CHAPTER – 2

Review of Literature
2.1 Therapeutic Importance of the Synthesized Heterocycles

2.1.1 Pyrazoline derivatives

2.1.1.1 Anticancer activity

Havrylyuk et al.\textsuperscript{146-148} examined several thiazolone and isatin based pyrazolines as anticancer agents against several human (leukemia, melanoma, lung, colon, CNS, ovarian, renal, prostate and breast) cancer cell lines. Amongst the tested compounds, 1 was found to be most active with mean GI\textsubscript{50} and TGI values of 0.071 \(\mu\text{M}\) and 0.76 \(\mu\text{M}\), respectively against all cancer cell lines. It demonstrated the highest antiproliferative influence on the non-small-cell lung cancer cell line HOP-92 (GI\textsubscript{50} < 0.01 \(\mu\text{M}\)), colon cancer line HCT-116 (GI\textsubscript{50} = 0.018 \(\mu\text{M}\)), CNS cancer cell line SNB-75 (GI\textsubscript{50} = 0.0159 \(\mu\text{M}\)), ovarian cancer cell line NCI/ADR-RES (GI\textsubscript{50} = 0.0169 \(\mu\text{M}\)), and renal cancer cell line RXF 393 (GI\textsubscript{50} = 0.0197 \(\mu\text{M}\)).

Several 17-pyrazolinyl derivatives of pregnenolones were developed and evaluated for anticancer activity against a panel of seven human cancer cell lines.\textsuperscript{149-150} Amongst the compounds which exhibited promising anticancer activity, analogues of 2 exhibited significant cytotoxicity (IC\textsubscript{50} = 0.24-2.37 \(\mu\text{M}\)) against HT-29 (colon), HCT-15 (colon) and 502713 (lung) cancer cell lines, while 3 was found to be significantly potent (IC\textsubscript{50} = 0.91 \(\mu\text{M}\)) against MDA-MB-231 (human breast) cancer cell line.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{images.png}
\caption{(1) (2) (3) (4)}
\end{figure}
Demirayak et al. prepared 2,3,4-tri-substituted pyrazolone-5-one derivatives and evaluated them for their \textit{in vitro} cytotoxicity against 66 human tumor cell lines derived from nine neoplastic diseases.\textsuperscript{151} The min-graph midpoint values shown that 4 provides an acceptable activity (log$GI_{50} = -4.39$) as compared to melphanin (-5.09) and cisplatin (-6.20).

2.1.1.2 Steroidal activity

For decades, androgen receptor (AR) has been a target for drug development focused upon the treatment of pathological conditions arising from abnormal androgen levels or altered target tissue responsiveness, the improvement of physical performance, and the regulation of male fertility. A new class of molecules targeting androgen receptors called selective androgen receptor modulators (SARMs) has been developed which has antagonist or weak agonist activity in the prostate (androgenic organ) while presenting strong agonist activity in the muscle and bone (anabolic organ). This suggests plausible application of these molecules for the treatment of androgen mediated diseases such as prostate cancer, hirsutism, acne, muscle-wasting conditions, hypogonadism, or age-related frailty while preventing potential risks for nascent or undetected prostate cancer. A novel series of pyrazolines (5) have been synthesized and evaluated as tissue-selective androgen receptor modulators (SARMs) \textit{in vitro}.\textsuperscript{152} SAR data suggested that the presence of strong electron-withdrawing groups at R\textsubscript{1} and R\textsubscript{2} positions and small groups at R\textsubscript{3} and R\textsubscript{6} positions are optimal for AR agonist activity.
Several androstano[14,16-c] pyrazolines and their oxidized derivatives were evaluated by Amr et al., for their anti androgenic properties. Compound 6 exhibited significant potency in vitro and in vivo when compared with the standard drug cyproterone.

Antiprogesterone agents are type of progesterone inhibitors which antagonizes or suppress the action of progesterone in body and is a classical strategy in treating brain meningiomas, breast cancer, uterine fibroids and endometriosis. Jones and his co-workers reported the discovery of novel pyrazolines as progesterone receptor ligands. Compound 7 exhibits high affinity for the progesterone receptor and functional profile comparable to the steroidal progesterone antagonist mifepristone.

2.1.1.3 Monoamine oxidase (MAO) inhibition

It has been well established that the two MAO isomers (A and B) plays a key role in the metabolism of neurotransmitters. As a result, MAO inhibitors (MAOI) are useful in the treatment of several psychiatric and neurological diseases. Reversible, selective MAO-A inhibitors are used as antidepressant and antianxiety drugs, and selective MAO-B inhibitors are co adjuvant in the treatment of Parkinson’s disease and Alzheimer’s disease.

A series of 1-acetyl pyrazolines were investigated for their ability to selectively inhibit the activity of MAO-A. Among all compounds, (-) stereo isomer of 8 showed the highest antiMAO-A activity (K_i= 2.0 nM) with maximum A selectivity (SI= 165000). In a similar series of pyrazolines, compound 9 showed selective MAO-A inhibition activity as compared to the standard drug Meclobemide.
In another series of selective and reversible pyrazoline based MAO-B inhibitors, homologues of 10 showed significant activity for human (Ki ~ 0.31-1.7 nM) MAO-B; (Ki ~ 32 - 300 nM) MAO-A.

2.1.1.4 Nitric oxide synthase (NOS) inhibition

Nitric oxide (NO) is an important bioregulator and ubiquitous messenger that mediates several normal and physiological processes. The biosynthesis of NO is mediated by nitric oxide synthase (NOS), which catalyses the L-arginine oxidation. Classically, three isoforms of NOS have been identified, two of them are constitutively expressed and are known as neural NOS (nNOS) and endothelial cell NOS (eNOS), while the third one is induced by several factors, and hence is called inducible NOS (iNOS). The NO action can be physiological or pathological. An excessive NO production by nNOS is involved in different neurological disorders such as Alzheimer’s disease, the amyotrophic lateral sclerosis and Huntington’s disease. The induction of iNOS produces a continuous and elevated NO production, and is involved in a variety of diseases, including septic shock, inflammatory arthritis, and inflammatory bowel disease. Development of NOS inhibitors constitutes a current strategy in the research of new compounds showing interesting properties.

Carrion et al., evaluated a new class of pyrazolines for their inhibition activity against both the neural and inducible isoforms of NOS. Compounds 11 exhibited the 50% inhibition of iNOS at 1μM.
In an attempt to search new pyrazoline derivatives with neuroprotective activity attributed to nNOS inhibition, Camacho et al. discovered new compounds 12 and 13 showing 70% and 62% inhibition of nNOS, respectively.\textsuperscript{169}

2.1.1.5 Antiamoebic activity

Amoebiasis is a protozoan infection caused by intestinal parasite Entamoeba histolytica. The prevalence of amoebic colitis and liver abscess is greater in developing countries. Amoebic abscesses of the brain are a dreadful complication of E. histolytica infection.\textsuperscript{170} The nitroimidazoles such as metronidazole are the cornerstones of amoebic liver abscess treatment. But due to mutagenic property of metronidazole, associated with serious side effects, and resistance of some E. histolytica strains to this drug,\textsuperscript{171-173} search for new lead molecules, which can be effectively used against amoebiasis, are the need of time.

Several pyrazolines of type 14 and 15 have been synthesized and evaluated for their antiamoebic activity, amongst which most of the derivatives showed higher potency (IC\textsubscript{50} = 0.42-0.67 μM) than the standard drug metronidazole.\textsuperscript{174-175} Recently, Hayat et al. reported synthesis and evaluation of novel pyrazolines bearing quinoline tail (16).\textsuperscript{176} Several of these pyrazoline derivatives exhibited high potency (IC\textsubscript{50} ~ 0.05-0.31μM).
2.1.1.6 Anti-inflammatory activity

Conventional non-steroidal anti-inflammatory drugs (NSAIDs) that non-selectively inhibit both major cyclooxygenase isoforms (COX-1 and COX-2), are widely used for the treatment of inflammation, mainly arthritic pain. COX-1 isoform is mainly responsible for the synthesis of cytoprotective prostaglandins in gastrointestinal (GI) tract, whereas COX-2 is inducible and plays a major role in biosynthesis of prostaglandin in inflammatory cells. It is believed that the inhibition of COX-1 causes unfavorable GI side effects. Hence, a high level of selective COX-2 inhibition is a therapeutic strategy to alleviate pain and inflammation. Therefore selective COX-2 inhibitors (coxibs) with better safety profile have been marketed as a new generation of NSAIDs. However, the withdrawal of rofecoxib, valdecoxib, due to serious cardiovascular toxicities raised alarming concerns to discover newer anti-inflammatory agents.177-178

A numbers of 1,3,5-trisubstituted-2-pyrazolines were synthesized and examined for their anti-inflammatory activity.179-180 Among the compounds examined, 17 and 18 showed potent anti-inflammatory and analgesic activity with low ulcerogenic potencial.

![Chemical structures](image)

 Recently, Bashir et al., reported a series of substituted pyrazolines and screened them for their anti-inflammatory and analgesic profile.181 Compound 19 was found to have potent anti-inflammatory activity than celecoxib with low ulcerogenic toxicity. The in vitro enzymatic assay of this compound revealed no inhibition of COX-I and II.
Hence, it was assumed that these derivatives might inhibit enzymatic activity of 5-lipooxygenase (LOX).

2.1.1.7 Antiviral activity

The West Nile Virus (WNV) is a member of the *Flavivirus* genus, which belongs to the *Flavividae* family. The *Flavivirus* genus includes a number of additional significant human pathogens such as yellow fever virus (YF), dengue virus (DEN), Japanese encephalitis virus (JE), Murray Valley encephalitis virus (MVE), St. Louis encephalitis virus (SLE), and tick-borne encephalitis virus (TBE). Viruses belonging to the *Flavivirus* genus are typically transmitted to vertebrates by mosquitoes or ticks and are responsible for severe morbidity and mortality in both humans and animals.\(^{182-184}\) Therefore, discovering compounds that can be used for prevention and treatment of WNV infections is becoming a general public health priority as there are no viable drugs or vaccines available to treat or prevent WNV infection in humans at present.

Goodell *et al.* examined several pyrazolines for their anti WNV activity using HTS developed in their laboratory using three separate systems; a cell line containing a persistently replicating subgenomic replicon, a full length reporting virus, and packaged virus-like particles (VLPs) containing replicon RNA.\(^{185}\) On the basis of activity and therapeutic indexes, these compounds 20-21 were identified as viable leads. The further evaluation of mechanism of action accounts that the pyrazoline inhibits RNA synthesis, pointing viral RNA polymerase, RNA helicase or other viral replication enzymes as potential targets. Later on Baragoiti *et al.* found that, besides
the inhibition of WNV, these compounds also inhibited other flaviviruses, an alphavirus, a corona virus and rhabdovirus. However, they were found to be inactive in suppressing orthomyxovirus or retrovirus.\textsuperscript{186}

2.1.1.8 Dipeptidyl peptidase inhibition

Dipeptidyl peptidase IV (DPP-IV) modulates the biological activity of several peptide hormone, chemokines and neuropeptides by specific mode of action. The inhibition of DP-IV increases the level of circulating GLP-1 and thus increases the insulin secretion which could ameliorate hyperglycemia in type-2 diabetes treatment.

Recently compound 22 has been reported as potential DPP-IV inhibitor with IC\textsubscript{50} value of 0.22, 0.13 and 0.08\textmu M in rat plasma, caco-2 and porcine kidney, respectively.\textsuperscript{187} Ahn et al. reported a series of cyanopyrazolines evaluated for their ability to inhibit DPP-IV. Amongst the series compounds, compound 23 was found to be most active with IC\textsubscript{50} 0.8, 0.41 and 0.96\textmu M in rat plasma, caco-2 and porcine kidney, respectively.\textsuperscript{188} Compound 23 also revealed good in vivo efficacy (ED\textsubscript{50}: 4.1mg/kg in vivo DPP-IV inhibition). The chemical modification at 5\textsuperscript{th} position resulted in discovery of more potent pyrazolines, amongst which compound 24 exhibited the most potent inhibitory activity against DPP-IV and also overcame the CYP3A4 enzyme inhibition.\textsuperscript{189}
2.1.1.9 Cannabinoid (CB) receptor antagonist activity

The CB₁ cannabinoid receptor is expressed at high levels in several brain areas including hippocampus, cortex, cerebellum and basal ganglia as well as in some peripheral tissues including urinary bladder, testis and ileum. The CB₂ receptor is principally found in immune system. CB₁ receptor antagonist may have potential in treatment of number of diseases such as neuroinflammatory disorders, cognitive disorders, septic shock, obesity, psychosis and gastrointestinal disorders.

Apart from the well known CB₁ receptor antagonist rimonabaunt (25), Lange et al. reported 3,4 diarylpyrazolines as potent CB₁ antagonist.190 The highest CB₁ receptor affinity (7.8 nM) was found for the compound 26 which is in same order of magnitude as that for rimonabaunt (11.5nM). The compound also exhibited highest CB₁/CB₂ receptor selectivity (~1000), which is ~7 fold higher than that for rimonabaunt. Number of analogues of diaryl dihydro pyrazole-3-carboxamide have been synthesized and evaluated for appetite suppression and body weight reduction in animal models.191 The bisulphate salts of racemic mixture and pure (-) enantiomer of compound 27 showed significant body weight reduction in vivo, which was attributed to CB₁ antagonist activity.

Later on the detailed findings suggested that the bio-isosteric replacement of pyrazoline nucleus of compound 27 by imidazole and oxazole resulted in complete
lose of required conformation of the molecules, necessary for the binding interactions, thus pyrazoline nucleus is necessary for CB1 binding.\textsuperscript{192}

2.1.1.10 Anti-insecticidal activity

The first generation of pyrazoline-type insecticides (28) was discovered in the 1970s as a result of research on the insecticidal activities of compounds related to the 1-benzoyl-3-phenylureas, such as diflubenzuron.\textsuperscript{193} Synthesis around the benzoylphenylurea structure identified a related group of compounds, the 1-phenyl carbamoyl-2-pyrazolines, that also displayed high levels of insecticidal activity but exhibited completely different symptoms of intoxication.\textsuperscript{194} The 1-phenylcarbamoyl-2-pyrazolines cause uncoordinated movements, cessation of feeding, convulsions, and death without the insect ever reaching the molting stage.\textsuperscript{194} Exploration of the 3-phenyl and 3,4- and 4,5-diphenyl variants of this series of compounds led to the discovery of structures far more active than 28, such as compound 29.\textsuperscript{195-197} Further development of dihydropyrazole insecticides resulted in the discovery of a second generation of compounds having both structural and functional similarity. These compounds retained a pyrazoline ring as the center of the molecule, and alterations were limited to the surrounding structural components.

Pyrazoline 30 is an example of a dihydropyrazole compound with high insecticidal activity. The insecticidal activity of this compound, like other dihydropyrazoles, is stereoselective. In American cockroaches and house flies, the S isomer is 10 and 100 times more insecticidally active than its R counterpart, respectively.\textsuperscript{198}
2.1.1.11 Antifungal activity

Zampiri et al. synthesized 4-imidazolyl-3,5-diaryl-2-pyrazolines and tested for their antifungal activities against clinical strains of *Candida albicans*.\(^{199}\) Compound 31 showed remarkable antifungal activity reaching the MIC values of 0.06g/mL which was superior to the reference drugs miconazole and amphotericin B. Ozdemir et al. also evaluated antifungal activity of series of new pyrazolines against various *Candida* species.\(^{200}\) Most of the compound exhibited appreciable activity against these species amongst which some compounds of type 32 were found to be even more potent than standard drug ketoconazole. Recently, examination of a series of pyrazolines and pyrazoles for their antifungal activity resulted in an identification of compounds 33 and 34 which were found to be equipotent to clotrimazole.\(^{201}\)

\[
\begin{align*}
\text{(31)} & \quad \text{Br} \\
\text{(32)} & \quad \text{X} = \text{NH, O, S} \\
\text{(33)} & \quad R_1 = \text{H, Br} \\
\text{(34)} & \quad R_2 = \text{4-Br-C}_6\text{H}_4
\end{align*}
\]

2.1.1.12 Anti-hepatotoxic activity

Khalillullah et al. studied the anti-hepatotoxic effect of pyrazoline derivatives on serum enzymatic activity in carbon tetrachloride induced liver damage in rats.\(^{202}\)

\[
\begin{align*}
\text{(35)} & \quad \text{R} = \text{H, 2,4-di-OH, 4-Cl, 4-OCH}_3
\end{align*}
\]
The administration of compound 35 significantly reduced the elevated enzyme level to values in the range comparable to that achieved with standard drug silymarin.

2.1.1.13 Acyl-co enzyme A: cholesterol transferase (ACAT) inhibition

Acyl-co enzyme A: cholesterol transferase (ACAT) is an integral membrane protein localized in the endoplasmic reticulum. ACAT catalyses the formation of cholesteryl ester from cholesterol fatty acyl co-A. The cholesteryl esters are stored as cytoplasmic lipid droplet inside the cell. The inhibition of ACAT activity has been associated with decreased plasma cholesterol levels by suppressing cholesterol adsorption and by diminishing the assembly and secretion of apolipoprotein β containing lipoproteins. Jeong et al. discovered novel class of hACAT-1 and hACAT-2 enzyme inhibitors of type 36.203 These compounds showed appreciable inhibition of hACAT-1 and -2 with values 0.14 μM and 0.17 μM respectively.

2.1.1.14 5-Hydroxytryptamine 6 receptor (5-HT6R) antagonist activity

In efforts to discover novel, neutral 5-HT6R antagonists to improve off-target selectivity towards 5-HT6 receptor, Loevenzijn et al. identified N-(sulfonyl) pyrazoline-1-carboxamide scaffold as promising neutral core from HTS results.204 Unique structural features of these compounds include a pseudoaromatic system and an internal hydrogen bond, freezing the bioactive conformation. The continuous development from hit to lead resulted in discovery of compound 37 which was found to be potent, selective, hERG free and CNS available neutral 5-HT6R.
2.1.1.15 Antitubercular activity

Shaharyar et al. reported the synthesis and in vitro antitubercular evolution of novel series of 4-(substituted-2-pyrazolino)phenoxy acetic acid derivatives (38-40) and N-nicotinoyl-3,5-di(substitutedphenyl)-2-pyrazolines (41), against both the viral (H37Rv) and multi drug resistant (MDR) strains of MTB. Most of the compounds were found to be potent with MIC < 1 μM; H37Rv and MDR-TB. The active compounds were also subjected to cytotoxicity assay using Vero cell line where these compounds were found to be non-toxic at 62.5 μg/mL, showing good selectivity index (>10).

Almeida da Silva et al. studied the synthesis and antitubercular evolution of 3-substituted 5-hydroxy-5-trihalomethyl-1-isonicotinoyl-2-pyrazoline derivatives (42,
43) against viral strain H₃₇Rv and various isoniazid resistant clinical isolates of MTB and non-tubercular mycobacteria.²⁰⁷

\[
\begin{align*}
\text{Zampieri et al. reported 1-(3,5-diaryl-2-pyrazoline)-1H-imidazole derivatives (44) as potential antitubercular agents.¹⁹⁹ The series of the compounds exhibited good activity against the viral strain (H₃₇Rv) of MTB.}
\end{align*}
\]

\[
\begin{align*}
\text{Eswaran et al. reported the synthesis and antitubercular activity of novel series of 4-pyrazolino quinoline derivatives.²⁰⁸ Compound 45 was found to possess higher activity amongst the other homologues of the series.}
\end{align*}
\]

\[
\begin{align*}
\text{Manna et al. synthesized novel 1,3,5-trisubstituted-2-pyrazoline derivatives. Though none of the compounds was found to be more potent than isoniazid against H₃₇Rv}
\end{align*}
\]
strain, compound 46 exhibited appreciable potency MIC~3.2 μg/mL against MDR-TB strain.²⁰⁹

Ahsan et al. reported novel 3a,4-dihydro-3H-indene[1,2-c]pyrazol-2-carboxamide analogues as potential antitubercular agents.²¹⁰ Amongst the series compounds, compound 47 (MIC$_{H37Rv}$=0.78μM, MIC$_{MDR-TB}$=0.78 μM), 48 (MIC$_{H37Rv}$=0.78μM, MIC$_{MDR-TB}$=3.12 μM) and 49 (MIC$_{H37Rv}$=0.83μM, MIC$_{MDR-TB}$=3.32 μM) showed good potency against both viral (H$_{37}$Rv) and MDR-TB strains.

Recently, Khunt et al. reported N-phenyl-3-(4-fluorophenyl)-4(substituted-2-pyrazolinio) pyrazole derivatives as potential antitubercular agents.²¹¹ The synthesized compounds were evaluated in vitro for their antitubercular activity against H$_{37}$Rv strain of MTB using BACTEC 460 radiometric system. Amongst the synthesized compounds, compound 50 (IC$_{50}$ = 0.47μM) and 51 (IC$_{50}$ = 2.65μM) showed appreciable results.
2.1.2 Benzoxazole derivatives

2.1.2.1 Anticancer activity

UK-1 (52), a bis(benzoxazole) natural product, displays a wide spectrum of potent anticancer activity against a wide range of human cancer cell lines. A series of truncated analogues of UK-1 have been prepared and evaluated for their anticancer activity. Within the series of analogues examined, only one analogue, compound 53, demonstrated cancer cell cytotoxicity.

In an effort to establish new candidates with improved antineoplastic, anti-HIV-1 and antimicrobial activities Rida et al. reported the synthesis and in vitro biological evaluation of various series of 2-substituted benzoxazoles. The most active compound 54 was found to possess GI50 MG-MID value 15.8 μM.

2.1.2.2 Antiviral activity

Based on molecular modeling simulation study including docking into the NS5B polymerase active site, Ismail et al. designed and synthesized inhibitors of Hepatitis C virus (HCV) based on benzoxazole scaffold. A number of the synthesized compounds showed significant inhibitory activity ranging from (52.2% inhibition up to 98% at < 50 μg/mL). Compound 55 demonstrated good HCV inhibitory activity with EC50 values of 41.6 and 24.5 μg/mL, respectively.

The highly conserved internal ribosome entry site (IRES) of HCV regulates translation of the viral RNA genome and is essential for the expression of HCV proteins in infected host cells. The structured subdomain IIa of the IRES element is the target site of inhibitors that selectively block viral translation through capture of
an extended conformation of an RNA internal loop. Zhou et al. identified a benzoxazole 56 as a ligand that bound selectively to IIa IRES target and confirmed as an inhibitor of in vitro viral translation by screening of a small pilot set of potential RNA binders.215

Synthesis and antiviral properties of new difluoromethylbenzoxazole (DFMB) pyrimidine thioether derivatives as non-nucleoside HIV-1 reverse transcriptase inhibitors have been reported by Boyer et al.216 By use of a combination of structural biology study and traditional medicinal chemistry, several members of this novel class were synthesized using a single electron transfer chain process (radical nucleophilic substitution, SRN1) and were found to be potent against wild-type HIV-1 reverse transcriptase, with low cytotoxicity but with moderate activity against drug-resistant strains. The most promising compound 57 showed a significant EC₅₀ value close to 6.4 nM against HIV-1 IIIB, a moderate EC₅₀ value close to 54 μM against an NNRTI resistant double mutant (K103N + Y181C), but an excellent selectivity index >15477 (CC₅₀ > 100 μM).

2.1.2.3 Anti-inflammatory activity

Paramashivappa synthesised a series of benzoxazoles from anacardic acid and investigated their ability to inhibit human cyclooxygenase-2 enzyme (COX-2). The active compounds were screened for cyclooxygenase-1 (COX-1) inhibition. Compound 58 was found to be more than 470-fold selective towards COX-2 compared to COX-1.

Inhibition of mPGES-1, the terminal enzyme in the arachidonic acid/COX pathway to regulate the production of pro-inflammatory prostaglandin PGE2, is considered an
attractive new therapeutic target for safe and effective anti-inflammatory drugs. The benzoxazole scaffold has been described in the patent literature to yield inhibitors of the MAPEG family enzymes including mPGES-1.\textsuperscript{217} The most potent compounds against mPGES-1 disclosed are 59 (IC$_{50}$ = 1.3 μM) and 60 (IC$_{50}$ = 1.5 μM) although no selectivity and efficacy data was reported.

![Chemical structures](image)

Arhancet et al. disclosed a novel class of benzoxazoles as inhibitors of mPGES-1. The initial HTS hit 61, showed modest activity in the enzyme assay (850 nM), but lacked potency in the cell-based assay. Structure–activity optimization of lead 61 with cyclohexyl carbinols resulted in compound 62, which showed excellent in vitro potency and selectivity against COX-2, and reasonable pharmacokinetic properties. Further SAR studies of the benzoxazole ring substituents led to a novel series of highly potent compounds with improved PK profile, including PF-4693627, which was effective in a carrageenan-stimulated guinea pig air pouch model of inflammation. Based on its excellent in vitro and in vivo pharmacological, pharmacokinetic and safety profile, compound PF-4693627 has advanced to clinical studies.\textsuperscript{218}
2.1.2.4 Cholesteryl ester transfer protein (CETP) inhibition

Cholesteryl ester transfer protein (CETP) mediates the exchange of cholesteryl ester (CE) from high density lipoprotein (HDL) with triglycerides primarily from very low-density lipoprotein (VLDL). Inhibition of CETP would therefore be expected to increase serum HDL-C levels.

A screening campaign, carried out on the BMS compound collection using a BODIPY fluorescence assay identified 2-arylbenzoxazole 63 as one of the most potent analogs in the HTS (SPA IC$_{50}$ = 0.28 μM) which also possessed a reasonable level of activity in human plasma (WPA IC$_{50}$ = 10 μM). An array synthesis approach was taken to rapidly optimize the in vitro CETP inhibition profile for this novel series of compounds. In accordance with the predictions derived from molecular modeling of 63, substitutions on either the A- or B- ring resulted in modulation of CETP inhibitory potency. In particular a compound 64 found to be the most potent (SPA IC$_{50}$ = 0.010 μM and WPA IC$_{50}$ = 0.91 μM). Smith and his coworkers reported a series of 2-arylbenzoxazole as inhibitors of CETP. The substitution at the
5- and 7-positions of the benzoxazole moiety was found to be beneficial for CETP inhibition. Compound 65 was found to be the most potent inhibitor in this series and inhibited CETP with an IC$_{50}$ of 28 nM.\textsuperscript{220} Another 2-aryl benzoxazole analogue 66 has been reported by Kallashi \textit{et al} as a potent CETP inhibitor with an IC$_{50}$ of 16 nM.\textsuperscript{221} Reversing the connectivity of the central aniline lead to a new class of 2-(4-carbonylphenyl)benzoxazoles. Structure–activity studies at the C-7 and terminal pyridine ring allowed for the optimization of potency and HDLc-raising efficacy in this new class of inhibitors. These efforts lead to the discovery of benzoxazole 67 with an IC$_{50}$ of 13 nM.\textsuperscript{222}

\subsection{2.1.2.5 Antitubercular activity}

Klimesova \textit{et al.}\textsuperscript{223} synthesized a set of 2-benzylsulfanyl derivatives of benzoxazole and evaluated for their \textit{in vitro} antimycobacterial activity against \textit{Mycobacterium tuberculosis}, non-tuberculous mycobacteria and multidrug-resistant \textit{M. tuberculosis}. The lead compounds in the set, dinitro derivatives 68 and 69 exhibited significant activity against both sensitive and resistant strains of \textit{M. tuberculosis} and also against non-tuberculous mycobacteria.

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|}
\hline
 & R$_1$ & R$_2$ & R$_3$ & R$_4$ \\
\hline
(68) & NO$_2$ & H & NO$_2$ & H \\
(69) & H & NO$_2$ & H & NO$_2$ \\
\hline
\end{tabular}
\caption{}
\end{table}

Vinsonova and his co-workers\textsuperscript{224} synthesized a series of novel 2-substituted 5,7-di-tert-butylbenzoxazole derivatives and evaluated for their potency towards \textit{Mycobacterium tuberculosis}. Benzoxazole 70 showed antituberculotic activity higher than the standard drug isoniazide. \textit{In vitro} cytotoxicity testing of this most active benzoxazole revealed lowest cytotoxicity of the compound amongst the other compounds from the series.
Screening of marine natural products with anti-tuberculosis activity, by Rodriguez et al., from the West Indian gorgonian coral *Pseudopterogorgia elisabethae* resulted in the isolation of two active diterpenoid alkaloids, pseudopteroxazole (71) and seco-pseudopteroxazole (72). Biological screening studies indicated that pseudopteroxazole (71) is a potent growth inhibitor of *Mycobacterium tuberculosis* H37Rv, while seco-pseudopteroxazole (72) shows moderate to strong inhibitorial activity.

In a continuous efforts for the search of novel alkaloid with high antitubercular potency, Rodriguez and his co-workers also isolated Ileabethoxazole (73), a new perhydroacenaphthene-type diterpene alkaloid containing the uncommon benzoazole moiety from the Caribbean sea whip *Pseudopterogorgia elisabethae*. Ileabethoxazole showed 92% inhibition of *Mycobacterium tuberculosis* (H37Rv) at the concentration range of 128–64 μg/mL.
2.2 Chemistry of Synthesized Heterocycles

2.2.1 Pyrazoline

2.2.1.1 Introduction

Pyrazolines (dihydro pyrazole) are five membered heterocyclic compounds containing two nitrogen atoms in an adjacent position having only one endocyclic double bond as shown in 74, 75 or 76. The prefixes 1- (Δ^1), 2- (Δ^2) and 3- (Δ^3) are used to distinguish between 74, 75 and 76 respectively.

Despite their scarcity in nature, the title compounds have found use in many applications including pharmaceuticals, agricultural chemicals and dyes. The literature prior to 1997 is surveyed in a comprehensive review and other useful reviews are available. Amongst all these non-aromatic reduced systems Δ^2-pyrazolines are biologically active, synthetically useful and important molecules. For these reasons, the chemistry of Δ^2-pyrazolines have been the subject of many investigations.

2.2.1.2 Molecular geometry and dimensions of 1,3,5-trisubstituted-2-pyrazolines

The X-ray diffraction analysis studies carried out at -173°C proves that the pyrazoline ring is quite planner and adopts an envelope conformation with the C5 atom bonded to R3 as the flap atom. The molecular dimensions of pyrazoline nucleus have been shown in Table 2.1.
Table 2.1 Molecular geometry of 1,3,5-trisubstituted-2-pyrazolines at -173°C

<table>
<thead>
<tr>
<th>Bond distance (Å)</th>
<th>Bond angle (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N₁-N₂</td>
<td>1.39</td>
</tr>
<tr>
<td>N₂-C₃</td>
<td>1.29</td>
</tr>
<tr>
<td>C₃-C₄</td>
<td>1.50</td>
</tr>
<tr>
<td>C₄-C₅</td>
<td>1.54</td>
</tr>
<tr>
<td>C₅-N₁</td>
<td>1.49</td>
</tr>
<tr>
<td>N₁-N₂-C₃</td>
<td>108</td>
</tr>
<tr>
<td>N₂-C₃-C₄</td>
<td>114</td>
</tr>
<tr>
<td>C₃-C₄-C₅</td>
<td>103</td>
</tr>
<tr>
<td>C₄-C₅-N₁</td>
<td>101</td>
</tr>
<tr>
<td>C₅-N₁-N₂</td>
<td>113</td>
</tr>
</tbody>
</table>

2.2.1.3 Synthesis of 4,5-dihydro-pyrazole-2-en (2-pyrazoline or Δ²-pyrazoline)

The first representatives of pyrazoline were synthesized in the last century. These nitrogen containing heterocyclic compounds became important in the development of different bioactive substances. For these reasons various procedures have been worked out for their synthesis.

The first synthesis of pyrazoline dates back to more than hundred years ago. The synthesis was carried out by Fischer and Knoevenagel by the reaction of phenylhydrazine and acrolein. This is probably the first example of pyrazoline synthesis by the cyclic condensation of an α,β-unsaturated carbonyl compound with a hydrazine derivative. Another milestone in the discovery synthesis was the reaction of diazoalkanes with α,β-unsaturated carboxylic acids and α,β-enones in the early twentieth century. Especially the diazomethane as a nitrogen source made possible the preparation of numerous 1-pyrazolines which could then be converted into their corresponding 2-pyrazoline isomers or cyclopropanes on denitrogenation.
2.2.1.3.1 Preparation of 2-pyrazoline by [3+2] cyclo addition of diazo compounds with α,β-unsaturated carbonyl compound

Dipolar cycloadditions are a powerful class of synthetic reactions for the construction of diversely functionalized five membered heterocycles.244-247

The synthesis of 2-pyrazolines by [3+2] cyclo addition of diazomethane with α,β-unsaturated carbonyl compounds was driven with the first synthesis of diazomethane by Pechmann in 1894. Upon reaction of diazomethane with dimethyl fumarate afforded a pyrazoline-type compound.248 Besides Pechmann also anticipated the mechanism of this reaction, viz. that the primary product of this cycloaddition is 1-pyrazoline (77) which then spontaneously isomerizes into the thermodynamically more stable 2-pyrazoline isomer (78) by 1,3-H shift as shown below.

Another early investigation of synthesis of 2-pyrazoline by reaction of α,β-unsaturated ketone with diazomethane was performed by Azzerallo.249 Formation of 2-pyrazoline derivatives (80) was observed in the reaction of benzalacetone (79) with diazomethane derivatives in anhydrous ether.
Later on the synthesis of 2-pyrazolines by the reaction of chalcones and related α,β-unsaturated ketone with diazomethane has been investigated by several laboratories in early eighteen century.\textsuperscript{250-252}

In developing an approach of synthesizing novel class of amino acids, Mish \textit{et al.} synthesized chiral pyrazoline derivatives \textbf{(82)} by means of asymmetric dipolar cycloaddition of Me\textsubscript{3}SiCHN\textsubscript{2} with α,β-unsaturated carbonyl compounds of type \textbf{81}.\textsuperscript{253} The typical reaction is carried out in less polar hexane-toluene followed by acid work up.

\[
\begin{align*}
\text{R}_1\begin{array}{c}
\text{N} \\
\text{O}
\end{array} \\
\text{O}_2\text{S}
\end{align*}
\begin{array}{c}
\text{N} \\
\text{O}
\end{array} \\
\text{R}_2\begin{array}{c}
\text{O} \\
\text{O}_2\text{S}
\end{array}
\]

\textbf{(81)} \rightarrow \text{Me}_3\text{SiCHN}_2, \begin{array}{c}
23^\circ\text{C}, \text{hexane-toluene} \\
\text{CF}_3\text{COOH}, \text{CH}_2\text{Cl}_2
\end{array}
\begin{array}{c}
\text{R}_1 \\
\text{R}_2
\end{array}
\begin{array}{c}
\text{N} \\
\text{O}
\end{array} \\
\text{O}_2\text{S}
\]

\textbf{(82)}

Analysis of unpurified mixture revealed that the products were obtained in 90-94% diastereoselectivity as single regioisomers. In all cases the diastereomeric products were readily separated by chromatography on silica gel. The configuration of the diastereomers was confirmed by single crystal X-ray crystallographic analysis.

Later, Yamauchi synthesized relatively stable pyrazoline \textbf{(84)} by condensation of 2-methylene-1,3-dicarbonyl compounds \textbf{83} with ethyl diazoacetate in dichloromethane at room temperature.\textsuperscript{254} With the earlier success in enantioselective cycloaddition using chiral lewis acid in Diels-alder reaction, they also examined the possibility of synthesis of optically active pyrazolines.

\[
\begin{array}{c}
\text{O} \\
\text{O}
\end{array} \\
\text{R}
\]

\textbf{(83)} \rightarrow \text{N}_2\text{CHCO}_2\text{C}_2\text{H}_5, \text{CH}_2\text{Cl}_2
\begin{array}{c}
\text{C}_2\text{H}_5\text{O}_2\text{C} \\
\text{N}\text{N}
\end{array} \\
\text{R}
\]

\textbf{(84)}

However, the products show no specific optical rotation. This means the reactivity of \textbf{83} is too high to introduce the chirality into the cyclo adduct.
In continuous efforts to synthesize optically active pyrazolines (86), Taichi Kano and his co-workers, developed enantioselective 1,3-dipolar cycloaddition reaction between diazoacetates and α-substituted acroleins (85) using chiral titanium BINOLate lewis acids. The reaction resulted in formation of highly functionalized pyrazolines with quaternary stereogenic center with good enantiomeric excess.

Radhakrishna et al. reported the facile one-pot synthesis of functionalized pyrazolines from different electron poor olefins/alkenes including Baylis–Hillman adducts (87). The 1,3-dipolar cycloaddition of ethyl diazoacetate with various olefins were mediated through InCl3 and/or DABCO, under solvent free conditions at ambient temperature to afford 3,5-disubstituted pyrazoline (88) in good yields.

### 2.2.1.3.2 Preparation of 2-pyrazolines by reaction of α,β-unsaturated ketones with hydrazines or their derivatives

As described earlier, the first reactions of hydrazines with α,β-unsaturated carbonyl compounds to form pyrazolines were performed in the late nineteenth and early twentieth centuries. One group of such carbonyl compounds was the α,β-unsaturated aldehydes (89) which gave hydrazones (90) on reactions with hydrazines. Hydrazones (90) can then be easily converted into 2-pyrazolines on treatment with acid. The hydrazone type intermediates were not isolated in each
case, but they were assumed to be the primary reaction products of various $\alpha,\beta$-unsaturated aldehydes and hydrazines. As a rate-controlling step in pyrazoline formation, the addition of the NH group to the carbon-carbon double bond of hydrazones (90) was considered. The solubility, stereochemistry, and electron distribution of hydrazones may influence this ring closure reaction providing 2-pyrazolines.

![Chemical structure](image)

Reaction of $\alpha,\beta$-unsaturated carbonyl compound with hydrazine in acetic acid with or without the isolation of the intermediate hydrazone (90) is a commonly used efficient procedure to prepare 2-pyrazolines in high yields.\textsuperscript{260-261} Hydrazines and $\alpha,\beta$-unsaturated ketones have also been allowed to react in a hot alcohol solution\textsuperscript{262,263} or in a boiling mixture of benzene with ethanol.\textsuperscript{264} Reaction of $\alpha,\beta$-unsaturated ketones with phenylhydrazine in refluxing pyridine afforded 2-pyrazolines as well.\textsuperscript{265}

![Chemical structure](image)

The mechanism of these reactions providing 2-pyrazolines has also been investigated under various reaction conditions. On the basis of numerous experimental results, it can be concluded that the reaction of $\alpha,\beta$-unsaturated ketones and hydrazines in acidic medium leads to 2-pyrazolines via hydrazones (92) as intermediates.\textsuperscript{266}
However, in the presence of piperidine, Michael addition takes place providing β-hydrazinoketones (91) instead of hydrazones formed from the same starting materials in acetic acid.\(^{267}\)

The reaction of hydrazines with ketones possessing more than one double bond conjugated with the carbonyl group (93) has also been investigated.\(^{237,268}\) The reaction resulted in the formation of similar kind of 2-pyrazolines (94) as was obtained by the reactions of hydrazines with α,β-unsaturated ketones.

Thus, the formation of 2-pyrazolines is based mainly on one of the α,β-unsaturated moieties of the substrate, and the other parts of the molecule have almost no influence on the result of the reaction.

Another root of synthesis of pyrazoline of type 96 has been developed by the condensation of acylhydrazines with 2-yne-1-ones (95) derivatives at refluxing temperature in an organic solvent.\(^{269,270}\)

Katritzky \textit{et al.} synthesized 4-benzotriazolylpyrazoline derivatives (98) by condensation of hydrazine with benzotriazolyl unsaturated ketone derivatives (97).\(^{271}\)

In this intermediate 98, a high leaving ability of benzo triazolyl moiety renders the adjacent ring hydrogen acidic and thus allows for its replacement using an
electrophilic reagent, so that $\text{98}$ can be further functionalized to afford tetrasubstituted pyrazolines ($\text{99}$).

Gembus in 2010, investigated polymer supported (PS) one-pot synthesis of pyrazolines ($\text{103}$) by means of site isolated resin bound methodology. The method involves base-catalysed aza-Michael addition of hydrazones ($\text{100}$) to terminal enones ($\text{101}$), followed by an acid-catalysed transimination reaction of $\text{102}$ to yield $\text{103}$.

Typical reaction was carried out using a mixture of commercially available polymer supported heterogeneous tosic acid (PS-TsOH, Amberlyst-A15) and PS-TBD (1,5,7-triazabicyclo[4.4.0]dec-5-ene) providing the isolation of product by simple filtration, thus facilitating the construction of chemical libraries.
2.2.1.3.3 Preparation of 2-pyrazolines by reaction of chalcone dibromide with hydrazine

2-Pyrazolines have also been synthesized by reaction of chalcone dibromides (104) with hydrazines. The mechanism of the reaction has not been investigated, but it was supposed that chalcones are generated in situ by debromination, which then react with hydrazines to afford 2-pyrazolines.

\[
\begin{align*}
\text{Br} & \quad \text{Br} \\
R_2 & \quad R_1 \\
\text{O} & \quad \text{R}_2 \\
R_2 & \quad \text{R}_1 \\
\end{align*}
\]

(104)

2.2.1.3.4 Preparation of 2-pyrazolines by the reaction of chalcone epoxide with hydrazine

Neubauer et al. synthesized 3,5-disubstituted-2-pyrazolines by the reaction of chalcone epoxides (105) and hydrazine. Based on the experiments it has been concluded that hydrazone (106) is formed as a stable intermediate, the ring closure of which affords 3,5-disubstituted-4-hydroxy-2-pyrazolines.

\[
\begin{align*}
\text{O} & \quad \text{O} \\
\text{R}_2 & \quad \text{R}_1 \\
\text{R}_2 & \quad \text{R}_1 \\
\text{H}_2\text{N} & \quad \text{NH}_2 \\
\text{H} & \quad \text{N} \\
\end{align*}
\]

(105)

\[
\begin{align*}
\text{N} & \quad \text{N} \\
\text{R}_2 & \quad \text{R}_1 \\
\text{NH}_2 & \quad \text{OH} \\
\text{H} & \quad \text{N} \\
\end{align*}
\]

(106)

2.2.1.3.5 Preparation of 2-pyrazolines by the reaction of flavanone with hydrazine

Kallay and co-workers developed a very facile procedure for the preparation of 2-pyrazolines by the reaction of flavanones (107) with hydrazines.
However, as flavanones were obtained from the 2-hydroxychalcones as starting material, this procedure did not get much attention.

2.2.1.3.6 Preparation of 2-pyrazolines by some of the other recently developed methods

Singh et al. developed unique route for the synthesis of pyrazoline spiro adducts (111) with high yields and excellent regio- and diastereoselectivities by [3+2] dipolar cycloaddition reaction of nitrile imines (109) with 3-alkylidene oxindoles (110). An optimum method involves the reaction of 3-alkylidene oxindoles (110) with hydrozonooyl bromide (108) in dichloridemethane solvent and in presence of organic base triethylamine. The nitrile imine (109) generated in situ (by dehydrohalogenation) undergoes [3+2] cycloaddition with 3-alkylidene oxindoles (110) giving pyrazoline spiroadducts of indole (111).
Ju and Varma described a one-pot synthesis of pyrazoline from alkyl halides (112) with hydrazine via efficient cycloaddition in an alkaline aqueous media under microwave irradiation.\(^{280}\)

![Chemical structure](attachment:image)

Important feature of this reaction is, though the reaction is not homogeneous, the selective absorption of MW by polar molecules and intermediate in multiple system substitutes as phase transfer catalyst without using any phase transfer reagent. Even in large scale experiments, the phase separation of the desired product in either solid or liquid form from the aqueous media can facilitate the product formation by simple filtration or decantation.

Alex et al. developed novel regioselective synthesis of aryl substituted pyrazolines.\(^{281}\) A typical method involves the reaction of aryl hydrazine with 3-butynol (113) in presence of zinc triflate to give pyrazoline derivatives in excellent yields. As per the proposed mechanism hydrohydrazination of 3-butynol gave the corresponding arylhydrazone. In general, the arylhydrazone undergoes Fischer indole cyclization in the presence of a stochiometric amount of Lewis acid, such as ZnCl\(_2\).\(^{282}\) However, in the case of 3-butynol, the pyrazoline was formed by an unusual nucleophilic substitution of the hydroxy group.

![Chemical structure](attachment:image)
2.2.2 Benzoxazole

2.2.2.1 Introduction

Benzoxazoles (114) is a heterocyclic compound where the benzene ring is fused with oxazole ring structure. Benzoxazole is a planar molecule and the lone pair of electrons on nitrogen, which is coplanar with heterocyclic ring and therefore involved in delocalization, confers weakly basic properties to the parent ring.

\[
\begin{array}{c}
\text{\textbf{(114)}} \\
\end{array}
\]

Naturally occurring benzoxazole derivatives, either isolated from plants or produced by total synthesis have shown remarkable biological activities. It is also found in the chemical structures of various pharmaceutical drugs such as flunoxaprofen, chloroxazone, calcimycin, etc.

2.2.2.2 Synthesis of 2-substituted benzoxazole derivatives

2.2.2.2.1 Synthesis of 2-alkyl/aryl benzoxazole derivatives

The condensation of 2-aminophenol (115) derivatives with respective carboxylic acid (116) is a common method for the synthesis of 2-substituted benzoxazole (117) derivatives. As shown in the scheme below, a large number of literature highlights the use of different reagents and reaction conditions for the synthesis of benzoxazoles.\textsuperscript{283-290} The typical reaction routes are themselves simple, one step with high conversion of reactants into products. Also, the chemists find an ease in removing the unconsumed starting materials and reagents by simple work up. The inorganic reagents can be removed simply by treatment with water while the starting materials can be removed through a simple acid-base purification.
Nieddu and Giacomelli synthesized a series of various poly-substituted benzoxazoles (117) from carboxylic acids (116) and 2-aminophenols (115) using a method based on cyanuric chloride (118)/microwave acid activation.291

A mild activation process allows the use of a wide range of starting materials, thus allowing the functionalization of final products, and avoids extra steps to prepare
starting materials. Furthermore, cheap and easy-to-handle reagents, and simple workup make this method very practical.

Fengjiang Wang and co-workers carried out a solid phase synthesis of benzoxazole. In this typical reaction, 2-amidophenol attached to a solid support (118) is converted to the corresponding benzoxazole derivative by treatment with triphenylphosphine (Ph₃P) and diethyl azodicarboxylate (DEAD) in THF at room temperature in high yields.²⁹²

\[
\text{Ph₃P, DEAD, THF} \xrightarrow{\text{1,4-dioxane, MW, reflux}} \text{N} \hspace{1cm} O
\]

The microwave irradiated solution-phase synthesis of benzoxazoles has been developed by Pottorf et al. using readily available reagents. Short reaction times, no additives and simple workup are the main advantages of this method. The key reaction involves the condensation between 2-aminophenol with substituted benzoyl chloride (119).²⁹³

\[
\text{R} \hspace{1cm} \text{O} \hspace{1cm} \text{N} \hspace{1cm} \text{R} \hspace{1cm} \text{Cl} \hspace{1cm} \text{1,4-dioxane} \hspace{1cm} \text{MW, reflux}
\]

Chang and co-workers synthesized benzoxazoles by the condensation arylaldehydes with 2-aminophenols and subsequent 2,3-dichloro-5,6-dicyano-1,4-benzoquinone (DDQ) promoted oxidative cyclization reactions. The resulting 2-arylbenzoxazoles
(120) were isolated with good yields from the reduced DDQ byproduct by treatment of reaction mixture with a strongly basic ion-exchange resin. Thus, the combination of this procedure with the use of basic ion-exchange resin can be useful for the combinatorial library synthesis.294

\[
\begin{align*}
\text{NH}_2 \quad & \text{OH} \\
\text{R}_1 \quad & \text{N} \quad \text{Ar} \\
\text{OH} \\
\text{DDQ} \quad & \text{CH}_2\text{Cl}_2 \\
\text{R}_1 \quad & \text{N} \quad \text{Ar}
\end{align*}
\]

Gadakh et al. synthesized benzoazole derivatives (122) from oximes (121). In a typical reaction, the oxime, upon refluxing with phosphorous oxychloride, undergoes Beckmann rearrangement with a ring closure afford benzoazole.295

\[
\begin{align*}
\text{POCl}_3, \quad & \text{110}^\circ\text{C} \\
\text{R}_1 \quad & \text{N} \quad \text{Ar} \\
\text{R}_2 \\
\text{HO} \\
\text{R}_1 \quad & \text{N} \quad \text{Ar}
\end{align*}
\]

The 2-pyranosyl benzoazole derivatives (125) were obtained by reaction of per-O-acetylated-β-d-pyranosyl nitrile oxides (124), generated by dehydrochlorination of the corresponding hydroximoyl chlorides (123), with 2-aminophenol.296
Yoshizumi et al. reported direct arylation of 1,3-benzoazole compounds with aryl iodides (126) to produce corresponding 2-arylated benzoxazoles with good yields. The reaction is promoted by CuI with use of PPh₃ and Na₂CO₃ or K₂CO₃ as ligand and base, respectively, in DMF or DMSO as solvents.

![Reaction Scheme 126](image)

The direct arylation of 1,3-benzoazole with aryl halides (127) through the palladium catalysed C-C coupling to produce the corresponding 2-arylated benzoxazoles has also been reported by Derridj et al.

![Reaction Scheme 127](image)

Wang et al. developed a protocol for the palladium-catalyzed direct desulfitative arylation of azoles at C-2 position using sodium arylsulfimates (128) as the aryl sources. Azoles including benzoxazoles, oxazoles, benzothiazole reacted with sodium arylsulfimates smoothly to generate the corresponding products in good to excellent yields.

![Reaction Scheme 128](image)

Another palladium catalyzed arylation at 2nd position of benzoxazole using iodo benzene diacetate (129) has been reported by Peng and co-workers. The reaction provides a novel methodology allowing for a wide functional group tolerance.
Hamdy et al. reported a new approach to arylated benzoxazoles by site-selective Suzuki–Miyaura cross-coupling reactions of 2,6-dichlorobenzoxazole (130) with arylboronic acids (131). The reaction proceeds with excellent site-selectivity in favour of position C-2, which is more electron deficient than position C-6.\textsuperscript{301}

An oxidative alkynylation of benzoxazole with terminal alkynes (132) under solvent and ligand-free condition has been described by Patil et al.\textsuperscript{302} A typical reaction was carried out in presence of ferrous chloride as catalyst and (t-BuO)\textsubscript{2} as an oxidant. Many literature reports the coupling of terminal alkynes with azoles under oxidative catalytic conditions using Au,\textsuperscript{303} Cu,\textsuperscript{304} Ni\textsuperscript{305} and Pd\textsuperscript{306} catalytic systems. However, these catalysts are derived from heavy or rare metals and their toxicity and prohibitive prices constitute severe drawbacks for their applications. In contrast, iron being the most abundant metal on earth, and does not require coordination with expensive and toxic ligands, thus making the reaction much efficient and versatile.

Gerelle et al. demonstrated a straightforward procedure for the palladium-catalyzed direct functionalization of benzoxazoles with alkenyl iodides (133).\textsuperscript{307} The use of palladium catalysis and ‘on water’ conditions allows relatively short reaction times, and moderate to good yields to be achieved across a variety of substrates.
2.2.2.2 Synthesis of 2-aminobenzoxazole derivatives

Toyoyuki and his co-workers prepared 2-amino benzoxazole derivatives by condensing 2-aminophenol derivatives with cyanoguanidine (134).\(^{308}\)

\[
\begin{align*}
\text{NH}_2 & \text{N} \quad \text{NH} \\
\text{H}_2 & \text{N} \quad \text{NH} \\
\text{O} & \text{N} \\
\text{R}_1 & \text{R}_2 \\
\text{NH}_2 & \text{N} \\
\end{align*}
\]

(134)

Berg and Parnell reported the synthesis of 2-aminobenzoxazole by a simple intramolecular cyclization of 1-(2-hydroxy5-nitrophenyl)guanidine (135).\(^{309}\)

\[
\begin{align*}
\text{O}_2 & \text{N} \\
\text{N} & \text{N} \\
\text{H}_2 & \text{N} \\
\text{O}_2 & \text{N} \\
\text{R}_1 & \text{R}_2 \\
\text{R}_3 & \text{R}_4 \\
\end{align*}
\]

(135)

Joseph and James reported one of the most efficient methods for the synthesis of 2-aminobenzoxazole. The typical reaction involves the condensation between 2-aminophenols and cyanogen bromide (136) in presence of a base.\(^{310}\)

\[
\begin{align*}
\text{R} & \text{R} \\
\text{NH}_2 & \text{N} \\
\text{O} & \text{N} \\
\text{Br} & \text{N} \\
\text{NaHCO}_3 & \text{R} \\
\end{align*}
\]

(136)

Due to high lachrymatory property of cyanogen bromide, the method has now been modified making use of other mild reagents. Yong-Qian Wu et al. reported the use of di(imidazole-1-yl)methanimine (137) for the synthesis of 2-amino benoxazole.\(^{311}\)
Wishart et al. reported the anhydrous copper sulphate mediated intra-molecular cyclization of 1-(2-hydroxy-5-substitutedphenyl) thiourea (138) into 2-amino benzoxazole derivative.\textsuperscript{312}

\[ \text{R-NH}_2 + \text{NH}_2\text{NH} \rightarrow \text{R-}N\text{O}_2\text{H} \]

Ji Young et al. prepared 2-amino benzoxazole derivatives through cobalt and manganese catalysed direct amination of benzoxazole in presence of tert-butyl hydrogen peroxide.\textsuperscript{313}

\[ \text{Co(OAc)}_2/\text{Mn(OAc)}_2/\text{t-BuOOH} \rightarrow \text{R-}N\text{H}_2 \]

Zhenping et al. reported an efficient method for the synthesis of the 2-amino benzoxazoles through cyclodesulfurization of N-substituted-2-hydroxy-phenylthioureas (139).\textsuperscript{314} These thiourea intermediates (139) are easily derived from the condensation of an appropriately substituted 2-aminophenol and alkyl or arylisothiocyanates.

\[ \text{R-NH}_2 + \text{R-NCS} \rightarrow \text{R-NH}_2 \]

A rapid and efficient one-pot method for the synthesis of 2-aminobenzimidazoles mediated by a polymer-supported (PS) carbodimide has been reported by Cee and
Downing.\textsuperscript{315} This procedure is successful with a wide range of structurally diverse 2-aminophenoles and isothiocyanate substrates (140), and does not require the isolation of the intermediate thiourea.

\[
\text{OH} \quad \text{NH}_2 \quad \text{O} \quad \text{N} \quad \text{NH} \\
\begin{array}{c}
\text{PS-Carbodimide,} \\
70^\circ\text{C, THF}
\end{array}
\]

Benjamin \textit{et al} reported the synthesis of 2-amino benzoazoles by amine displacement of 2-halogenated precursor of benzoxazole.\textsuperscript{316} The procedure involves deprotonation at the 2-position of the benzoxazole and quenching the intermediate organolithium species with a halogen electrophile. The 2-halobenzoxazole is then treated in the same pot with an amine nucleophile to afford the desired product.

\[
\begin{array}{c}
\text{O} \\
\text{N} \\
R_1 R_2 \\
\text{H} \\
\text{N} \\
\text{O} \\
\end{array}
\]

\[
\text{i. LiHMDS, THF, -20}^\circ\text{C} \\
\text{ii. NBS, THF, -20}^\circ\text{C} \\
\text{iii. } R_1 N R_2
\]

Yogesh \textit{et al}. reported a mild, and metal-free protocol for the synthesis of 2-aminobenzoxazoles using direct oxidative C–H bond amination of benzoxazoles with secondary or primary amines.\textsuperscript{317} The reaction was performed using catalytic amount of \textit{N}-iodosuccinimide (NIS) and aqueous hydrogen peroxide as a green oxidant. Reaction proceeds smoothly at ambient temperature and requires shorter reaction time to furnish excellent yield of the desired products.

\[
\begin{array}{c}
\text{NIS, AcOH} \\
\text{Aq. H}_2\text{O}_2 \\
\text{ACN, rt}
\end{array}
\]

Another facile method has also been developed which proceeds through C–H bond amination of benzoxazoles with amines through a ring-opening and subsequent ring-closure approach.\textsuperscript{318}
The reaction proceeds in two subsequent steps wherein, initially a nucleophilic addition of amines across benzoxazoles takes place in the absence of any reagent or catalyst under solvent-free condition, followed by oxidative ring closure using 2-iodoxybenzoic acid (141) as a hypervalent iodine reagent.

Joon Kim et al reported the synthesis of 2-amido benzoxazole (144) derivatives by condensing N-(bis-methylsulfanylmethylene) amide derivatives (143) (produced in situ by reacting substituted benzamides (142) with carbon disulfide, iodomethane and sodium hydride at room temperature) with 2-aminophenol.

2.2.2.2.3 Synthesis of 2-mercaptobenzoxazole derivatives

One efficient method for the synthesis of 2-mercaptobenzoxazole is the condensation of 2-aminophenol with potassium ethyl xanthate (145) (produced in situ by treating CS$_2$ with KOH in ethanol). Hydrogen sulfide gas evolves during the course of the reaction. The isolation of final product is carried out by acidic work up which also facilitates the removal of unconsumed 2-aminophenol.
In cases where the amine is weak nucleophile due to the presence of electron withdrawing groups present in the ring the above method produces poor yield. In these cases strong reagent, thiophosgene (146) is used. The reaction is usually carried out in aprotic solvents.324,325

Although the above method produces good practical yield, one finds the difficulty in handling the thiophosgene because of its high volatility, lachrymatory and carcinogenic properties. Thus, the method has now been modified by making use of mild reagent (1H-imidazol-1-yl)methanethione (147).

Arisawa et al. reported a rhodium-catalyzed phenylthiolation reaction of benzoxazole using α-(phenylthio)isobutyrophenone (148).326 In the presence of catalytic amounts of RhH(PPh3)4 and 1,2-bis(diphenylphosphino)ethane (dppe), benzoxazoles reacted with α-(phenylthio)isobutyrophenone (148) giving 2-phenylthio derivatives (149).
Review of Literature

\[
\text{N} + \text{PhCHO} \quad \xrightarrow{\text{RhH(PPh}_3\text{)}_4, \text{ dppe}} \quad \begin{array}{c}
\text{N} \quad \text{Ph} \\
\text{O}
\end{array}
\]

(149)

(148)
2.3 Quantitative Structural Activity Relationship (QSAR)

2.3.1 Introduction

Most molecular discoveries today are the results of an iterative, three-phase cycle of design, synthesis and test. Analysis of the results from one iteration provides information and knowledge that enables the next cycle of discovery to be initiated and further improvement to be achieved. A common feature of this analysis stage is the construction of some form of computational model which enables the observed activity or properties to be related to the molecular structure. Such models are often referred to as Quantitative Structure Activity Relationships.\textsuperscript{327}

Quantitative structure-activity relationship studies unquestionably are of great importance in modern chemistry and biochemistry. The concept is to transform searches for compounds with desired properties using chemical intuition and experience into a mathematically quantified and computerized form. QSAR methods are characterized by two assumptions with respect to the relationship between chemical structure and the biological potency of compounds. The first is that one can derive a quantitative measure from the structural properties significant to the biological activity of a compound. The properties are assumed to be physicochemical such as partition coefficient or sub structural such as presence or absence of certain chemical features. The other assumption is that one can mathematically describe the relationship between biological property and physico-chemical properties.\textsuperscript{328} QSAR’s general mathematical form is represented by the following general equation (1).

$$\text{Biological Activity} = f(\text{Physicochemical Property}) \quad (1)$$

2.3.2 Objectives of QSAR

QSAR attempts to correlate structural, chemical and physical properties with biological activity by various statistical approaches.\textsuperscript{329} QSAR models are scientific credible tool for predicting and classifying biological activities of untested chemical entities. QSAR is an essential tool for lead development (optimization). The growing
trend is to use QSAR early in drug discovery process as a screening and enrichment tool to eliminate development of those chemicals lacking “drug like” properties or those predicted to elicit a toxic response.330

2.3.3 Historical development

QSAR modeling was born in the toxicology field. In 1863, A.F.A Cros observed that the toxicity of alcohols in mammals increased as the solubility of alcohols in water decreased.331 Crum-Brown and Fraser expressed the idea that the physiological action of substance was function of its chemical composition and constitution and published equation (2) in 1868 which is considered to be the first formulation of a quantitative structure-activity relationship: the “physiological activity” $\Phi$ was expressed as a function of the chemical structure $C$.332

$$\Phi = f(C)$$

A few decades later, in 1893, Richet showed that the cytotoxicities of a diverse set of simple organic molecules were inversely related to their water solubilities.333 In 1900’s, Meyer and Overton independently suggested that narcotic (depressant) action of a group of organic compounds paralleled their oil/water partition coefficient.334 L. Hammet in mid 1930s, correlated the electronic properties of organic acids and bases with their equilibrium properties. He found that a relationship resulted when different substitutions were made to aromatic compounds. The technique was introduced by Hansch et al. in the early1960s.335-336 The approach stemmed from linear free-energy relationship in general and the Hammett equation in particular.337 It is based on the assumption that differences in physicochemical properties accounts for the differences in biological activities of compounds. According to this approach, the changes in physicochemical properties that affect the biological activities of a set of congeners are of major three types viz. electronic, steric, and hydrophobic.338
2.3.3.1 Hansch analysis

Corwin Hansch, may be the best known as the father of concept of QSAR, described the quantitative correlation of the physic-chemical properties of molecules with their biological activities. It is also known as linear free energy relationship (LFER) or extrathermodynamic approach. Its basic assumption is that the effect of substituents on the strength of interactions between drug and its receptor and other biomolecules is an additive combination of the effect of the substituents on various types of simpler model intermolecular interactions. From physical chemistry, the interactions were assumed to be electrostatic, steric and hydrophobic in nature. He suggested the linear (equation (3)) and non-linear (equation (4)) dependence of biological activity on different parameters.

\[
\log BA = a(\log P) + b\sigma + cEs + d \quad \text{............linear (3)}
\]

\[
\log BA = a(\log P)^2 + b(\log P) + c\sigma + dEs + e \quad \text{............non-linear (4)}
\]

Where \(\log BA\) is the logarithmic form of biological activity, a-e are constants determined for particular biological activity by multiple linear regression analysis. \(\log P, \sigma, Es\) represents lipophilic, electronic and steric properties, respectively. The Equation (4) was developed from the two assumptions as proposed by the Hansch; (i) The transport of a drug from the site of application to its site of action depends on the lipophilicity of the drugs (in a non-linear manner) as shown in Figure 2.1.

![Figure 2.1 Non-linear trend in biological activity with respect to lipophilicity](image)
(ii) The binding affinity (interaction) of drug to its biological counterpart, such as an enzyme or receptor depends on the lipophilicity, the electronic properties and other linear free energy related properties.

2.3.3.2 Free Wilson analysis

The Free-Wilson approach is truly a structure-activity based methodology because it incorporates the contribution made by various structural fragments to the overall biological activity. Indicator variables are used to denote the presence or absence of a particular structural feature. It is represented by equation (5).

\[
BA = \Sigma a_i X_i + \mu
\]

Where BA is the biological activity, \( \mu \) is the overall activity, \( a_i \) is the contribution of each structural feature, \( X_i \) denotes the presence (\( X_i = 1 \)) or absence (\( X_i = 0 \)) of a particular structural fragment. This mathematical model incorporated symmetry equation to minimize linear dependence between variables. This approach was easy to apply; it had its drawbacks, mostly centered on the large number of parameters and subsequent loss of the statistical degree of freedom. In 1971, in an attempt to deal with limitations of this approach, Fujita and Ban proposed a simplified approach that solely focused on the additivity of group contribution (equation (6)).

\[
\log \frac{A}{A_0} = \Sigma G_i X_i
\]

where \( A \) and \( A_0 \) represents the biological activity of the substituted and unsubstituted compounds respectively, while \( G_i \) was the activity of the \( i^{th} \) substituent and \( X_i \) had the value of 1 or 0 that corresponded to the presence or absence of that substituent.

2.3.4 QSAR methodology

QSAR is a method of finding a simple mathematical equation that can be used to calculate some biological property from the physico-chemical and structural features of a compound. QSAR attempts to correlate structural molecular features and physicochemical properties (descriptors) with biological activities for a set of compounds, by means of various statistical methods. As a result, a simple
mathematical relationship is established which is known as QSAR model. Applications of QSAR can be extended to any molecular design purpose, including prediction of different kinds of biological activities of unsynthesized compounds, lead compound optimization, prediction of novel structural leads in drug discovery and molecular database searching for the hits.

The process of building a QSAR model is similar, apart from what type of property is being predicted. This consists of several steps as shown in the Figure 2.2.

![Figure 2.2 Schematic diagram of a typical QSAR guided drug design](image)

### 2.3.4.1 Experimental data

The first step in building a model is to select the training set and test set of compounds with their experimental biological activities. Ideally, each of these activities should cover the range of possible values (potent to least potent) for that activity. The training set data is used to formulate the statistical model and the test set is used to validate the model.
2.3.4.2 Molecular descriptors used in QSAR

Molecular descriptors (Figure 2.3) can be defined as a numerical representation of chemical information encoded within a molecular structure via mathematical procedure.\(^{343}\) This mathematical representation has to be invariant to the molecule’s size and number of atoms to allow model building with statistical methods.

![Diagram of molecular descriptors](image)

**Figure 2.3** Various physico-chemical and structural descriptors of query molecule

The information content of structural descriptors depends on two major factors: (i) The molecular representation of compounds. (ii) The algorithm which is used for the calculation of the descriptor.\(^{344}\) The molecular descriptors can be distinguished in two broad families depending on the information about 3D orientation and conformation of the molecule. These are 2D QSAR Descriptors and 3D QSAR Descriptors. It is generally assumed that 3D approaches are superior to 2D in drug design. Yet, studies show such an assumption may not always hold.

### 2.3.4.2.1 2D QSAR descriptors

2D-QSAR descriptors are defined as numerical properties that can be calculated from the connection table representation of molecule. These descriptors are independent of conformation of molecule. Following table enlists commonly used 2D descriptors.
Table 2.2 List of some of the representative 2D QSAR descriptors

<table>
<thead>
<tr>
<th>Type</th>
<th>Representation of properties</th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constitutional descriptors</td>
<td>Capture properties of the molecule that are related to elements constituting its structure</td>
<td>H bond donor count, H bond acceptor count, total no. of single, double, triple or aromatic bonds, total no. of aromatic rings, etc.</td>
</tr>
<tr>
<td>Electronic descriptors</td>
<td>Electrostatic descriptors capture information on electronic nature of the molecule</td>
<td>Dipole moment, energy of highest occupied molecular orbital (HOMO), energy of lowest occupied molecular orbital (LUMO), superdelocalizability, sum of atomic polarizability, etc.</td>
</tr>
<tr>
<td>Thermodynamic descriptors</td>
<td>Capture information on thermodynamic properties of the molecule</td>
<td>Log of partition coefficient, heat of formation, molar refractivity, etc.</td>
</tr>
<tr>
<td>Topological descriptors</td>
<td>Topological descriptors are derived from hydrogen-suppressed molecular graphs, in which the atoms are represented by vertices and the bonds by edges. The connections between the atoms can be described by various types of topological matrices (e.g., distance or adjacency matrices)</td>
<td>Wiener index$^{345}$, Randic indices$^x$$^{346}$, Balaban's J index$^{347}$, Schultz index$^{348}$, Kier and Hall indices$^{xv}$$^{349}$, etc.</td>
</tr>
<tr>
<td>Fragment-based descriptors</td>
<td>Generalisation of feature counts based on 2D fingerprints</td>
<td>BCI fingerprints$^{350}$, Hologram QSAR$^{351}$, etc.</td>
</tr>
</tbody>
</table>
2.3.4.2.2 3D QSAR descriptors

There are mainly two types of molecular descriptors.

(i) Those which depend on internal co-ordinate only. These are classified as “i3D”.

(ii) Those which depend on absolute orientation. These are classified as “α3D”.

The lists of 3D QSAR Descriptors have been described in the following Table 2.3.

Table 2.3 List of some of the representative 3D QSAR descriptors

<table>
<thead>
<tr>
<th>Type</th>
<th>Representation of Properties</th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potential energy descriptors</td>
<td>The energy descriptors use the MOE potential energy model to calculate energetic quantities from stored conformation</td>
<td>Angle bend potential energy, out of plane potential energy, electronic interaction potential energy, solvation energy, etc</td>
</tr>
<tr>
<td>Surface area, volume and shape descriptors</td>
<td>Depends upon the structure connectivity and conformation</td>
<td>Molecular surface area(^{352}), molecular volume(^{353}), molecular shadow(^{354}), principle moment of inertia(^{355}), etc</td>
</tr>
<tr>
<td>Conformation dependant charge descriptors</td>
<td>Depends upon the stored partial charges of molecules and their conformations</td>
<td>Dipole moments and it’s X, Y and Z components(^{356}), Jurs descriptors(^{357}), total solvent accessible surface area(^{358}), etc</td>
</tr>
<tr>
<td>WHIM descriptors(^{359-361})</td>
<td>Based on statistical indices calculated on the projections of atom along principle axe</td>
<td>Directional WHIM descriptors, global WHIM descriptors</td>
</tr>
</tbody>
</table>
2.3.4.3 Selection of molecular descriptors

Nowadays it is possible to calculate numerous descriptors using computer-based algorithms. Once descriptors have been calculated using various empirical and semi-empirical protocols, it is necessary to pick which should be included in the QSAR model.

However, it is difficult to predict in advance which descriptors will be valuable as to be accounted for in the study. The purpose of descriptor selection is to remove descriptors irrelevant or negligible to the activity of the compounds, so as to improve computation speed, performance and interpretability of predictive models. Irrelevant and redundant descriptors are removed either by using a filter or a wrapper approach or a combination of both of these approaches.

2.3.4.4 Statistical methods for the generation of QSAR models

Given the selected descriptors, the final step in building the QSAR model is to derive the mapping between the activity and the values of the descriptors. Simple, yet useful methods model the activity as a linear function of the descriptors. Other, non-linear, methods extend this approach to more complex relations. Another important division of the mapping methods is based on the nature of the activity variable. In case of predicting a continuous value a regression problem is encountered. When only some categories of classes of the activity need to be predicted, e.g. partitioning compounds
into active and inactive, the classification problem occurs. In regression, the dependent variable is modeled as a function of the descriptors, as noted above. In classification framework, the resulting model is defined by a decision boundary, separating the classes in the descriptor space.

2.3.4.4.1 Linear models

Linear models have been the basis of QSAR analysis since its beginning. They predict the activity as a linear function of molecular descriptors. In general, linear models are easily interpretable and sufficiently accurate for small datasets of similar compounds, especially when the descriptors are carefully selected for a given activity.

**Multiple linear regression (MLR)**

Multiple linear regression (MLR) models the activity to be predicted as a linear function of all descriptors. Based on the examples from the training set, the coefficients of the function are estimated. These free parameters are chosen to minimize the squares of the errors between the predicted and the actual activity. The main restriction of MLR analysis is the case of large descriptors-to-compounds ratio or multi collinear descriptors in general. This makes the problem ill-conditioned and makes the results unstable. Multiple linear regression is among the most widely used mapping methods in QSAR in last decades.

**Partial least squares (PLS)**

Partial least squares (PLS)\(^{362-364}\) linear regression is a method suitable for overcoming the problems in MLR related to multi collinear or over-abundant descriptors. The technique assumes that despite the large number of descriptors the modeled process is governed by a relatively small number of latent independent variables. The PLS indirectly tries to obtain knowledge on the latent variables by decomposing the input matrix of descriptors into two components, the scores and the loadings. The scores are orthogonal and, while being able to capture the descriptor information, allow also for good prediction of the activity. The estimation of score
vectors is done iteratively. The first one is derived using the first eigenvector of the activity descriptor combined variance-covariance matrix. Next, the descriptor matrix is deflated by subtracting the information explained by the first score vector. The resulting matrix is used in the derivation of the second score vector, which followed by consecutive deflation, closes the iteration loop. In each iteration step, the coefficient relating the score vector to the activity is also determined.

**Linear discriminant analysis (LDA)**

The aim of discriminant analysis (LDA) is to try and separate the molecules into their constituent classes. Discriminant analysis finds a linear combination of factor that best discriminate between different classes. Linear discriminant analysis was used for the analysis rather than multiple linear regression since the biological activity data were not on a continuous scale of activity but rather were classified into two groups: active and inactive. It is used to obtain a qualitative association between molecular descriptor and biological property.\(^{365}\)

**Non-linear models**

Non-linear models extend the structure-activity relationships to non-linear functions of input descriptors. Such models may become more accurate, especially for large and diverse datasets. However, usually, they are harder to interpret.

**Artificial neural networks (ANN)**

Artificial neural networks (ANN)\(^{366}\) are biologically inspired prediction methods based on the architecture of a network of neurons. A wide range of specific models based on this paradigm have been analyzed, amongst which perceptron based and the radial basis function based are the prevailing ones. Both these methods fall into the category of feed-forward networks, in which, during the prediction, the information flows only in direction from the input descriptors, through a set of layers, to the output of the network.
2.3.4.4.3 Decision trees

Decision trees (DT)\textsuperscript{367-368} classification model consists of a tree-like structure consisting of nodes and links. DT is a powerful data mining-method suitable for knowledge discovery from a small set of data while conveniently accepting both categorical and numerical data. DT makes no underlying assumptions regarding the distribution of the values of the predictors. The construction of a tree involves growing the tree to facilitate discoveries, stopping tree growth, and then pruning the tree to offer statistical protection so that the resulting tree from the training set does not have a biased form.\textsuperscript{369} Nodes are linked hierarchically, with several child nodes branching from a common parent node. A node with no children nodes is called a leaf. Typically, in each node, a test using a single descriptor is made. The decision tree asks a binary question at each node, yes or no, concerning a specific feature, and partitions the data into 2 subsets resulting in 2 child nodes while minimizing the mean “impurity” of the two partitions at each node in the tree. Based on the result of the test, the algorithm is directed to one of the child nodes branching from the parent. In the child node, another test is performed and further traversal of the tree towards leaves is carried out. The final decision is based on the activity class associated with the leaf.

![Decision Tree Diagram]

\textbf{Figure 2.4} Classification using decision tree based on three molecular descriptors
Thus, the whole decision process is based on the series of simple tests, with results guiding the path from the root of the tree to a leaf, as depicted in Figure 2.4.

Due to the reliance on single feature in each node, DT methods usually offer suboptimal error rates compared to other non-linear methods. The tree effectively combines the training process with descriptor selection which limits the complexity of the model to be analyzed. Since several leaves in the tree may correspond to a single activity class, they allow for inspection of different paths leading to the same activity.

2.3.4.4 Support vector machine (SVM)

In support vector regression (SVR)\(^{370}\), the basic idea is to map the data \(X\) into a higher-dimensional feature space via a nonlinear mapping and then to do linear regression in this space. Therefore, regression approximation addresses the problem of estimating a function based on a given data set.

2.3.4.5 Other techniques in 3D QSAR modeling

2.3.4.5.1 Alignment-independent 3D QSAR

The group of methods that do not require molecule alignment prior to the calculation of descriptors is strongly independent to molecule rotation and translation in space. Thus, no superposition of compounds is required.

2.3.4.5.2 Comparative molecular moment analysis (CoMMA)

The method is based on 3D geometrical representation of molecule which calculates different molecular moments with respect to the center of mass, center of charge and center of dipole of molecule.\(^{371}\) The moments relate to center of the mass and center of the dipole. The CoMMA descriptors include principal moments of inertia, magnitudes of dipole moment and principal quadrupole moment. CoMMA overcomes the problem due to the molecular alignment by calculating molecular descriptors based on 3D geometry without common orientation frame.
2.3.4.5.3 Alignment-dependent 3D QSAR

The group of methods that require molecule alignment prior to the calculation of descriptors is strongly dependent on the information on the receptor-ligand interactions. If in case where such data is available, the alignment can be guided by studying the receptor-ligand complexes. These methods mainly rely e.g. on atom-atom or substructure-substructure mapping.

2.3.4.5.4 Self organizing molecular field analysis (SOMFA)

SOMFA utilizes a self-centered activity, i.e., dividing the molecule set into actives (+) and inactives (-), and a grid probe process that penetrates the overlaid molecules, the resulting steric and electrostatic potentials are mapped onto the grid points and are correlated with activity using linear regression.\textsuperscript{372}

2.3.4.5.5 Comparative molecular field analysis (CoMFA)

The hypothesis of CoMFA\textsuperscript{373} is that the differences among chemicals in a target property, such as binding affinity, are often correlated with the differences in the non covalent fields surrounding their structure. These fields, i.e. the steric (Lennard–Jones) and electrostatic (Coulombic) fields, are calculated at regular intervals throughout a defined region. For this the aligned molecule is placed in a 3D grid. In each point of the grid lattice a probe atom with unit charge is placed and the potentials (Coulomb and Lennard-Jones) of the energy fields are computed. Then, they serve as descriptors in further analysis, typically using partial least squares regression. This analysis allows for identifying structure regions positively and negatively related to the activity in question.

Despite of offering many advantages over classical QSAR and good performance in various practicable applications, CoMFA has several limitations as given below:

1. Uncertainty in selection of compounds and variables.
2. Fragmented contour maps with variable selection procedures.
3. Hydrophobicity not well-quantified.
4. Cut-off limits used.
5. Low signal to noise ratio due to many useless field variables.
6. Imperfections in potential energy functions.
7. Various practicable problems with PLS.
8. Applicable only to \textit{in vitro} data.

2.3.4.5.6 \textbf{Comparative molecular similarity indices analysis (CoMSIA)}

CoMSIA was developed at BASF Ludwigshafen, Germany by Klebe \textit{et al.} \cite{374} CoMSIA is similar to CoMFA in the aspect of atom probing throughout the regular grid lattice in which the molecules are immersed. The similarity between probe atom and the analyzed molecule are calculated. This technique is most commonly used in drug discovery to find the common features that are important in binding to the relevant biological receptor. In CoMSIA, both steric and electrostatic features, hydrogen bond donor, hydrogen bond acceptor and hydrophobic fields are considered. The equation used to calculate the similarity indices is as given in equation (10).

\begin{equation}
A^q_{F,K,(i)} = -\sum W_{probe,k} W_{ik} e^{-\alpha r_{iq}}
\end{equation}

where \( A \) is the similarity index at grid point \( q \), summed over all atoms, \( i \), of the molecule \( j \). \( W_{probe,k} \) is the probe atom with a radius 1 Å charge +1, hydrophobicity +1, hydrogen bond donating +1, hydrogen bond accepting +1. \( W_{ik} \) is the actual value of the physicochemical property, \( k \), of atom \( i \). \( r_{iq} \) is the mutual distance between the probe atom at grid point \( q \) and atom \( i \) of the test molecule. \( \alpha \) is the attenuation factor. A larger value results in a steeper Gaussian function and a strong attenuation of the distance-dependent effects of molecular similarity.

2.3.4.5.7 \textbf{Hypothetical active site lattice (HASL)}

Inverse grid based methodology developed in 1986-88, allows mathematical construction of a hypothetical active site lattice which can model enzyme-inhibitor interaction and provides predictive structure-activity relationship for a set of competitive inhibitors. Computer-assisted molecule to molecule match makes the use of multidimensional representation of inhibitor molecules. The result of such
matching are used to construct a hypothetical active site by means of a lattice of points which is capable of modelling enzyme-inhibitor interactions.\textsuperscript{375}

2.3.4.6 Recent advances in QSAR

2.3.4.6.1 3D Pharmacophore modeling

Pharmacophore modeling is powerful method to identify new potential drugs. Pharmacophore models are hypothesized on the 3D arrangement of structural properties such as hydrogen bond donor and acceptor properties, hydrophobic groups and aromatic rings of compounds that bind to the biological target.\textsuperscript{376} The pharmacophore concept assumes that structurally diverse molecules bind to their receptor site in a similar way, with their pharmacophoric elements interacting with the same functional groups of the receptor.\textsuperscript{377} Pharmacophore generation utilizes Hiphop and Hypogen hypothesis (Discovery Studio 2.1, Accelrys Inc. USA). Hiphop is common feature-based pharmacophore modeling and Hypogen is activity-based pharmacophore modeling, it is used in virtual screening of databases in lead identification/lead optimization phase.

2.3.4.6.2 4D QSAR

4D QSAR can be interpreted as a feasible extension of 3D QSAR to address uncertainties during the alignment process. 4D QSAR concept approaches the alignment issue by incorporating molecular and spatial variety by representing each molecule in different conformation, orientations, tautomers, stereoisomers or protonation states. Two different class of 4D concept have been developed: one class of QSAR makes use of large ensemble of structurally similar conformations. In the second QSAR approach a small set of diverse ligand configurations represents independent alternatives for the QSAR modelling.\textsuperscript{378}

2.3.4.6.3 5D QSAR (induced fit modeling)

Flexible-protein docking involves allowance for the flexibility of the binding pocket, while docking a small molecule-ligand is now-a-days considered as state of the art. The adaptation of this philosophy to the area of QSAR is still in its infancy. To
simulate induced fit in an explicit manner, simulation of a topological adaptation of the model of the binding site surface to the individual ligand molecule has been devised where the surface of the binding site model can slightly shrink or expand depending on the size and topology of the ligand binding to it. As the identification of the correct magnitude and mechanism of induced fit is not possible in absence of the structure of the true target protein, different induced fit protocols (e.g. magnitude dependent on steric, electrostatic, hydrogen bond or lipophilic potential) are presented as alternative scenarios (5D QSAR) to the QSAR.\textsuperscript{378}

2.3.4.6.4 6D QSAR

6D QSAR is an extension of 5D QSAR. It allows for the evaluation of multiple representation of different solvation models.
2.3.4.7 Genetic function approximation (GFA)

GFA algorithm offers a new approach to build structure-activity models. It automates the search for QSAR models by combining genetic algorithm and statistical modelling tools. GFA algorithm was developed by David Rogers. GFA is genetic based method which combines Holland’s genetic algorithm and Friedman’s multivariate adaptive regression splines (MARS). Application of GFA algorithm may allow the construction of higher quality predictive models and make available additional information not provided by standard regression techniques.379-380

Genetic algorithm was derived from analogy with the evolution of DNA. In this analogy, individuals are represented by a one dimensional string of bits. An initial population is created of individuals, usually with random initial bits. A fitness function is used to estimate the quality of individuals, so that the ‘best’ individuals receive the best fitness score. Individuals with the best fitness score are more likely to be chosen for mating and to propagate their genetic material to offspring through the crossover operation in which piece of genetic material is taken from each parents and recombined to create child. The typical crossover operation is mentioned in Figure 2.5. After many mating steps, the average fitness of individuals in the population increases as ‘good’ combinations of genes are discovered and spread through the population.

\[
\begin{align*}
\{\text{LOGP; ATCH4; DIPV}_Y; \text{MOFI}_Z; \text{LOGP}^2\} & \quad \text{First parent} \\
\{\text{MOFI}_Y; \text{MOFI}_Z; DIPV}_Y; \text{ESDL3}\} & \quad \text{Second parent} \\
\{\text{LOGP; ATCH4; DIPV}_Y; \text{ESDL3}\} & \quad \text{New model}
\end{align*}
\]

**Figure 2.5** The crossover operation, generation of new model from parents

The GFA algorithm accomplishes the breeding of the best equations and elimination of the poorer equation by genetic algorithm. The genetic algorithm cuts and separates individual equations and recombines the fragments to form new equations. The genetic algorithm utilizes Freidman’s Lack-of-fit (LOF), as defined by the equation
(11), parameter to select equations for breeding and survival. Use of the LOF measure drives the population toward more parsimonious, simple model and avoid over-fitting of the data. As generations of equations are bred and mutated, the population evolves to a series of ever-increasing quality of equations.

\[ \text{LOF} = \frac{\text{LSE}}{\{1 - \frac{(c + dp)}{m}\}^2} \]  

(11)

Where LSE is least square error, c is number of basis function in the model, d is smoothing parameter, p is number of descriptors and m is number of observations in the training set. The smoothing parameter d controls the scoring bias between equations of different sizes.

The GFA method does not require the assumption that the relationship between independent and dependent variables operates over the entire variable range. GFA circumvents the need for this assumption by using spline-based terms for the construction of its regression equation.

The GFA algorithm approach has number of important advantages over the other techniques: it build multiple models rather than a single model. It automatically selects which basis functions are to be used and determines appropriate number of basis functions by testing full-size model rather than incrementally building them; it is better at discovering combination of basis functions that take advantage of correlations between descriptor; it incorporates the LOF error measure that resist over-fitting. GFA can builds model using not only linear polynomial, but also higher-order polynomials, spline and gaussians.
Validation of generated QSAR models
Statistical significance of relationship between molecular descriptors and biological activity is analyzed by three independent analyses: variance inflation factor (VIF); cross validation or internal validation with training set; and external validation with test set.

Multicollinearity of descriptors
For a statistically significant model, it is necessary that the descriptors evolved in the equation should not be inter-correlated with each other. Variance inflation factor (VIF) analysis is performed in order to check the inter-correlation of descriptors, which is calculated from equation (12). A VIF value larger than 10 indicates that the information of descriptors can be hidden by correlation of descriptors and multicollinearity.

\[
VIF = \frac{1}{1 - r^2} \quad (12)
\]

Where, ‘r²’ is the multiple correlation coefficient of one descriptor’s effect regressed on the remaining molecular descriptors. VIF analysis for the selected descriptors should preferably be <5.

Cross validation or Internal validation of the QSAR model
QSARs are developed to make predictions, not merely for the ability to reproduce the results in the training set. Cross-validation techniques by leave-one out/leave-some out/leave-many out employing the training set molecules ensure the robustness or predictiveness of QSAR models. In this test, each sample is systematically removed from the data set, and new regression coefficients are generated for a given model. This newly-regressed model is used to predict the removed sample. This procedure is performed on each sample in sequence. Correlation-coefficient of cross validation greater than 0.7 indicates robustness and significance of the model as interpretative model and it is calculated by equation (13).

\[
r^2_{cv} = 1 \left[ \frac{\sum(Y_{obs} - Y_{pred})^2}{\sum(Y_{obs} - Y_{mean})^2} \right] \quad (13)
\]
Where, ‘Y_{obs}’ is the observed activity of training set compound, ‘Y_{pred}’ is the predicted activity of training set compound and ‘Y_{mean}’ is the average activity of all molecules in the training set.

**External validation of QSAR model**

Predictive power of the models can also be estimated using an external dataset by examination of the statistical parameters such as $r^2_{pred}$ and $r^2_m$. The predictive ability of each analysis is determined from a set of compounds those are not included in the training set. The predictive $r^2$ ($r^2_{pred}$) value is based on molecules of the test set only and is defined as equation (14).\(^{383}\)

$$r^2_{pred} = \frac{SD - PRESS}{SD}$$  \hspace{1cm} (14)

Where, ‘SD’ is the sum of the squared deviations between the biological activities of the test set and mean activity of the training set molecules; and ‘PRESS’ is the sum of the squared deviation between predicted and actual activity values for every molecule in the test set. The value of $r^2_{pred}$ between 0.5 and 1.0 indicates good predictive power of the model while that between 0.5 and 0.0 indicates poor predictive power of the model. For better understanding the external predictability of models, modified $r^2$ ($r^2_m$) can be determined by equation (15).

$$r^2_m = r^2[1-\sqrt{(r^2-r_0^2)}]$$  \hspace{1cm} (15)

Where, ‘$r^2$’ is the squared correlation coefficient between observed and predicted values and ‘$r_0^2$’ is the squared correlation coefficient between observed and predicted values without intercept.