imidazole	25371-96-4	FC(F)(F)c1ccccc1n2cncc2	active antagonist	6.943857	Diazole
6-Thioguanine	154-42-7	n1c[nH]c2c1N=C(N)NC2=S	active antagonist	0.470732	Diazole
Bifonazole	60628-96-8	n1ccn(c1)C(c2ccccc2)c3ccc(cc3)c4ccccc4	active antagonist	0.87418	Diazole
Butoconazole nitrate	64872-77-1	n1ccn(c1)CC(CCc2ccc(cc2)Cl)Sc3c(cccc3Cl)Cl	active antagonist	13.76149	Diazole
Climbazole	38083-17-9	O=C(C(Oc1ccc(cc1)Cl)n2cncc2)C(C)(C)C	active antagonist	13.33319	Diazole
Clofarabine	123318-82-1	FC3C(O)C(OC3(n2cnc1c(nc(nc12)Cl)N))CO	active antagonist	1.584653	Diazole
Econazole nitrate	24169-02-6	n1ccn(c1)CC(OCc2ccc(cc2)Cl)c3ccc(cc3Cl)Cl	active antagonist	0.679556	Diazole
Fadrozole hydrochloride	102676-31-3	N#Cc1ccc(cc1)C3n2cncc2CCC3	active antagonist	0.009023	Diazole
Imazalil, Imazalil sulfate	35554-44-0, 58594-72-2	n1ccn(c1)CC(OCC=C)c2ccc(cc2Cl)Cl	active antagonist	0.683702	Diazole
Imazodan	84243-58-3	O=C3NN=C(c1ccc(cc1)n2cncc2)CC3	active antagonist	7.215516	Diazole
Isoconazole	27523-40-6	n1ccn(c1)CC(OCc2c(cccc2Cl)Cl)c3ccc(cc3Cl)Cl	active antagonist	10.12353	Diazole

Lanoconazole	101530-10-3	N#CC(=C2SCC(c1ccccc1Cl)S2)n3cncc3	active antagonist	3.101707	Diazole
Liarozole hydrochloride	145858-50-0	n1ccn(c1)C(c2cccc(c2)Cl)c3ccc4nc[nH]c4(c3)	active antagonist	0.510819	Diazole
Luliconazole	187164-19-8	N#CC(=C2SCC(c1ccc(cc1Cl)Cl)S2)n3cncc3	active antagonist	1.298099	Diazole
Medetomidine hydrochloride	86347-15-1	n1c[nH]cc1C(c2cccc(c2C)C)C	active antagonist	29.84929	Diazole
Metrifudil	23707-33-7	OCC4OC(n2cnc1c(ncnc12)NCc3ccccc3C)C(O)C4(O)	active antagonist	5.955717	Diazole
Miconazole	22916-47-8	n1ccn(c1)CC(OCc2ccc(cc2Cl)Cl)c3ccc(cc3Cl)Cl	active antagonist	7.738655	Diazole
Oxiconazole nitrate	64211-46-7	N(OCc1ccc(cc1Cl)Cl)=C(c2ccc(cc2Cl)Cl)Cn3cncc3	active antagonist	2.660321	Diazole
Prochloraz	67747-09-5	O=C(n1cncc1)N(CCOc2c(cc(cc2Cl)Cl)Cl)CCC	active antagonist	1.383639	Diazole
Sertaconazole nitrate	99592-39-9	n1ccn(c1)CC(OCc3csc2c3(cccc2Cl))c4ccc(cc4Cl)Cl	active antagonist	5.073786	Diazole
Sulconazole nitrate	61318-91-0	n1ccn(c1)CC(c2ccc(cc2Cl)Cl)SCc3ccc(cc3)Cl	active antagonist	3.080814	Diazole
Timiperone	57648-21-2	O=C(c1ccc(F)cc1)CCCN4CCC(N2c3ccccc3(NC2=S))CC4	active antagonist	18.12456	Diazole

Tioconazole	65899-73-2	n1ccn(c1)CC(OCc2ccsc2Cl)c3ccc(cc3Cl)Cl	active antagonist	5.515701	Diazole
Triclabendazole	68786-66-3	n2c3cc(c(Oc1cccc(c1Cl)Cl)cc3([nH]c2SC))Cl	active antagonist	23.71014	Diazole
Triflumizole	68694-11-1	FC(F)(F)c2cc(ccc2(N=C(n1cncc1)COCCC))Cl	active antagonist	0.050406	Diazole
Zaldaride	109826-26-8	O=C2Nc1cccccc1N2C3CCN(CC3)CC6(OCc4ccccc4n5cccc56)C	active antagonist	29.84929	Diazole
6-Mercaptopurine	50-44-2	N1=CNC(c2nc[nH]c12)=S	active agonist	6.263652	Diazole
Abacavir	136470-78-5	OCC4C=CC(n2cnc1c(nc(nc12)N)NC3CC3)C4	active agonist	2.660321	Diazole
Acadesine	2627-69-2	O=C(N)c1ncn(c1(N))C2OC(CO)C(O)C2(O)	active agonist	24.63773	Diazole
Adenine	73-24-5	n1cnc2[nH]cnc2(c1N)	active agonist	5.515701	Diazole
Albendazole	54965-21-8	O=C(OC)Nc1nc2cc(ccc2([nH]1))SCCC	active agonist	1.180317	Diazole
Azathioprine	446-86-6	O=[N+]([O-])c1ncn(c1Sc3ncnc2nc[nH]c23)C	active agonist	10.42257	Diazole
Cambendazole	26097-80-3	O=C(OC(C)C)Nc1ccc2[nH]c(nc2(c1))c3ncsc3	active agonist	3.480172	Diazole
Carbendazim	10605-21-7	O=C(OC)Nc1nc2ccccc2([nH]1)	active agonist	43.27713	Diazole

Fenbendazole	43210-67-9	O=C(OC)Nc1nc2cc(ccc2([nH]1))Sc3ccccc3	active agonist		1.051959	Diazole
Furafylline	80288-49-9	O=C2c1[nH]c(nc1N(C(=O)N2C)Cc3occc3)C	active agonist		29.84929	Diazole
Lansoprazole	103577-45-3	O=S(c1nc2ccccc2([nH]1))Cc3nccc(OCC(F)(F)F)c3C	active agonist		12.34812	Diazole
N-Benzyl-9-(tetrahydro-2H-pyran-2-yl)adenine	2312-73-4	n3cnc1c(ncn1C2OCCCC2)c3NCc4ccccc4	active agonist		2.730602	Diazole
Nocodazole	31430-18-9	O=C(OC)Nc1nc2cc(ccc2([nH]1))C(=O)c3cccs3	active agonist		2.511505	Diazole
Parbendazole	14255-87-9	O=C(OC)Nc1nc2cc(ccc2([nH]1))CCCC	active agonist		2.181049	Diazole
Pifexole	27199-40-2	n1oc(nc1c2ccncc2)c3ccccc3Cl	active agonist		1.496009	Diazole
Thiabendazole	148-79-8	n1cscc1c2nc3ccccc3([nH]2)	active agonist		5.009364	Diazole
thioinosine	574-25-4	OCC3OC(n2cnc1c2(ncnc1S))C(O)C3(O)	active agonist		5.731488	Diazole
1,2,3-Benzotriazole	95-14-7	n1nc2ccccc2([nH]1)	inactive	NA		Triazole
1H-Benzotriazol-1-ol	2592-95-2	On2nnc1ccccc12	inactive	NA		Triazole
1H-Benzotriazole, 5-chloro-	94-97-3	n1nc2cc(ccc2([nH]1))Cl	inactive	NA		Triazole
Apafant	105219-56-5	O=C(N1CCOCC1)CCc4cc3C(=NCc2n	inactive	NA		Triazole

		nc(n2c3s4)C)c5ccccc5Cl			
Azaconazole	60207-31-0	n1cnn(c1)CC2(OCCO2)c3ccc(cc3Cl)Cl	inactive	NA	Triazole
Bumetrizole	3896-11-5	Oc1c(cc(cc1C(C)(C)C)C)n2nc3ccc(cc3(n2))Cl	inactive	NA	Triazole
Estazolam	29975-16-4	n1nc2n(c1)c4ccc(cc4(C(=NC2)c3ccccc3))Cl	inactive	NA	Triazole
Flucarbazone-sodium	181274-17-9	O=C(NS(=O)(=O)c1ccccc1(OC(F)(F)F))N2N=C(OC)N(C2(=O))C	inactive	NA	Triazole
Fluconazole	86386-73-4	Fc1ccc(c(F)c1)C(O)(Cn2ncnc2)Cn3ncnc3	inactive	NA	Triazole
Flumetsulam	98967-40-9	O=S(=O)(Nc1c(F)cccc1(F))c2nc3nc(ccn3(n2))C	inactive	NA	Triazole
Isazofos	42509-80-8	n1c(nn(c1Cl)C(C)C)OP(OCC)(OCC)=S	inactive	NA	Triazole
Itraconazole	84625-61-6	O=C1N(N=CN1c2ccc(cc2)N7CCN(c6ccc(OCC3OC(OC3)(c4ccc(cc4Cl)Cl)Cn5ncnc5)cc6)CC7)C(C)CC	inactive	NA	Triazole
Maraviroc	376348-65-1	O=C(NC(c1ccccc1)CCN2C4CCC2CC(n3c(nnc3C(C)C)C)C4)C5C	inactive	NA	Triazole

		CC(F)(F)CC5			
Octrizole	3147-75-9	Oc1ccc(cc1n2nc3ccccc3(n2))C(C)(C)CC(C)(C)C	inactive	NA	Triazole
Paclobutrazol	76738-62-0	OC(C(n1ncnc1)Cc2ccc(cc2)Cl)C(C)(C)C	inactive	NA	Triazole
Penoxsulam	219714-96-2	O=S(=O)(Nc1nc2c(OC)cnc(OC)n2(n1))c3c(OCC(F)F)cccc3C(F)(F)F	inactive	NA	Triazole
Rizatriptan	144034-80-0	n1cnn(c1)Cc2ccc3[nH]cc(c3(c2))CCN(C)C	inactive	NA	Triazole
Sulfentrazone	122836-35-5	O=C1N(N=C(N1C(F)F)C)c2cc(NS(=O)(=O)C)c(cc2Cl)Cl	inactive	NA	Triazole
Tazobactam sodium	89785-84-2	O=C(O)C3N1C(=O)CC1S(=O)(=O)C3(C)(Cn2nncc2)	inactive	NA	Triazole
Trapidil	15421-84-8	n1cnn2c1nc(cc2N(CC)CC)C	inactive	NA	Triazole
Trazodone hydrochloride	25332-39-2	O=C1N4C=CC=CC4(=NN1CCCN3CCN(c2cccc(c2)Cl)CC3)	inactive	NA	Triazole
Triazophos	24017-47-8	n1cn(nc1OP(OCC)(OCC)=S)c2ccccc2	inactive	NA	Triazole
Triticonazole	131983-72-7	OC2(C(=Cc1ccc(cc1)Cl)CCC2(C)(C))(Cn3ncnc3)	inactive	NA	Triazole

Voriconazole	137234-62-9	Fc1ccc(c(F)c1)C(O)(Cn2ncnc2)C(c3ncnc3(F))C	inactive	NA	Triazole
Zaprinast	37762-06-4	O=C1NC(=Nc2[nH]nnc12)c3ccccc3(OCCC)	inactive	NA	Triazole
propoxycarbazone-sodium	181274-15-7	O=C(OC)c1ccccc1S(=O)(=O)NC(=O)N2N=C(OCCC)N(C2(=O))C	inactive	NA	Triazole
Anastrozole	120511-73-1	N#CC(c1cc(cc(c1)C(C#N)(C)C)Cn2ncnc2)(C)C	active antagonist	0.29668	Triazole
Carfentrazone-ethyl	128639-02-1	O=C(OCC)C(Cc1cc(c(F)cc1Cl)N2N=C(N(C2(=O))C(F)F)C)Cl	active antagonist	54.48267	Triazole
Etaconazole	60207-93-4	n1cnn(c1)CC2(OCC(O2)CC)c3ccc(cc3Cl)Cl	active antagonist	26.27797	Triazole
Fenbuconazole	114369-43-6	N#CC(c1ccccc1)(Cn2ncnc2)CCc3ccc(cc3)Cl	active antagonist	33.082	Triazole
Flutriafol	76674-21-0	Fc1ccc(cc1)C(O)(c2ccccc2(F))Cn3ncnc3	active antagonist	48.55773	Triazole
Ipconazole	125225-28-7	OC3(Cn1ncnc1)(C(Cc2ccc(cc2)Cl)CCC3(C(C)C))	active antagonist	3.967251	Triazole
Israpafant	117279-73-9	n1nc4n(c1C)c3c(cc(CCc2ccc(cc2)CC(C)C)s3)C(=NC4C)c5ccccc5Cl	active antagonist	23.71014	Triazole

Letrozole	112809-51-5	N#Cc1ccc(cc1)C(c2ccc(C#N)cc2)n3ncnc3	active antagonist	0.017325	Triazole
Myclobutanil	88671-89-0	N#CC(c1ccc(cc1)Cl)(Cn2ncnc2)CCCC	active antagonist	5.015372	Triazole
Penconazole	66246-88-6	n1cnn(c1)CC(c2ccc(cc2Cl)Cl)CCC	active antagonist	14.22083	Triazole
Propiconazole	60207-90-1	n1cnn(c1)CC2(OCC(O2)CCC)c3ccc(cc3Cl)Cl	active antagonist	31.19954	Triazole
Tebuconazole	107534-96-3	OC(Cn1ncnc1)(CCc2ccc(cc2)Cl)C(C)(C)C	active antagonist	33.41299	Triazole
Tetraconazole	112281-77-3	FC(F)C(F)(F)OCC(c1ccc(cc1Cl)Cl)Cn2ncnc2	active antagonist	7.36097	Triazole
Triadimefon	43121-43-3	O=C(C(Oc1ccc(cc1)Cl)n2ncnc2)C(C)(C)C	active antagonist	54.42727	Triazole
Triadimenol	55219-65-3	OC(C(Oc1ccc(cc1)Cl)n2ncnc2)C(C)(C)C	active antagonist	25.70798	Triazole
2-(2H-Benzotriazol-2-yl)-4-methylphenol	2440-22-4	Oc1ccc(cc1n2nc3ccccc3(n2))C	active agonist	1.673448	Triazole
Alprazolam	28981-97-7	n4nc1n(c3ccc(cc3(C(=NC1)c2ccccc2))Cl)c4C	active agonist	4.216325	Triazole
Brotizolam	57801-81-7	n1nc3n(c1C)c2c(cc(Br)s2)C(=NC3)c4ccccc4Cl	active agonist	11.88322	Triazole
Etizolam	40054-69-1	n1nc3n(c1C)c2c(cc(CC)s2)	active agonist	23.71014	Triazole

Ribavirin	36791-04-5	C(=NC3)c4cccc4Cl O=C(N)c1ncn(n1)C2OC(CO)C(O)C2(O)	active agonist	2.564569	Triazole

References

1. Chan, H.J., K. Petrossian, and S. Chen, *Structural and functional characterization of aromatase, estrogen receptor, and their genes in endocrine-responsive and–resistant breast cancer cells.* The Journal of steroid biochemistry and molecular biology, 2016. **161**: p. 73-83.
2. Hong, Y., et al., *Molecular characterization of aromatase.* Annals of the New York Academy of Sciences, 2009. **1155**: p. 112.
3. Ghosh, D., J. Lo, and C. Egbuta, *Recent progress in the discovery of next generation inhibitors of aromatase from the structure–function perspective.* Journal of medicinal chemistry, 2016. **59**(11): p. 5131-5148.
4. Ghosh, D., et al., *Novel aromatase inhibitors by structure-guided design.* Journal of medicinal chemistry, 2012. **55**(19): p. 8464-8476.
5. Jiang, W. and D. Ghosh, *Motion and flexibility in human cytochrome p450 aromatase.* PloS one, 2012. **7**(2): p. e32565.
6. Ghosh, D., et al., *Structural basis for androgen specificity and oestrogen synthesis in human aromatase.* Nature, 2009. **457**(7226): p. 219-223.
7. Hong, Y., R. Rashid, and S. Chen, *Binding features of steroidal and nonsteroidal inhibitors.* Steroids, 2011. **76**(8): p. 802-806.
8. Adhikari, N., et al., *Combating breast cancer with non-steroidal aromatase inhibitors (NSAIs): Understanding the chemico-biological interactions through comparative SAR/QSAR study.* European journal of medicinal chemistry, 2017. **137**: p. 365-438.
9. Berman, H., K. Henrick, and H. Nakamura, *Announcing the worldwide protein data bank.* Nature Structural & Molecular Biology, 2003. **10**(12): p. 980-980.
10. *Protein Data Bank.*
11. Suvannang, N., et al., *Molecular docking of aromatase inhibitors.* Molecules, 2011. **16**(5): p. 3597-3617.
12. Ghosh, D., C. Egbuta, and J. Lo, *Testosterone complex and non-steroidal ligands of human aromatase.* The Journal of steroid biochemistry and molecular biology, 2018. **181**: p. 11-19.
13. Spinello, A., I. Ritacco, and A. Magistrato, *Recent advances in computational design of potent aromatase inhibitors: open-eye on endocrine-resistant breast cancers.* Expert opinion on drug discovery, 2019. **14**(10): p. 1065-1076.
14. *UniProt: The universal protein knowledgebase in 2021.* Nucleic Acids Research, 2021. **49**(D1): p. D480-D489.

15. Stocco, C., *Tissue physiology and pathology of aromatase.* Steroids, 2012. **77**(1-2): p. 27-35.
16. Evans, C.T., et al., *Isolation and characterization of a complementary DNA specific for human aromatase-system cytochrome P-450 mRNA.* Proceedings of the National Academy of Sciences, 1986. **83**(17): p. 6387-6391.
17. Honda, S.-i., N. Harada, and Y. Takagi, *Novel exon 1 of the aromatase gene specific for aromatase transcripts in human brain.* Biochemical and biophysical research communications, 1994. **198**(3): p. 1153-1160.
18. Mahendroo, M.S., et al., *Tissue-specific expression of human P-450AROM. The promoter responsible for expression in adipose tissue is different from that utilized in placenta.* Journal of Biological Chemistry, 1991. **266**(17): p. 11276-11281.
19. Simpson, E.R., et al., *Tissue-specific promoters regulate aromatase cytochrome P450 expression.* Clinical chemistry, 1993. **39**(2): p. 317-324.
20. Chumsri, S., et al., *Aromatase, aromatase inhibitors, and breast cancer.* The Journal of steroid biochemistry and molecular biology, 2011. **125**(1-2): p. 13-22.
21. Lu, Q., et al., *Expression of aromatase protein and messenger ribonucleic acid in tumor epithelial cells and evidence of functional significance of locally produced estrogen in human breast cancers.* Endocrinology, 1996. **137**(7): p. 3061-3068.
22. Santen, R., et al., *History of aromatase: saga of an important biological mediator and therapeutic target.* Endocrine reviews, 2009. **30**(4): p. 343-375.
23. Bulun, S., et al., *Regulation of aromatase expression in breast cancer tissue.* Annals of the New York Academy of Sciences, 2009. **1155**(1): p. 121-131.
24. Chayawan, C., et al., *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, 2020. **25**(3): p. 739.
25. Avendan, C. and C. Menendez, *Anticancer Drugs That Modulate Hormone Action.* Medicinal Chemistry of Anticancer Drugs, eds. Carmen Avendaño and José Carlos Menéndez. Amsterdam, Netherlands: Elsevier Science, 2015: p. 81-131.
26. Almeida, C.F., et al., *Estrogen receptor-positive (ER+) breast cancer treatment: Are multi-target compounds the next promising approach?* Biochemical pharmacology, 2020. **177**: p. 113989.
27. Carr, B.R., P.C. MacDonald, and E.R. Simpson, *The role of lipoproteins in the regulation of progesterone secretion by the human corpus luteum.* Fertility and Sterility, 1982. **38**(3): p. 303-311.
28. Miller, W.L., *Steroidogenesis: unanswered questions.* Trends in Endocrinology & Metabolism, 2017. **28**(11): p. 771-793.
29. Miller, W.L. and J.F. Strauss III, *Molecular pathology and mechanism of action of the steroidogenic acute regulatory protein, StARProceedings of Xth International Congress on Hormonal Steroids, Quebec, Canada, 17–21 June 1998.* The Journal of steroid biochemistry and molecular biology, 1999. **69**(1-6): p. 131-141.
30. Belfiore, C.J., et al., *Regulation of cytochrome P450scc synthesis and activity in the ovine corpus luteum.* The Journal of steroid biochemistry and molecular biology, 1994. **51**(5-6): p. 283-290.

31. Jones, M., *Aromatase--a brief overview.* Annu Rev Physiol, 2002. **64**: p. 93-127.
32. Barakat, R., et al., *Extra-gonadal sites of estrogen biosynthesis and function.* BMB reports, 2016. **49**(9): p. 488.
33. Fuentes, N. and P. Silveyra, *Estrogen receptor signaling mechanisms.* Advances in protein chemistry and structural biology, 2019. **116**: p. 135-170.
34. Simpson, E.R., *Sources of estrogen and their importance.* The Journal of steroid biochemistry and molecular biology, 2003. **86**(3-5): p. 225-230.
35. Matta, J., et al., *Estrogen receptor expression is associated with DNA repair capacity in breast cancer.* PloS one, 2016. **11**(3): p. e0152422.
36. Koos, R.D., *Minireview: putting physiology back into estrogens' mechanism of action.* Endocrinology, 2011. **152**(12): p. 4481-4488.
37. Voogt, J., *Control of hormone release during lactation.* Clinics in obstetrics and gynaecology, 1978. **5**(2): p. 435-455.
38. Schulster, M., A.M. Bernie, and R. Ramasamy, *The role of estradiol in male reproductive function.* Asian journal of andrology, 2016. **18**(3): p. 435.
39. Brodie, A.M., et al., *Inactivation of aromatase in vitro by 4-hydroxy-4-androstene-3, 17-dione and 4-acetoxy-4-androstene-3, 17-dione and sustained effects in vivo.* Steroids, 1981. **38**(6): p. 693-702.
40. SCHWARZEL, W.C., W.G. KRUGGEL, and H.J. BRODIE, *Studies on the mechanism of estrogen biosynthesis. VIII. The development of inhibitors of the enzyme system in human placenta.* Endocrinology, 1973. **92**(3): p. 866-880.
41. Sjoerdsma, A., *Suicide enzyme inhibitors as potential drugs.* Clinical Pharmacology & Therapeutics, 1981. **30**(1): p. 3-22.
42. Brodie, A. and B. Long, *Aromatase inhibition and inactivation.* Clinical cancer research, 2001. **7**(12): p. 4343s-4349s.
43. Brueggemeier, R.W., J.C. Hackett, and E.S. Diaz-Cruz, *Aromatase inhibitors in the treatment of breast cancer.* Endocrine reviews, 2005. **26**(3): p. 331-345.
44. Siraki, A.G., *Free radical metabolites in arylamine toxicity*, in *Advances in Molecular Toxicology.* 2013, Elsevier. p. 39-82.
45. Brodie, A.M., *Aromatease, its inhibitors and their use in breast cancer treatment.* Pharmacology & therapeutics, 1993. **60**(3): p. 501-515.
46. Samojlik, E., R. Santen, and S. Wells, *Adrenal suppression with aminoglutethimide. II. Differential effects of aminoglutethimide on plasma androstenedione and estrogen levels.* The Journal of Clinical Endocrinology & Metabolism, 1977. **45**(3): p. 480-487.
47. Dutta, U. and K. Pant, *Aromatase inhibitors: past, present and future in breast cancer therapy.* Medical Oncology, 2008. **25**(2): p. 113-124.
48. Fabian, C., *The what, why and how of aromatase inhibitors: hormonal agents for treatment and prevention of breast cancer.* International journal of clinical practice, 2007. **61**(12): p. 2051-2063.
49. Macedo, L.F., G. Sabnis, and A. Brodie, *Aromatase inhibitors and breast cancer.* Annals of the New York Academy of Sciences, 2009. **1155**(1): p. 162-173.
50. Hamadeh, I.S., et al., *Personalizing aromatase inhibitor therapy in patients with breast cancer.* Cancer treatment reviews, 2018. **70**: p. 47-55.

51. Jänicke, F., *Are all aromatase inhibitors the same? A review of the current evidence.* The breast, 2004. **13**(1): p. 10-18.
52. Buzdar, A.U., et al., *An overview of the pharmacology and pharmacokinetics of the newer generation aromatase inhibitors anastrozole, letrozole, and exemestane.* Cancer: Interdisciplinary International Journal of the American Cancer Society, 2002. **95**(9): p. 2006-2016.
53. Hong, S., et al., *The expanding use of third-generation aromatase inhibitors: what the general internist needs to know.* Journal of general internal medicine, 2009. **24**(2): p. 383-388.
54. Rani, S., et al., *A Review on Diverse Heterocyclic Compounds as the Privileged Scaffolds in Non-steroidal Aromatase Inhibitors.* Bioorganic Chemistry, 2021: p. 105017.
55. Ahmad, I., *Recent developments in steroidal and nonsteroidal aromatase inhibitors for the chemoprevention of estrogen-dependent breast cancer.* European journal of medicinal chemistry, 2015. **102**: p. 375-386.
56. Kang, H., et al., *Potent aromatase inhibitors and molecular mechanism of inhibitory action.* European journal of medicinal chemistry, 2018. **143**: p. 426-437.
57. Adhikari, N., et al., *Integrating regression and classification-based QSARs with molecular docking analyses to explore the structure-antiaromatase activity relationships of letrozole-based analogs.* Canadian Journal of Chemistry, 2017. **95**(12): p. 1285-1295.
58. Caporuscio, F., et al., *Structure-based design of potent aromatase inhibitors by high-throughput docking.* Journal of medicinal chemistry, 2011. **54**(12): p. 4006-4017.
59. Cai, J., et al., *Computational insights into inhibitory mechanism of azole compounds against human aromatase.* RSC advances, 2015. **5**(110): p. 90871-90880.
60. Kjeldsen, L.S., M. Ghisari, and E.C. Bonefeld-Jørgensen, *Currently used pesticides and their mixtures affect the function of sex hormone receptors and aromatase enzyme activity.* Toxicology and Applied Pharmacology, 2013. **272**(2): p. 453-464.
61. Andersen, H.R., et al., *Effects of currently used pesticides in assays for estrogenicity, androgenicity, and aromatase activity in vitro.* Toxicology and applied pharmacology, 2002. **179**(1): p. 1-12.
62. Hong, Y. and S. Chen, *Aromatase inhibitors: structural features and biochemical characterization.* Annals of the New York Academy of Sciences, 2006. **1089**(1): p. 237-251.
63. Awan, A. and K. Esfahani, *Endocrine therapy for breast cancer in the primary care setting.* Current Oncology, 2018. **25**(4): p. 285-291.
64. Augusto, T.V., et al., *Acquired resistance to aromatase inhibitors: where we stand!* Endocrine-related cancer, 2018. **25**(5): p. R283-R301.
65. Eicher, T., S. Hauptmann, and A. Speicher, *The chemistry of heterocycles: structures, reactions, synthesis, and applications*. 2013: John Wiley & Sons.
66. Aromí, G., et al., *Triazoles and tetrazoles: Prime ligands to generate remarkable coordination materials.* Coordination Chemistry Reviews, 2011. **255**(5-6): p. 485-546.

67. Zhao, H., et al., *In situ hydrothermal synthesis of tetrazole coordination polymers with interesting physical properties.* Chemical Society Reviews, 2008. **37**(1): p. 84-100.
68. Huang, J. and G. Yu, *Structural Engineering in Polymer Semiconductors with Aromatic N-Heterocycles.* Chemistry of Materials, 2021. **33**(5): p. 1513-1539.
69. Gao, H. and J.n.M. Shreeve, *Azole-based energetic salts.* Chemical Reviews, 2011. **111**(11): p. 7377-7436.
70. Wozniak, D.R. and D.G. Piercey, *Review of the current synthesis and properties of energetic pentazolate and derivatives thereof.* Engineering, 2020.
71. Xhanari, K. and M. Finšgar, *Organic corrosion inhibitors for aluminum and its alloys in chloride and alkaline solutions: a review.* Arabian Journal of Chemistry, 2019. **12**(8): p. 4646-4663.
72. Easton, M.E., et al., *Controlling the interface between salts, solvates, co-crystals, and ionic liquids with non-stoichiometric protic azolium azolates.* Crystal Growth & Design, 2020. **20**(4): p. 2608-2616.
73. Ahmed, F. and H. Xionga, *Recent developments in 1, 2, 3-triazole-based chemosensors.* Dyes and Pigments, 2020: p. 108905.
74. Fontana, G., *Current bioactive azole-containing natural products.* Current Bioactive Compounds, 2010. **6**(4): p. 284-308.
75. Gaware, V.M., et al., *Thiazolo-Triazole a nucleus possessing range of pharmacological activities: a review.* Der Pharmacia Lettre, 2010. **2**(2): p. 35-40.
76. Sykes, J.E. and M.G. Papich, *Antifungal drugs.* Canine and feline infectious diseases, 2013: p. 87-96.
77. Jørgensen, L.N. and T.M. Heick, *Azole Use in Agriculture, Horticulture, and Wood Preservation–Is It Indispensable?* Frontiers in Cellular and Infection Microbiology, 2021: p. 806.
78. Campoy, S. and J.L. Adrio, *Antifungals.* Biochemical pharmacology, 2017. **133**: p. 86-96.
79. Cai, W., et al., *Biodegradation of typical azole fungicides in activated sludge under aerobic conditions.* journal of environmental sciences, 2021. **103**: p. 288-297.
80. H Zhou, C. and Y. Wang, *Recent researches in triazole compounds as medicinal drugs.* Current medicinal chemistry, 2012. **19**(2): p. 239-280.
81. Perez-Cantero, A., et al., *Azole resistance mechanisms in Aspergillus: Update and recent advances.* International journal of antimicrobial agents, 2020. **55**(1): p. 105807.
82. Haroon, M., et al., *Synthesis, antioxidant, antimicrobial and antiviral docking studies of ethyl 2-(2-(arylidene) hydrazinyl) thiazole-4-carboxylates.* Zeitschrift für Naturforschung C, 2021.
83. Das, R., et al., *Recent developments in azole compounds as antitubercular agent.* Mini-Reviews in Organic Chemistry, 2019. **16**(3): p. 290-306.
84. Gonzalez, F.J. and H.V. Gelboin, *Human cytochromes P450: evolution and cDNA-directed expression.* Environmental Health Perspectives, 1992. **98**: p. 81-85.

85. Mermer, A., et al., *Synthesis, biological activity and structure activity relationship studies of novel conazole analogues via conventional, microwave and ultrasound mediated techniques.* Bioorganic Chemistry, 2018. **81**: p. 55-70.
86. Guerrero-Perilla, C., F.A. Bernal, and E. Coy-Barrera, *Insights into the interaction and binding mode of a set of antifungal azoles as inhibitors of potential fungal enzyme-based targets.* Molecular Diversity, 2018. **22**(4): p. 929-942.
87. Zarn, J.A., B.J. Brüschweiler, and J.R. Schlatter, *Azole fungicides affect mammalian steroidogenesis by inhibiting sterol 14 alpha-demethylase and aromatase.* Environmental health perspectives, 2003. **111**(3): p. 255-261.
88. Trösken, E.R., et al., *Comparative assessment of the inhibition of recombinant human CYP19 (aromatase) by azoles used in agriculture and as drugs for humans.* Endocrine research, 2004. **30**(3): p. 387-394.
89. Petkov, P., et al., *Mechanism-based categorization of aromatase inhibitors: a potential discovery and screening tool.* SAR and QSAR in Environmental Research, 2009. **20**(7-8): p. 657-678.
90. Marín-Luna, M., et al., *Theoretical and spectroscopic characterization of API-related azoles in solution and in solid state.* Current Pharmaceutical Design, 2020. **26**(38): p. 4847-4857.
91. Mori, A., *Structure-and Functionality-Based Molecular Design of Azoles and Thiophenes.* Bulletin of the Chemical Society of Japan, 2020. **93**(10): p. 1200-1212.
92. De Oliveira Silva, A., J. McQuade, and M. Szostak, *Recent advances in the synthesis and reactivity of isothiazoles.* Advanced Synthesis & Catalysis, 2019. **361**(13): p. 3050-3067.
93. Lauko, J., P.H. Kouwer, and A.E. Rowan, *1H-1, 2, 3-Triazole: From structure to function and catalysis.* Journal of Heterocyclic Chemistry, 2017. **54**(3): p. 1677-1699.
94. Narasimhamurthy, K.H., et al., *An overview of recent developments in the synthesis of substituted thiazoles.* ChemistrySelect, 2020. **5**(19): p. 5629-5656.
95. Reinus, B. and S.M. Kerwin, *Preparation and utility of N-alkynyl azoles in synthesis.* Molecules, 2019. **24**(3): p. 422.
96. Zyryanov, G.V., et al., *Rational synthetic methods in creating promising (hetero) aromatic molecules and materials.* Mendeleev Communications, 2020. **30**(5): p. 537-554.
97. Gnant, M., et al., *Duration of adjuvant aromatase-inhibitor therapy in postmenopausal breast cancer.* New England Journal of Medicine, 2021. **385**(5): p. 395-405.
98. Lepailleur, A., G. Poezevara, and R. Bureau, *Automated detection of structural alerts (chemical fragments) in (eco) toxicology.* Computational and structural biotechnology journal, 2013. **5**(6): p. e201302013.
99. Pizzo, F., et al., *Identification of structural alerts for liver and kidney toxicity using repeated dose toxicity data.* Chemistry Central Journal, 2015. **9**(1): p. 1-11.
100. Abernethy, D., J. Woodcock, and L. Lesko, *Pharmacological mechanism-based drug safety assessment and prediction.* Clinical Pharmacology & Therapeutics, 2011. **89**(6): p. 793-797.

101. Milan, C., et al., *Comparison and possible use of in silico tools for carcinogenicity within REACH legislation.* Journal of Environmental Science and Health, Part C, 2011. **29**(4): p. 300-323.
102. Valerio Jr, L.G., *In silico toxicology for the pharmaceutical sciences.* Toxicology and applied pharmacology, 2009. **241**(3): p. 356-370.
103. Ekins, S., *Computational toxicology: risk assessment for pharmaceutical and environmental chemicals*. Vol. 1. 2007: John Wiley & Sons.
104. Raies, A.B. and V.B. Bajic, *In silico toxicology: computational methods for the prediction of chemical toxicity.* Wiley Interdisciplinary Reviews: Computational Molecular Science, 2016. **6**(2): p. 147-172.
105. Lester, C., et al., *Structure activity relationship (SAR) toxicological assessments: The role of expert judgment.* Regulatory Toxicology and Pharmacology, 2018. **92**: p. 390-406.
106. McKinney, J.D., et al., *The practice of structure activity relationships (SAR) in toxicology.* Toxicological Sciences, 2000. **56**(1): p. 8-17.
107. Bohacek, R.S., C. McMartin, and W.C. Guida, *The art and practice of structure-based drug design: a molecular modeling perspective.* Medicinal research reviews, 1996. **16**(1): p. 3-50.
108. Cruz-Monteagudo, M., F. Borges, and M.N.D. Cordeiro, *Desirability-based multiobjective optimization for global QSAR studies: application to the design of novel NSAIDs with improved analgesic, antiinflammatory, and ulcerogenic profiles.* Journal of computational chemistry, 2008. **29**(14): p. 2445-2459.
109. Nicolotti, O., et al., *Improving quantitative structure− activity relationships through multiobjective optimization.* Journal of chemical information and modeling, 2009. **49**(10): p. 2290-2302.
110. Nicolotti, O., et al., *Multiobjective optimization in quantitative structure− activity relationships: deriving accurate and interpretable QSARs.* Journal of Medicinal Chemistry, 2002. **45**(23): p. 5069-5080.
111. Escher, S.E., et al., *Towards grouping concepts based on new approach methodologies in chemical hazard assessment: the read-across approach of the EU-ToxRisk project.* Archives of Toxicology, 2019. **93**(12): p. 3643-3667.
112. Schultz, T., et al., *A strategy for structuring and reporting a read-across prediction of toxicity.* Regulatory Toxicology and Pharmacology, 2015. **72**(3): p. 586-601.
113. Patlewicz, G., et al., *Use of category approaches, read-across and (Q) SAR: general considerations.* Regulatory Toxicology and Pharmacology, 2013. **67**(1): p. 1-12.
114. Cronin, M., *An introduction to chemical grouping, categories and read-across to predict toxicity.* Chemical Toxicity Prediction, 2013: p. 1-29.
115. Patlewicz, G., et al., *Navigating through the minefield of read-across tools: A review of in silico tools for grouping.* Computational Toxicology, 2017. **3**: p. 1-18.
116. Agency, E.C., *Read-Across Assessment Framework (RAAF)*, E.C. Agency, Editor. 2017.
117. Committee, E.S., et al., *Guidance on the use of the weight of evidence approach in scientific assessments.* Efsa Journal, 2017. **15**(8): p. e04971.

118. Waigmann, E., et al., *Risk assessment of genetically modified organisms (GMOs).* EFSA Journal, 2012. **10**(10): p. s1008.
119. Madden, J., *Tools for grouping chemicals and forming categories.* Chemical toxicity prediction: category formation and read-across, 2013: p. 72-97.
120. Enoch, S. and D. Roberts, *Approaches for grouping chemicals into categories.* 2013: Royal Society of Chemistry.
121. Hewitt, M., et al., *Hepatotoxicity: a scheme for generating chemical categories for read-across, structural alerts and insights into mechanism (s) of action.* Critical reviews in toxicology, 2013. **43**(7): p. 537-558.
122. Stepan, A.F., et al., *Structural alert/reactive metabolite concept as applied in medicinal chemistry to mitigate the risk of idiosyncratic drug toxicity: a perspective based on the critical examination of trends in the top 200 drugs marketed in the United States.* Chemical research in toxicology, 2011. **24**(9): p. 1345-1410.
123. Przybylak, K. and T. Schultz, *Informing chemical categories through the development of Adverse Outcome Pathways.* Chemical toxicity prediction: category formation and read-across, 2013. **17**.
124. Jeliazkova, N., et al., *Linking LRI AMBIT chemoinformatic system with the IUCLID substance database to support read-across of substance endpoint data and category formation.* Toxicol Lett, 2016. **258**: p. S114-S115.
125. Patlewicz, G., et al., *Building scientific confidence in the development and evaluation of read-across.* Regulatory Toxicology and Pharmacology, 2015. **72**(1): p. 117-133.
126. Gadaleta, D., et al., *A new semi-automated workflow for chemical data retrieval and quality checking for modeling applications.* Journal of cheminformatics, 2018. **10**(1): p. 1-13.
127. Cereto-Massagué, A., et al., *Molecular fingerprint similarity search in virtual screening.* Methods, 2015. **71**: p. 58-63.
128. Bajusz, D., A. Rácz, and K. Héberger, *Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?* Journal of cheminformatics, 2015. **7**(1): p. 1-13.
129. Ferrari, T., et al., *Automatic knowledge extraction from chemical structures: the case of mutagenicity prediction.* SAR and QSAR in Environmental Research, 2013. **24**(5): p. 365-383.
130. Berthold, M.R., et al., *KNIME-the Konstanz information miner: version 2.0 and beyond.* AcM SIGKDD explorations Newsletter, 2009. **11**(1): p. 26-31.
131. Gadaleta, D., et al., *Automated integration of structural, biological and metabolic similarities to improve read-across.* ALTEX-Alternatives to animal experimentation, 2020. **37**(3): p. 469-481.
132. Alfonso, A.Y.C., et al., *Exploration of structural requirements for azole chemicals towards human aromatase CYP19A1 activity: Classification modeling, structure-activity relationships and read-across study.* Toxicology in Vitro, 2022: p. 105332.
133. Caballero-Alfonso, A.Y., et al., *Ensemble-based modeling of chemical compounds with antimalarial activity.* Current Topics in Medicinal Chemistry, 2019. **19**(11): p. 957-969.

134. Caballero, A., et al. *Assessment of a framework to identify analogues for read-across: case study.* in *TOXICOLOGY LETTERS*. 2019. ELSEVIER IRELAND LTD ELSEVIER HOUSE, BROOKVALE PLAZA, EAST PARK SHANNON, CO ….
135. *Tox21Assay*. 2018.
136. Franco, P., et al., *The use of 2D fingerprint methods to support the assessment of structural similarity in orphan drug legislation.* Journal of cheminformatics, 2014. **6**(1): p. 1-10.
137. Duan, J., et al., *Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods.* Journal of Molecular Graphics and Modelling, 2010. **29**(2): p. 157-170.
138. Hert, J., et al., *Comparison of fingerprint-based methods for virtual screening using multiple bioactive reference structures.* Journal of chemical information and computer sciences, 2004. **44**(3): p. 1177-1185.
139. Willett, P., J.M. Barnard, and G.M. Downs, *Chemical similarity searching.* Journal of chemical information and computer sciences, 1998. **38**(6): p. 983-996.
140. Willighagen, E.L., et al., *The Chemistry Development Kit (CDK) v2. 0: atom typing, depiction, molecular formulas, and substructure searching.* Journal of cheminformatics, 2017. **9**(1): p. 1-19.
141. Todeschini, R., V. Consonni, and R. Mannhold, *Methods and principles in medicinal chemistry.* Kubinyi H, Timmerman H (Series eds) Handbook of molecular descriptors. Wiley-VCH, Weinheim, 2000.
142. Ertl, P., B. Rohde, and P. Selzer, *Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties.* Journal of medicinal chemistry, 2000. **43**(20): p. 3714-3717.
143. Alagga, A.A. and V. Gupta, *Drug Absorption.* StatPearls [Internet], 2021.
144. Prasanna, S. and R. Doerksen, *Topological polar surface area: a useful descriptor in 2D-QSAR.* Current medicinal chemistry, 2009. **16**(1): p. 21-41.
145. Leo, A., C. Hansch, and D. Elkins, *Partition coefficients and their uses.* Chemical reviews, 1971. **71**(6): p. 525-616.
146. Vosooghi, M., et al., *Design, synthesis, docking study and cytotoxic activity evaluation of some novel letrozole analogs.* DARU Journal of Pharmaceutical Sciences, 2014. **22**(1): p. 1-7.
147. Shoombuatong, W., N. Schaduangrat, and C. Nantasenamat, *Towards understanding aromatase inhibitory activity via QSAR modeling.* EXCLI journal, 2018. **17**: p. 688.
148. Neidle, S., *Design Principles for Quadruplex-binding Small Molecules.* Therapeutic Applications of Quadruplex Nucleic Acids, 2012: p. 151-174.
149. Autino, J.C., G.P. Romanelli, and D.M. Ruiz, *Introducción a la química orgánica.* Series: Libros de Cátedra, 2013.
150. McMurry, J., *Química Orgánica. Cap. 22-23.* International Thomson, 2001.
151. Turner, J. and S. Agatonovic-Kustrin, *In silico prediction of oral bioavailability*, in *Comprehensive Medicinal Chemistry II ADME Tox Approaches*. 2007, Elsevier Ltd.

152. Becker, H., M. Naaman, and L. Gravano. *Learning similarity metrics for event identification in social media.* in *Proceedings of the third ACM international conference on Web search and data mining.* 2010.
153. Kokare, M., B. Chatterji, and P. Biswas. *Comparison of similarity metrics for texture image retrieval.* in *TENCON 2003. Conference on convergent technologies for Asia-Pacific region.* 2003. IEEE.
154. Reisen, F., et al., *Benchmarking of multivariate similarity measures for high-content screening fingerprints in phenotypic drug discovery.* Journal of biomolecular screening, 2013. **18**(10): p. 1284-1297.
155. Strehl, A., J. Ghosh, and R. Mooney. *Impact of similarity measures on web-page clustering.* in *Workshop on artificial intelligence for web search (AAAI 2000).* 2000.
156. Bender, A. and R.C. Glen, *Molecular similarity: a key technique in molecular informatics.* Organic & biomolecular chemistry, 2004. **2**(22): p. 3204-3218.
157. Maggiora, G., et al., *Molecular similarity in medicinal chemistry: miniperspective.* Journal of medicinal chemistry, 2014. **57**(8): p. 3186-3204.
158. Gillet, V.J., P. Willett, and J. Bradshaw, *Similarity searching using reduced graphs.* Journal of chemical information and computer sciences, 2003. **43**(2): p. 338-345.
159. Salim, N., J. Holliday, and P. Willett, *Combination of fingerprint-based similarity coefficients using data fusion.* Journal of chemical information and computer sciences, 2003. **43**(2): p. 435-442.
160. Gadaleta, D., et al., *SAR and QSAR modeling of a large collection of LD 50 rat acute oral toxicity data.* Journal of cheminformatics, 2019. **11**(1): p. 1-16.
161. Gini, G., et al., *A new QSAR model for acute fish toxicity based on mined structural alerts.* J Toxicol Risk Assess, 2019. **5**(1): p. 016.
162. Yang, H., et al., *In silico prediction of chemical toxicity for drug design using machine learning methods and structural alerts.* Frontiers in chemistry, 2018. **6**: p. 30.
163. Lombardo, A., et al., *A new in silico classification model for ready biodegradability, based on molecular fragments.* Chemosphere, 2014. **108**: p. 10-16.
164. Yang, H., et al., *Evaluation of different methods for identification of structural alerts using chemical ames mutagenicity data set as a benchmark.* Chemical research in toxicology, 2017. **30**(6): p. 1355-1364.
165. Selvestrel, G., et al., *Skin sensitization quantitative QSAR models based on mechanistic structural alerts.* Toxicology, 2022: p. 153111.
166. Vian, M., et al., *In silico model for mutagenicity (Ames test), taking into account metabolism.* Mutagenesis, 2019. **34**(1): p. 41-48.
167. Team, R.C., *R: A language and environment for statistical computing.* 2013.
168. Blakemore, J. and F. Naftolin, *Aromatase: contributions to physiology and disease in women and men.* Physiology, 2016.
169. Caciolla, J., et al., *Targeting orthosteric and allosteric pockets of aromatase via dual-mode novel azole inhibitors.* ACS medicinal chemistry letters, 2020. **11**(5): p. 732-739.

170. Sari, S., et al., *Antibacterial azole derivatives: Antibacterial activity, cytotoxicity, and in silico mechanistic studies.* Drug Development Research, 2020. **81**(8): p. 1026-1036.
171. Peng, X.-M., G.-X. Cai, and C.-H. Zhou, *Recent developments in azole compounds as antibacterial and antifungal agents.* Current topics in medicinal chemistry, 2013. **13**(16): p. 1963-2010.
172. Bajusz, D., A. Rácz, and K. Héberger, *Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?* Journal of Cheminformatics, 2015. **7**(1): p. 20.
173. Thiel and Bernd Wiswedel, M.R.B.a.N.C.a.F.D.a.T.R.G.a.T.K.o.t.a.T.M.a.P.O.a.C.S.a.K., *KNIME*. 2007, Springer.
174. Chattaraj, P.K., U. Sarkar, and D.R. Roy, *Electrophilicity index.* Chemical Reviews, 2006. **106**(6): p. 2065-2091.
175. Feher, M. and J.M. Schmidt, *Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry.* Journal of chemical information and computer sciences, 2003. **43**(1): p. 218-227.
176. Guégan, F., et al., *Dual descriptor and molecular electrostatic potential: complementary tools for the study of the coordination chemistry of ambiphilic ligands.* Physical Chemistry Chemical Physics, 2014. **16**(29): p. 15558-15569.
177. Kassimi, N.E.-B., R.J. Doerksen, and A.J. Thakkar, *Polarizabilities of aromatic five-membered rings: azoles.* The Journal of Physical Chemistry, 1995. **99**(34): p. 12790-12796.
178. Lata, S. and Vikas, *Externally predictive quantum-mechanical models for the adsorption of aromatic organic compounds by graphene-oxide nanomaterials.* SAR and QSAR in Environmental Research, 2019. **30**(12): p. 847-863.
179. Levine, I., *Quantum chemistry vol. 6: Pearson Prentice Hall Upper Saddle River*. 2009, NJ.
180. Roy, K., *Quantitative structure-activity relationships in drug design, predictive toxicology, and risk assessment*. 2015: IGI Global.
181. Lučić, B., et al., *Estimation of random accuracy and its use in validation of predictive quality of classification models within predictive challenges.* Croatica Chemica Acta, 2019. **92**(3): p. 379-391.
182. Batista, J., D. Vikić-Topić, and B. Lučić, *The Difference Between the Accuracy of Real and the Corresponding Random Model is a Useful Parameter for Validation of Two-State Classification Model Quality.* Croatica Chemica Acta, 2016. **89**(4): p. 527-534.
183. Sahin, Z., et al., *Studies on non-steroidal inhibitors of aromatase enzyme; 4-(aryl/heteroaryl)-2-(pyrimidin-2-yl) thiazole derivatives.* Bioorganic & medicinal chemistry, 2018. **26**(8): p. 1986-1995.
184. Ertas, M., et al., *Pyridine-substituted thiazolylphenol derivatives: Synthesis, modeling studies, aromatase inhibition, and antiproliferative activity evaluation.* Archiv der Pharmazie, 2018. **351**(3-4): p. 1700272.

185. Seth, A., et al., *Thiazolidinediones (TZDs) affect osteoblast viability and biomarkers independently of the TZD effects on aromatase.* Hormone and Metabolic Research, 2013. **45**(01): p. 1-8.
186. Vosooghi, M., et al., *Design, synthesis, docking study and cytotoxic activity evaluation of some novel letrozole analogs.* DARU Journal of Pharmaceutical Sciences, 2014. **22**(1): p. 83.
187. Song, Z., et al., *Synthesis and aromatase inhibitory evaluation of 4-N-nitrophenyl substituted amino-4H-1, 2, 4-triazole derivatives.* Bioorganic & medicinal chemistry, 2016. **24**(19): p. 4723-4730.
188. Wang, R., et al., *Design, synthesis and aromatase inhibitory activities of novel indole-imidazole derivatives.* Bioorganic & medicinal chemistry letters, 2013. **23**(6): p. 1760-1762.
189. Di Matteo, M., et al., *Synthesis and biological characterization of 3-(imidazol-1-ylmethyl) piperidine sulfonamides as aromatase inhibitors.* Bioorganic & medicinal chemistry letters, 2016. **26**(13): p. 3192-3194.
190. Jeong, S., et al., *Inhibition of drug metabolizing cytochrome P450s by the aromatase inhibitor drug letrozole and its major oxidative metabolite 4, 4′-methanol-bisbenzonitrile in vitro.* Cancer chemotherapy and pharmacology, 2009. **64**(5): p. 867-875.
191. Ghodsi, R. and B. Hemmateenejad, *QSAR study of diarylalkylimidazole and diarylalkyltriazole aromatase inhibitors.* Medicinal Chemistry Research, 2016. **25**(5): p. 834-842.
192. Gianti, E. and R.J. Zauhar, *Structure–activity relationships and drug design*, in *Remington*. 2021, Elsevier. p. 129-153.
193. Schetz, J., *Structure-Activity Relationships: Theory, Uses and Limitations.* 2015.
194. Ferlin, M.G., et al., *Design, synthesis, and structure–activity relationships of azolylmethylpyrroloquinolines as nonsteroidal aromatase inhibitors.* Journal of medicinal chemistry, 2013. **56**(19): p. 7536-7551.
195. Papillon, J.P., et al., *Structure–Activity Relationships, Pharmacokinetics, and in Vivo Activity of CYP11B2 and CYP11B1 Inhibitors.* Journal of medicinal chemistry, 2015. **58**(11): p. 4749-4770.
196. Demchuk, E., et al., *SAR/QSAR methods in public health practice.* Toxicology and applied pharmacology, 2011. **254**(2): p. 192-197.
197. Mohapatra, A., *Software tools for toxicology and risk assessment*, in *Information Resources in Toxicology*. 2020, Elsevier. p. 791-812.
198. Wassermann, A.M., et al., *SAR matrices: automated extraction of information-rich SAR tables from large compound data sets.* Journal of chemical information and modeling, 2012. **52**(7): p. 1769-1776.
199. Information, N.C.f.B., *PubChem Substructure Fingerprint*, in *PubChem Data Specification Directory*. 2009.
200. Fernández-de Gortari, E., et al., *Database fingerprint (DFP): an approach to represent molecular databases.* Journal of cheminformatics, 2017. **9**(1): p. 1-9.
201. Rostkowski, M., O. Spjuth, and P. Rydberg, *WhichCyp: prediction of cytochromes P450 inhibition.* Bioinformatics, 2013. **29**(16): p. 2051-2052.

202. Webster, F., et al., *Predicting estrogen receptor activation by a group of substituted phenols: An integrated approach to testing and assessment case study.* Regulatory Toxicology and Pharmacology, 2019. **106**: p. 278-291.
203. Chicco, D. and G. Jurman, *The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation.* BMC genomics, 2020. **21**(1): p. 1-13.
204. Bubalo, M.C., et al., *Toxicity mechanisms of ionic liquids.* Arhiv za higijenu rada i toksikologiju, 2017. **68**(3): p. 171-179.
205. Cruz-Monteagudo, M., et al., *Chemoinformatics profiling of ionic liquids—automatic and chemically interpretable cytotoxicity profiling, virtual screening, and cytotoxicophore identification.* toxicological sciences, 2013. **136**(2): p. 548-565.
206. Ranke, J., et al., *Biological effects of imidazolium ionic liquids with varying chain lengths in acute Vibrio fischeri and WST-1 cell viability assays.* Ecotoxicology and environmental safety, 2004. **58**(3): p. 396-404.
207. Ranke, J., et al., *Lipophilicity parameters for ionic liquid cations and their correlation to in vitro cytotoxicity.* Ecotoxicology and environmental safety, 2007. **67**(3): p. 430-438.
208. Dong, X., et al., *Inhibitory effects of ionic liquids on the lactic dehydrogenase activity.* International journal of biological macromolecules, 2016. **86**: p. 155-161.
209. Na, L., et al., *Effect of imidazolium ionic liquids on the hydrolytic activity of lipase.* Chinese Journal of Catalysis, 2013. **34**(4): p. 769-780.
210. McConkey, B.J., V. Sobolev, and M. Edelman, *The performance of current methods in ligand–protein docking.* Current Science, 2002: p. 845-856.

Bibliography

Publications Related to the Thesis

Journal Articles

Alfonso, A. Y. C., Lagares, L. M., Novic, M., Benfenati, E., & Kumar, A. (2022). Exploration of structural requirements for azole chemicals towards human aromatase CYP19A1 activity: Classification modeling, structure-activity relationships and read-across study. Toxicology *in vitro*, 105332.

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.

Caballero-Alfonso, A. Y., Cruz-Monteagudo, M., Tejera, E., Benfenati, E., Borges, F., Cordeiro, M. N. D. S., ... & Perez-Castillo, Y. (2019). Ensemble-based modeling of chemical compounds with antimalarial activity. Current Topics in Medicinal Chemistry, 19(11), 957-969.

Conference Paper

Caballero, A. Y., Toma, C., Gadaleta, D., Perez, Y., & Benfenati, E. (2019, October). Assessment of a framework to identify analogues for read-across: case study. In TOXICOLOGY LETTERS (Vol. 314, pp. S272-S273). ELSEVIER HOUSE, BROOKVALE PLAZA, EAST PARK SHANNON, CO, CLARE, 00000, IRELAND: ELSEVIER IRELAND LTD.

Other Publications

Selvestrel, G., Baderna, D., Toma, C., Alfonso, A. Y. C., Gamba, A., & Benfenati, E. (2022). Skin sensitization quantitative QSAR models based on mechanistic structural alerts. Toxicology, 153111

Mora Lagares, L., Minovski, N., Caballero Alfonso, A. Y., Benfenati, E., Wellens, S., Culot, M., ... & Novič, M. (2020). Homology modeling of the human p-glycoprotein (Abcb1) and insights into ligand binding through molecular docking studies. International journal of molecular sciences, 21(11), 4058.

.

Biography

The author of this thesis Ana Yisel Caballero Alfonso was born on the 18th of November 1993 in Santa Clara, a small city in the center of Cuba. Since her childhood, she has been active, enthusiastic, and curious about the world and its mysteries. She has attended a high school specialized in sciences, which helped her to grow academically, and developing a passion for the natural sciences, especially a high vocation for Chemistry and Biology. Prior to initiating University studies, she was involved in several Chemistry concourses, and received many regional awards.

Ana Caballero started her bachelor's degree in Pharmaceutical Sciences in the Central University Marta Abreu de las Villas, where she associated with the Bioinformatics team as a researcher student by the start of the second year of the Degree course. Besides, she participated in local and national student forums and was honored with many awards. She obtained the bachelor's degree in Pharmaceutical Sciences in 2016 with honors and the award for the most comprehensive student.

To follow her dreams, upon the completion of her degree, she specialized in Molecular Biology, at the Central University Marta Abreu de las Villas. By the year 2017, she obtained her title in Applied Molecular Technics associated to the Biology Faculty and the Centro de Bioactivos Quimicos. The same year she received an international prize for the excellence in research for her participation in the IDIFARMA congress.

Afterwards, she has been awarded with the prestigious Marie Curie Fellowship funded by the EU's Marie Sklodowska-Curie Action - Innovative Training Network. This splendid opportunity allowed her to initiate her doctoral degree in the field of Computational Toxicology at the Jožef Stefan International Postgraduate School, Slovenia, and join the Mario Negri Institute for Pharmacological Research, in Milan, Italy. She worked at the Department of Environment and Health as a researcher. During this period, she participated in several secondments at institutes of excellence such as the Liverpool Jhon Moores University and L'Oreal industry. Ana Caballero was announced as a quality researcher by famous experts in the field of alternative methods and toxicology. Later, she joined the European Chemicals Agency (ECHA) for a period of six months for training in data sciences in the hazard assessment unit.

Currently, she is working at ProtoQSAR S.L., a company focused on the development and application of computational methods for the replacement of animal testing, as a researcher and business developer. We live in a world where reality may be overshadowed by the fallacy, notion, and misconception. Yet hope remains, as scientific discoveries can change all that is anticipated. Each scientific field has its own role to play, though computer science in toxicology has also a great potential for understanding and fixing fundamental problems in our society. And that is her dream, a new approach, and this new approach can only be achieved by taking risks to try something original and new, investing in open minded, ambitious visionaries who can invent them.