Chapter 2
Literature review

2. Literature review

2.1. Cholinesterase

2.1.1. Acetylcholinesterase

AChE (EC 3.1.1.7) is a serine hydrolase involved in the metabolism of ACh. It terminates the action of ACh at the neuronal synapse and neuromuscular junction (NMJ). AChE has a serine residue located deep down the end of the gorge in the active site, which acts as a nucleophile and is responsible for the initiation of the hydrolysis by attacking the carbonyl group of ACh. The oxyanion hole located adjacent to the catalytic site helps in the stabilisation of tetrahedral intermediate. The histidine residue acts like acid and base, and is involved in the proton transfer, leading to the formation of choline [113]. The aromatic anionic site residues are responsible for binding with quaternary nitrogen of choline on the active site and provide stability of protein-ACh complex through stacking interactions [114]. The acyl binding pocket, located near the active site, is crucial for maintaining the substrate specificity of the enzyme. The mutation of phenylalanine of acyl binding pocket with alanine leads to a 130-fold increase in enzyme turnover of butyrylthiocholine [115]. The structure of AChE revealed a group of mixed β-sheets surrounded by the α-helices, rendering similarity with other hydrolases such as serine carboxypeptidase-II and haloalkane dehalogenase [116]. The amino acid residues of the peripheral anionic site (PAS) provide a guidance mechanism to facilitate the entry and migration of ACh molecules in the enzyme (**Figure 2.1**). The PAS and enzyme tunnel is formed of aromatic and some anionic amino acid residues [117].

AChE is widely distributed in the human body, with its prominent presence observed in the cholinergic, non-cholinergic nerve fibres and myoneural junction. It is also present on the glial cells and cell membrane of the red blood cells. Conventionally, it is involved in the ACh metabolism into acetate and choline. It also facilitates the breakdown of leu-enkephaline, met-enkephaline, substance-P and other neuropeptides [118, 119]. The

presence of AChE in the embryonic heart signifies its role in morphogenesis and pulse conduction [120, 121]. Various forms of AChE are globular monomer (G1), dimer (G2) and tetramer (G4), tailed tetramer (A4), double-tailed tetramer (A8) and triple tailed tetramer (A12) [122].

2.1.2. Butyrylcholinesterase

BChE (EC: 3.1.1.8), also referred to as plasma cholinesterase and pseudocholinesterase, is a ChE enzyme that is involved in numerous hydrolytic processes. The gene located on the third chromosome (3q26) is responsible for encoding the BChE enzyme and displays poly-allelism [123]. The locus present on the second chromosome encodes a subunit X which is covalently bound to the tetrameric form of the enzyme [124, 125]. The BChE monomer is formed of 574 amino acid residues with a molecular weight of ~65.1 KDa. A complex glycan with nine asparagine residues appended with the enzyme led to an increase in molecular weight to ~ 85 KDa [126]. The polymorphic presence of the enzyme in the form of monomer (C1), two monomers linked via disulfide bonds to form a dimer (C2), trimer (C3) and tetramer (C4) are evident [127]. It is present significantly in plasma, leg muscles, small intestine, liver, skin and lungs. BChE is present, in small amounts, in the heart, spleen, stomach, and thyroid. It is also present in the cerebral cortex (3 mg) and in smaller quantities in the cerebellum and medulla oblongata in the brain [128].

BChE has a group of β-sheets encircled by the α-helices, similar to AChE. The enzymatic cavity is bowl shape with a larger volume of 500 $Å^3$, contrary to the narrow channel of AChE [129]. The catalytic site of the enzyme has three primary residues viz. Ser198, Glu325 and His438 form a catalytic triad located 20 Å deep in the enzyme cavity. The histidine residue, located close to the serine, is responsible for accepting the proton from the hydroxyl group of serine and initiates a nucleophilic attack on the acyl group of substrates. The oxyanion hole formed of Gly116, Gly117 and Ala119 residues is located near the catalytic site and is responsible for the stabilisation of substrate in the cavity

mediated through hydrogen bonding. The NH groups create a dipole moment in the cavity. The anionic site (AS), formed from Trp82, is the choline-binding site and is present opposite to the catalytic site. It is the crucial residue that forms a π-cation interaction with the positive charge present on the choline. The adjacent Met81 residue provides extra stability to tryptophan residue through π-sulfur interaction and to Tyr440 through hydrogen bonds.

The acyl binding pocket of the enzyme is larger in volume as compared to AChE and is lined by Val286 and Val288 residues in contrast to phenylalanine. Hence, it could accommodate the larger acyl groups besides the acetyl group. Asp70 and Tyr332 residues constitute the PAS and lack three aromatic residues present in AChE (**Figure 2.1**). The cavity of BChE is lined with fewer aromatic residues in comparison to AChE [130]. Tyr332 is the crucial residue that initially binds with cationic ligands of the enzyme [131]. BChE is responsible for the inactivation/metabolism of substances such as eserine, cocaine, choline-based esters, endogenous ACh and ghrelin in the plasma. It also protects the body against nerve gas, organophosphorus pesticides and neurotoxins such as anatoxin-a(s). The exposure or toxicity of the chemicals could be also be determined after a long time of about 12 days due to the longer half-life of the enzyme. These agents form a covalent bond with serine residue of the active site and remain in the blood circulation [132].

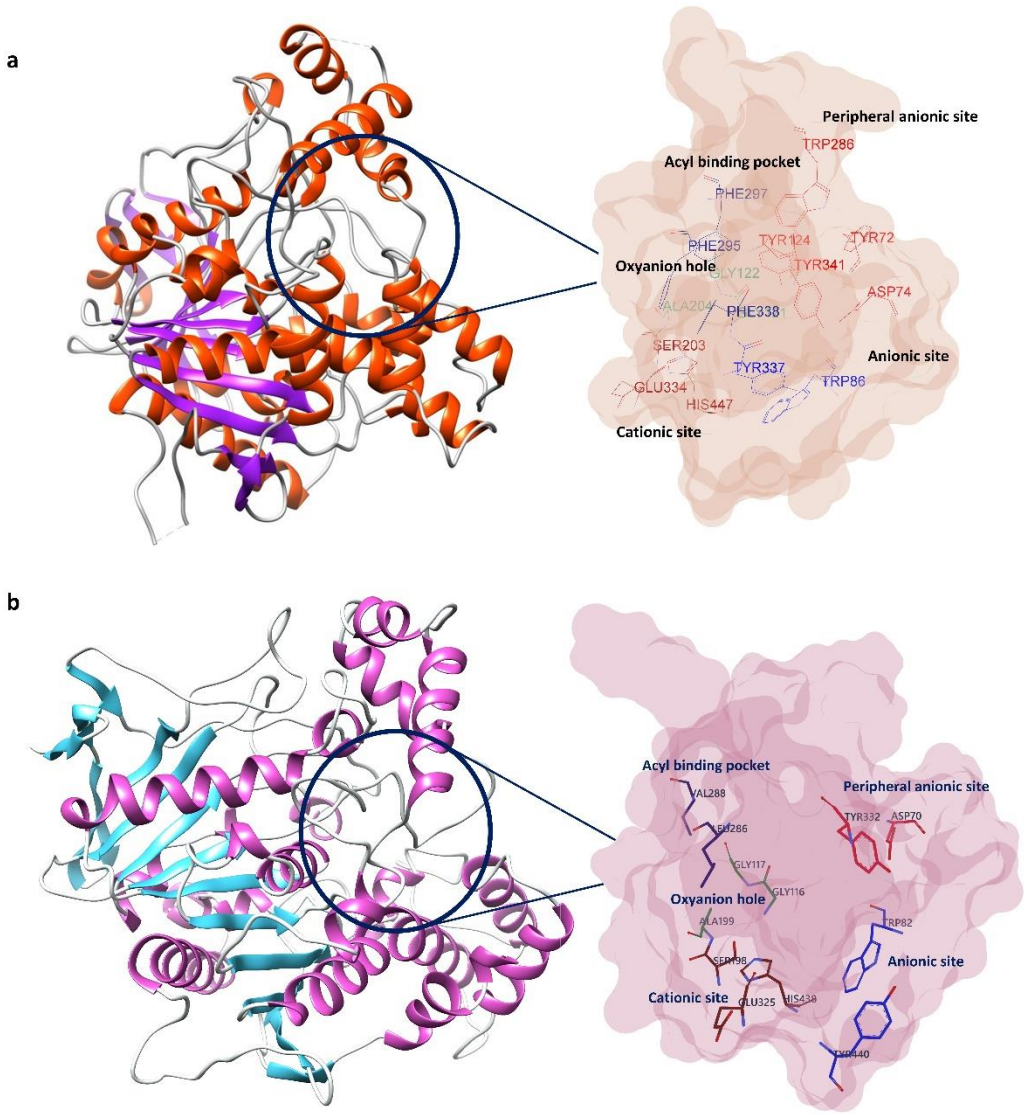


Figure 2.1 Structure of (a) AChE and (b) BChE

2.1.3. Role of cholinesterase in Alzheimer's disease

Cholinergic neuronal loss in the brain is one of the phenomena observed in AD and has a correlation with memory and cognitive decline. In AD, AChE is responsible for the hydrolysis of ACh and subsequent decrease in its level in the hippocampus, which is the centre of learning and memory. The inhibition of AChE provides therapeutic benefits by the improvement in memory and cognition in AD. The administration of scopolamine in the young population displayed memory deficit indicating the role of the cholinergic pathway in maintaining memory and cognition [51]. The action of ACh in the synaptic cleft is also terminated by BChE, besides AChE. In a healthy brain, this process is

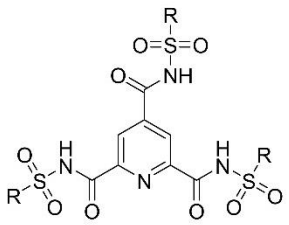
dominated by AChE, while the role of BChE is deemed to be nominal. BChE is present in the amygdala, neocortex and hippocampus, where it is associated with glial, vascular and neuronal cells [133]. A study on AChE knockout mice displayed the absence of cholinergic abnormality or deficit in the brain, as its function was performed by BChE [134]. Another study indicated that the reduction in AChE levels due to cholinergic neuronal loss resulted in an increase in BChE to AChE ratio by 0.6 – 11 in AD patients [135]. A study of a selective BChE inhibitor on AD patients indicated that withdrawal of ethopropazine led to cognitive impairment, while its re-administration resulted in improvement in cognition [136]. This established that BChE has excellent therapeutic potential in AD as a promising target, especially in severe AD in the later stage. Further, the selective BChE inhibitors may be beneficial due to the absence of classical cholinergic side effects observed with the currently available AChE inhibitors.

2.2. Sulfonamides

Sulfonamides are bioactive chemicals with a wide range of biological activities, including antibacterial, anticancer, anti-carbon anhydrase, anti-diabetic, anti-inflammatory, antithyroid, hypoglycaemic, proteases inhibition and diuretic properties [137-140].

2.2.1. Sulfonamides as cholinesterase inhibitors

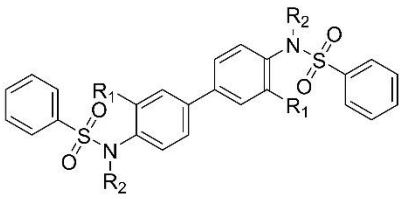
Riaz *et al.* synthesised *pyridine sulfonamide* derivatives for the treatment of diabetes and AD. The *pyridine-2,4,6-tricarbohydrazide* was used as the starting material for the synthesis of various sulfonamide derivatives in the presence of an aqueous *sodium carbonate* solution. Compounds **I** and **II**, i.e., *methyl* and *phenyl* derivatives, respectively, displayed the AChE and BChE inhibition. The aliphatic substitution (**I**) at the sulfonyl side-chain resulted in better AChE inhibition than the aromatic derivatives (**II**). On the other hand, compounds **I** and **II** did not display any trend in BChE inhibition [141].



Pyridine-2,4,6-tricarbohydrazide

Compound	R	AChE IC_{50} (μM)	BChE IC_{50} (μM)
I	CH_3	10.8 ± 0.9	44.8 ± 0.9
II	C_6H_5	50.2 ± 0.8	43.8 ± 0.9

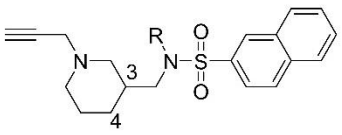
IC_{50} expressed in Mean ± SEM



Biphenyl bis-sulfonamide

Compound	R_1	R_2	AChE IC_{50} (μM)	BChE IC_{50} (μM)
III	H	$CH_2(CH_2)_{14}CH_3$	4.63 ± 0.05	7.74 ± 0.07
IV	CH_3	$CH_2(CH_2)_{14}CH_3$	2.45 ± 0.02	24.74 ± 0.19
V	CH_3	$CH_2(C_6H_5)$	2.27 ± 0.01	31.31 ± 0.05

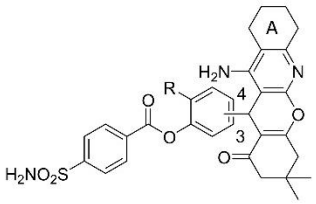
IC_{50} expressed in Mean ± SEM



N-propargyl piperidines

Compound	R	BChE IC_{50} (μM)
VI	$(CH_2)_2OCH_3$	0.137 ± 0.003
VII	$(CH_2)_3OCH_3$	0.127± 0.006

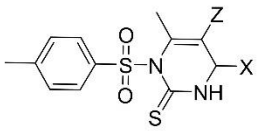
IC_{50} expressed in Mean ± SEM



Tacrine

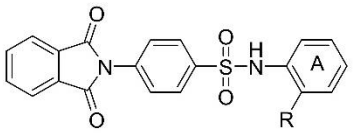
Compound	R	Substitution	AChE IC_{50} (μM)	BChE IC_{50} (μM)
VIII	H	4	0.01 ± 0.001	3.41 ± 0.332
IX	NO_2	3	0.02 ± 0.003	2.43 ± 0.270

IC_{50} expressed in Mean ± SEM



Tetrahydro-pyrimidinethiones

Compound	X	Z	AChE IC_{50} (μM)	BChE IC_{50} (μM)
X	$CH_3C_6H_4$	$COOCH_2CH_2CH_3$	0.76	5.87
XI	C_6H_5	$COOC_2H_4OCOCH_2CH_3$	0.95	5.25



4-phthalimidobenzene sulfonamide

Compound	R	AChE IC_{50} (μM)	BChE IC_{50} (μM)
XII	OCH_3	6.79 ± 0.41	44.24 ± 13.38
XIII	$OCH_2(CH_3)_2$	8.19 ± 0.13	13.41 ± 0.62

IC_{50} expressed in Mean ± SEM

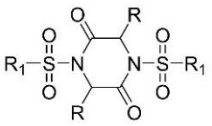
Figure 2.2 Chemical structure of BChE inhibitors

Mutahir *et al.* synthesised novel *biphenyl bis-sulfonamide* derivatives as dual cholinesterase inhibitors. The condensation of the *benzidines* with *benzene sulfonyl chloride* was carried out in the presence of *pyridine* as a base. The alkylation or acylation was carried using *sodium hydride* in *dimethylformamide* with the replacement of acidic

proton. Compounds **III** and **IV,** *hexadecyl* derivatives, and **V,** a *benzyl* derivative, displayed good inhibition of AChE and BChE. It indicated that the substitution of the R1 group with the alkyl chain resulted in reduced BChE inhibition but increased AChE inhibition potency (**Figure 2.2**) [142].

Košak *et al.* designed *piperidine*-based BChE inhibitors along with MAO-B inhibitory activity. The nitrogen of the *piperidine* ring was an essential feature for human BChE inhibition, as reported in their previous work. At the same time, the introduction of the *propargyl* group might help in MAO-B inhibition, as in selegiline and rasagiline. The *nipecotamide* was used as a starting compound and various *naphthyl amides* and *sulfonamides* were synthesised to obtain multifunctional anti-AD agents. Compounds **VI** and **VII** were *naphthyl sulfonamide* derivatives of *piperidine* with selective human BChE inhibition. The compounds displayed poor inhibition against mouse AChE and human MAO enzymes. Study established that the presence of the *sulfonamide* group, instead of the carboxamide, showed better BChE inhibition. Another interesting observation was that 1, 3-substituted *piperidine* was a better scaffold than 1, 4-substitution of *piperidine* for BChE inhibition [143]. Ulus *et al.* synthesised various sulfonamide derivatives of *aldehydes*, *chromenes* and *tacrines*. Initially, substituted benzaldehydes were condensed with the *4-sulfamoylbenzoic acid*, which was followed by condensation into chromenes by reaction with *malononitrile* and *dimendone* in the presence of K_2CO_3 and microwave irradiation (6 min. and 100˚ C). Finally, cyclic ketones (*pentanone*, *hexanone* and *heptanone*) were reacted with *chromenes* using $AlCl_3$ under microwave irradiation (10 min. and 120˚ C) to obtain *tacrine sulfonamides*. It was observed that the six-membered ring showed highest AChE inhibition, and a significant reduction in inhibition was observed with the five and seven-membered skeletons at position A. Compounds **VIII** and **IX** were the most potent compounds with significant AChE and BChE inhibition. It

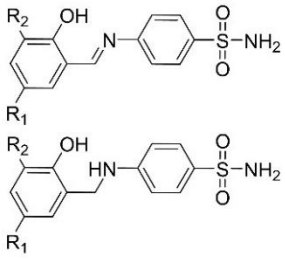
was observed that the *tacrine* derivatives showed better inhibition of AChE than BChE [144].



Piperazine-2,5-dione

Compound	R	R_1	AChE IC_{50} (nM)	BChE IC_{50} (nM)
XIV	H	4-Cl	61.62 ± 0.094	180.21 ± 0.230
XV	C_6H_5	4-Cl	28.65 ± 0.029	160.58 ± 0.082
XVI	$C_6H_5CH_2$	4-Cl	32.45 ± 0.044	157.95 ± 0.264

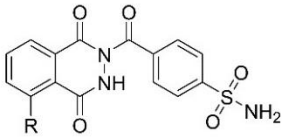
IC_{50} expressed in Mean ± SEM



Benzene sulfonamide

Compound	R_1	R_2	AChE Ki (μM)
XVII	Br	Br	2.54 ± 0.22
XVIII	Cl	Cl	5.19 ± 0.44
XIX	Br	Br	5.12 ± 0.10
XX	Cl	Cl	4.50 ± 0.31

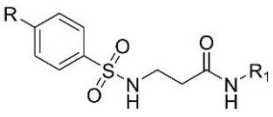
IC_{50} expressed in Mean ± SEM



N-substituted phthalazine sulfonamide

Compound	R	AChE IC_{50} (nM)
XXI	Cl	48.69 ± 0.57
XXII	NO_2	57.96 ± 0.56

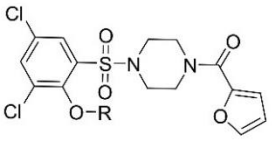
IC_{50} expressed in Mean ± SEM



Bi-phenyl sulfonamide

Compound	R	R_1	AChE IC_{50} (nM)	BChE IC_{50} (μM)
XXIII	4-Cl-C_6H_4	8-quinolinyl	18.24 ± 1.61	4.28 ± 0.14
XXIV	4-F-C_6H_4	6-quinolinyl	19.18 ± 0.19	3.58 ± 0.87

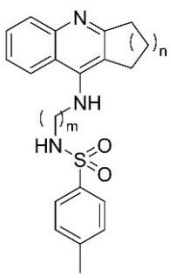
IC_{50} expressed in Mean ± SEM



N-substituted phthalazine sulfonamide

Compound	R	BChE IC_{50} (nM)
XXV	CH_2CHCH_2	9.51 ± 0.06
XXVI	o-$CH_3C_6H_4CH_2$	15.43 ± 0.09

IC_{50} expressed in Mean ± SEM



Conjugates of 4-amino-2,3-polymethylene-quinolines and p-tolylsulfonamide

Compound	n	m	AChE IC_{50} (μM)	BChE IC_{50} (μM)
XXVII	2	4	0.668 ± 0.17	0.061 ± 0.0003
XXVIII	2	5	0.131 ± 0.01	0.068 ± 0.0014
XXIX	3	3	2.76 ± 0.04	0.0431 ± 0.0011

IC_{50} expressed in Mean ± SEM

Figure 2.3 Chemical structure of BChE inhibitors.

Taslimi *et al.* developed hetero-aryl *sulfonamides* based ChE inhibitors from the reaction between various *tetrahydro-pyrimidine-thiones* and *toluyl sulfonyl chloride* in the

presence of *triethylamine* in ethanol. The heteroaryl *sulfonamides* displayed good inhibition against AChE with the IC_{50} reported in the nanomolar ranges. Further, these compounds have K_i value in the nanomolar range, indicating strong binding with AChE as compared to *tacrine*. However, the compounds displayed much higher IC_{50} values against the BChE. Compounds **X** and **XI** produced IC_{50} of 760 and 950 nM against AChE [145]. Soyer *et al.* developed a series of *4-phthalimido-benzenesulfonamide* derivatives as ChE inhibitors. Initially, *N-phenylphthalimide* was synthesised from *phthalic anhydride* and *aniline*, which was reacted with *chlorosulfonic acid* to yield *4-phthalimido-benzene sulfonyl chloride*. Further, *anilines* were substituted to form *sulfonamides*. Compounds **XII** and **XIII** displayed better AChE and BChE inhibition among all the synthesised compounds (**Figure 2.2**). It was observed that the substitution at the *ortho* position of the *aniline* ring (A) of the *sulfonamide* group displayed better AChE and BChE inhibition than the *para*-substitution. Substituting the *phenyl* with a heterocyclic ring did not improve the ChE inhibition [146]. Kumar *et al.* designed ChE and MMP dual inhibitors using pharmacophore and data mining techniques. The *3,6-Diphenyl-1,4-bis(phenylsulfonyl)piperazine-2,5-dione* derivatives were identified as novel multifunctional anti-AD agents. Amino acids viz. *glycine*, *phenylglycine* and *phenylalanine* were reacted with substituted aromatic *sulfonyl chlorides*. Further, the substituted *sulfonamide* was condensed in *ethylene glycol* at 120 – 150˚ C to obtain substituted *piperazine-2,5-dione*. Compounds **XIV**, **XV** and **XVI**, the *para-chloro* derivatives, were the most active and displayed inhibition of both AChE and BChE in the nanomolar concentration [147]. Isik *et al.* identified *4-(benzylideneamino)* and *4-(benzylamino)-benzenesulfonamide* derivatives as AChE inhibitors. Halogen-substituted aldehydes were treated with *sulfanilamide* to obtain *benzylidene* derivatives, which were further reduced in the presence of $NaBH_4$ to obtain *benzyl* derivatives. The *dibromo* compounds **XVII** and **XIX** were found to be most active with a mixed type

inhibition against AChE. The *dichloro* derivatives **XVIII** and **XX** were also found to be active [148]. Turkes *et al.* synthesised *N-substituted phthalazine sulfonamide* compounds for AChE and carbonic anhydrase dual inhibition. *4-Sulfonylamide* ester obtained from *4-Sulfamoylbenzoic acid* was converted to hydrazide in ethanol. The obtained hydrazide was reacted with substituted *phthalic anhydrides* to obtain *phthalazine sulfonamide* derivatives. The *chloro* (**XXI**) and *nitro* (**XXII**) derivatives displayed better IC_{50} than others [149]. Rayala *et al.* synthesised multifunctional hybrid *sulfonamides* as anti-AD agents. It involved the synthesis of *sulfonamides* of *glycine* and *β-alanine* using substituted *biphenyl sulfonyl chlorides*. Further, heterocyclic amines were used for the synthesis of amides. Compounds **XXIII** and **XXIV** displayed excellent inhibition for AChE and BChE (**Figure 2.3**). These compounds exhibited metal-chelating properties and anti-Aβ aggregation [150]. Hassan *et al.* synthesised sulfonamides of 2-*furoyl piperazine* as BChE inhibitors for the treatment of AD. *2-furyl(1-piperazinyl)methanone* was reacted with *3,5-dichloro-2-hydroxybenzenesulfonyl chloride* in basic pH condition. The *hydroxyl* present on the *phenyl* ring connected through the *sulfonyl* group was substituted with various aliphatic and aromatic side chains in the presence of LiH. The substitution of the R group with aromatic or allylic groups resulted in better BChE inhibition. Compound **XXV**, a *propenyl* derivative, displayed better inhibition among all the compounds, followed by compound **XXVI**, an *ortho-chloro benzyl* derivative [151]. Makhaeva *et al.* synthesised conjugates of *4-amino-2,3-polymethylene-quinolines* and *p-tosyl sulfonamide* as ChE inhibitors for AD. The conjugates were formed by condensation of *anthranilic acid* with cyclic ketones of varying ring sizes, followed by substituting variable *diamine* to yield *aminoquinolines*. The *aminoquinolines* were further reacted with *tosyl chloride* in the presence of *triethylamine*. The most active against AChE was compound **XXVII**, which had a *tacrine* ring connected with the *toluene sulfonamide* group with a linker of 5 carbons. Compound **XXIX** was the most active compound,

having a seven carbon-containing *tacrine* ring connected through a three-carbon linker to the *toluene sulfonamide* group [152].

2.3. Scoring function

Molecular docking helps in the prediction of ligand pose that suitably and stably binds in the active site of the protein. SF is the mathematical function that estimates the binding affinity between protein and ligand.

2.3.1. Classification of scoring function

SF is classified into four types:

2.3.1.1. Force-field based scoring function

Force field-based SF uses bond stretching, bending, torsional forces along with physical atomic interactions such as van der Waals and electrostatic interactions. These parameters are obtained from experimental data as well as *ab initio* quantum mechanical calculations [153]. The SF of DOCK6 docking program employs Amber based force-field that uses two energy components, i.e., an electrostatic and a Lennard-Jones VDW terms [154, 155].

$$E = \sum_{i} \sum_{j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i \, q_j}{\varepsilon(r_{ij}) r_{ij}} \right)$$

where r_{ij} stands for the distance between protein atom i and ligand atom j, A_{ij} and B_{ij} are the VDW parameters, and q_i and q_j are the atomic charges. Here, the term $\varepsilon(r_{ij})$ is a distance-dependent dielectric factor that treats the effect of solvent as an implicit parameter. The major limitation is that such dielectric factor could not consider the de-solvation effect produced by ligand binding to the protein surface. The other methods, such as free energy perturbation (FEP) and thermodynamic integration (TI) involve the treatment of water explicitly and suitably incorporate the solvent's effect. However, these methods are computationally much expensive [156]. Autodock, DOCK3.5(PB/SA),

DOCK/GBSA(SDOCK), GOLD, SYBYL/D-Score, are some of the examples of force-field based SF.

2.3.1.2. Empirical scoring function

Empirical SF uses a set of weighted energy components to calculate a protein-ligand complex's binding affinity.

$$\Delta G = \sum_i W_i \cdot \Delta G_i$$

where ΔG_i represents various energy components viz. de-solvation, electrostatics, entropy, hydrogen bond, hydrophobicity, VDW energies etc. The modelling of the binding affinities of the protein-ligand complexes with obtained poses helps to determine the coefficients Wi [157, 158]. The empirical SFs are computationally faster in the calculation of binding scores as they calculate simple energy terms compared to force field-based SF. SCORE1, developed by Bohm, uses hydrogen bonds, ionic interactions, the lipophilic protein-ligand contact surface and the number of rotatable bonds terms for calculation. PLIP, FlexX, Glide, SCORE, X-SCORE, LigScore, MedusaScore, SFScore are some of the empirical SFs [159].

2.3.1.3. Knowledge-based scoring function

Energy potentials obtained from inherent structural information in experimentally determined atomic structures are used in knowledge-based scoring systems. It is also known as the statistical-potential based SFs. The inverse Boltzmann relation is used to generate pairwise potentials directly from the occurrence frequency of atom pairs in a database consisting of experimentally determined protein-ligand complexes.

$$w(r) = -k_B T ln[g(r)]$$

$$g(r) = \rho(r)/\rho*(r)$$

where k_B is the Boltzmann constant, ρ(r) is the number density of the protein-ligand atom pair at distance r, T is the absolute temperature (K) and ρ*(r) is the pair density in a reference state, where the interatomic interactions are zero. MScore, BLEEP, ITSCore, DrugScore, SMoG, DFIRE are some of the knowledge-based SFs.

2.3.1.4. Machine learning-based scoring function

ML based SF does not assume a predefined relationship between binding affinity and structural characteristics. The function uses various user-defined features as input to improve the accuracy over the conventional one [160, 161]. Various ML algorithms such as RF, SVM, neural network, decision tree etc. are employed for the development of SFs.

Table 2.1 Selected machine learning-based scoring functions.

S.No.	Scoring function	ML algorithm	Source	Reference
1	BT-Score	Gradient boosted decision tree	-	[162]
2	MT-Net	Deep neural network	-	[162]
3	RI-Score	Random forest	http://weilab.math.msu.edu/RI-Score	[163]
4	RF_Score v3	Random forest	http://ballester.marseille.inserm.fr/rf-score-3.tgz	[164]
5	TopBP	Gradient boosted decision tree, Convolutional neural network	-	[165]
6	DLSCORE	Deep neural network	https://github.com/sirimullalab/dlscore	[166]
7	XGB-Score	XGBoost	https://github.com/HongjianLi/MLSF	[167]

2.4. Machine learning in drug discovery

Drug discovery is a multi-stage, long and tedious process with high chances of failure. The period of 2002 – 2012 reported a failure of 99.6 % in the discovery of anti-Alzheimer's agent. Since, drug discovery relies on various decision-making steps to obtain a hit against a target, the ML and Artificial intelligence(AI) could provide firm

support to the complex decision-making processes and increase the chance of the successful identification of a drug based on the previously acquired data [168].

2.4.1. Application of machine learning in pharmacophore modelling

Hot-Spots-Guided Receptor-Based Pharmacophores (HS-Pharma) is a pharmacophore model development program. It is involved in the development of ML-based filter to reduce the number of the pharmacophore features. In this, various atom-based cavity fingerprints are collected from more than 3500 protein-ligand complexes and classification algorithms are used to identify important ligand binding atoms in the protein cavity. RF performs well to identify important features that are used to develop a structure-based pharmacophore model [169]. Pharmacophore-based interaction fingerprint (Pharm-IF) are pharmacophore descriptors and are employed for training ML models. It is observed that the ML models trained on these pharmacophore fingerprints show better enrichment as compared to PLIF. Pharm-IF fingerprints take into account the distance between pharmacophore features [170]. Deep site is another program that involves a convolutional neural network (CNN) to detect cavities and predict binding affinities. The atomic-based pharmacophoric descriptors were used to train the ML model [171].

2.4.2. Application of machine learning in quantitative structure-activity relationship

QSAR is conventionally performed by regression techniques to identify the crucial molecular features. However, in the past decade, ML based models were reported to develop QSAR. Among the various algorithms, RF is a popular technique that deciphers the feature importance. Some of the application of ML in QSAR includes anticipation of the relationship between chemical structures and toxicities viz. *in vitro* toxicity [172], *in vivo* toxicity [173], hepatotoxicity [174] and mutagenesis [175].

2.4.3. Application of machine learning in molecular docking

Molecular docking involves two phases, i.e., searching the conformation of the ligands and identifying the suitable/correct poses. The conformational sampling is carried out by using various algorithms. The incremental connection uses small fragments to build the ligand in the protein cavity, while in the case of the Monte Carlo simulation, the ligand conformation is gradually built by rotation of bonds and translation of the position of the ligand in the protein. The genetic algorithm applies mutation and crossover over a fraction of the selected population of a generation provided by SF. The other important component of molecular docking is SF. ML based SF has recently gained popularity and provides better feature mapping. RF-score is one of the ML based SF that was developed using RF algorithm employing atom-type pair count as features [176]. ID-Score uses a set of 50 descriptors to describe the protein-ligand contacts that cover nine types of interactions. The SF uses a SVM algorithm was used to train the model over 220 protein-ligand complexes [177]. NNScore 2.0 is a neural network-based SF that performed better that SF of Vina and Autodock over 12 different protein targets [178].

2.4.4. Application of machine learning in molecular dynamics

The implementation of ML in MD helps in the generation of high-precision force fields. The on-the-fly ML method is one of the ways to generate an atomistic force field. The *ab initio* data are picked and added to the training data during the MD calculations. As long as the dynamic configuration is accurately represented in the existing database, no additional QM calculations are added. ML can also be used to accelerate the trajectory generation [179], analyse long simulation data [180], prediction of energies of the ground state [181] and prediction of solute-solvent interaction map [182].