

Table of contents

Table of contents	xi
List of tables	xvi
List of figures	xviii
List of abbreviations	xxii
Preface	xxiv
1. Introduction	2
1.1. Alzheimer's disease	2
1.1.1. Symptoms	2
1.1.2. Stages of Alzheimer's disease	3
1.1.3. Diagnosis of Alzheimer's disease	4
1.1.4. Pathology associated with Alzheimer's disease	5
1.1.5. Management of Alzheimer's disease	10
1.2. Structure-based drug design	11
1.2.1. Molecular docking	12
1.2.2. Homology modelling	13
1.2.3. Molecular dynamics	14
1.3. Ligand-based drug design	15
1.3.1. Pharmacophore modelling	15
1.3.2. Quantitative structure-activity relationship	15
1.4. Machine learning	16
1.4.1. Approaches to machine learning	16
1.4.2. Machine learning algorithms	17
1.4.3. Validation of machine learning models	19
2. Literature review	22
2.1. Cholinesterase	22
2.1.1. Acetylcholinesterase	22
2.1.2. Butyrylcholinesterase	23
2.1.3. Role of cholinesterase in Alzheimer's disease	25
2.2. Sulfonamides	26
2.2.1. Sulfonamides as cholinesterase inhibitors	26
2.3. Scoring function	32
2.3.1. Classification of scoring function	32
2.4. Machine learning in drug discovery	34
2.4.1. Application of machine learning in pharmacophore modelling	35
2.4.2. Application of machine learning in quantitative structure-activity relationship	35
2.4.3. Application of machine learning in molecular docking	36
2.4.4. Application of machine learning in molecular dynamics	36
3. Objective and plan of work	38

3.1. Objective	38
3.2. Plan of work	39
4. Development of sulfonamide based BChE inhibitors as anti-Alzheimer's agents through machine learning	42
4.1. Introduction	42
4.2. Materials and methods	43
4.2.1. Development of machine learning models for prediction of BChE inhibitors	43
4.2.2. Chemistry	44
4.2.3. <i>In vitro</i> cholinesterase inhibitory activity	60
4.2.4. <i>In vitro</i> blood-brain barrier permeation assay	62
4.2.5. Cell viability (MTT Assay)	62
4.2.6. <i>In silico</i> ADMET and molecular property analysis	63
4.2.7. Molecular docking	63
4.2.8. QSAR	63
4.2.9. Molecular dynamics	64
4.2.10. <i>In vivo</i> evaluation of compounds	65
4.3. Results and discussion	69
4.3.1. Development of machine learning models and prediction of BChE inhibitors	69
4.3.2. Chemistry	75
4.3.3. <i>In vitro</i> cholinesterase inhibitory activity	77
4.3.4. Cell Viability (MTT Assay)	82
4.3.5. <i>In vitro</i> blood-brain barrier permeation assay	82
4.3.6. <i>In silico</i> ADMET and molecular property analysis	84
4.3.7. Molecular docking	85
4.3.8. Quantitative structure-activity relationship	89
4.3.9. Molecular dynamics	94
4.3.10. <i>In vivo</i> evaluation of compounds	100
5. Development of sulfonamide based BChE inhibitors as anti-Alzheimer's agents through scaffold hopping	110
5.1. Introduction	110
5.2. Materials and methods	111
5.2.1. Chemistry	111
5.2.2. <i>In vitro</i> ChE inhibitory activity	122
5.2.3. <i>In vitro</i> blood-brain barrier permeation assay	123
5.2.4. Cell viability (MTT Assay)	123
5.2.5. <i>In silico</i> ADMET and molecular property analysis	123
5.2.6. Molecular docking	123
5.2.7. Molecular dynamics	123
5.2.8. <i>In vivo</i> evaluation of compounds	123

5.3. Results and discussion	124
5.3.1. Chemistry	124
5.3.2. <i>In vitro</i> ChE inhibitory activity	127
5.3.3. Cell viability (MTT Assay)	131
5.3.4. <i>In vitro</i> blood-brain barrier permeation assay	131
5.3.5. <i>In silico</i> ADMET and molecular properties analyses	131
5.3.6. Molecular docking	132
5.3.7. Molecular dynamics	135
5.3.8. <i>In vivo</i> evaluation of compounds	140
6. Identification of potential AChE inhibitors through structure-based drug design	150
6.1. Introduction	150
6.2. Material and methods	150
6.2.1. Pharmacophore hypothesis and database screening using Pharmit	150
6.2.2. Molecular property filters	151
6.2.3. Homology modelling and model validation	152
6.2.4. Protein preparation	152
6.2.5. Ligand preparation	153
6.2.6. Grid generation and validation	153
6.2.7. Virtual screening and molecular docking	153
6.2.8. ADMET property	154
6.2.9. MM-GBSA/MM-PBSA assay	154
6.2.10. Alanine scanning	154
6.2.11. Molecular dynamics	155
6.3. Results and discussion	155
6.3.1. Pharmacophore hypothesis and database screening using Pharmit	155
6.3.2. Molecular property filters	156
6.3.3. Homology modelling and model validation	156
6.3.4. Protein preparation	157
6.3.5. Ligand preparation	158
6.3.6. Grid generation and validation	158
6.3.7. Virtual screening and molecular docking	159
6.3.8. ADMET property	162
6.3.9. MM-GBSA/MM-PBSA assay	162
6.3.10. Alanine scanning	163
6.3.11. Molecular dynamics	165
7. Computational binding study of Anvyllic-3288 with α7-nicotinic acetylcholine receptor	172
7.1. Introduction	172
7.2. Material and methods	174

7.2.1. Protein preparation	174
7.2.2. Ligand preparation	174
7.2.3. Molecular docking	174
7.2.4. Alanine scanning	175
7.2.5. Molecular dynamics	175
7.2.6. MM-PBSA	175
7.3. Results and discussion	176
7.3.1. Binding mode in the top pocket	176
7.3.2. Binding mode in vestibule pocket	177
7.3.3. Binding mode in agonist sub-pocket	178
7.3.4. Alanine scanning	180
7.3.5. Molecular dynamics	182
7.3.6. MM-PBSA	185
8. Development of homology model, docking protocol and machine-learning based scoring functions for identification of <i>Electrophorus electricus</i>'s AChE inhibitors	188
8.1. Introduction	188
8.2. Materials and methods	189
8.2.1. Sequence alignment and analysis	189
8.2.2. Active site mapping	189
8.2.3. Homology modelling	189
8.2.4. Homology model refinement and protein preparation	190
8.2.5. Ligand preparation and grid generation	190
8.2.6. Molecular docking and validation of docking protocol and scoring function.	190
8.2.7. Development of the scoring functions for eeAChE	191
8.3. Results and discussion	192
8.3.1. Sequence alignment and analysis	192
8.3.2. Active site mapping	193
8.3.3. Homology modelling	196
8.3.4. Homology model refinement and protein preparation	196
8.3.5. Ligand preparation and grid generation	198
8.3.6. Molecular docking and validation of docking protocol and scoring function.	198
8.3.7. Development and validation of the scoring functions for eeAChE	205
8.3.7.2. Development of scoring function based on binary classification models	207
8.3.7.3. Development of scoring function based on multiclass classification models	210
8.3.7.4. Development of scoring functions based on regression-based models	212
8.3.8. Improved scoring function	215
8.3.9. Applicability domain	215
9. Development of homology model, docking protocol and machine-learning based scoring functions for identification of <i>Equus caballus</i>'s BChE inhibitors	218

9.1. Introduction	218
9.2. Material and methods	218
9.2.1. Homology modelling	218
9.2.2. Protein model refinement and preparation	218
9.2.3. Ligand preparation and grid generation	219
9.2.4. Validation of docking protocol and scoring function	219
9.2.5. Development and validation of the scoring function	220
9.3. Results and discussion	222
9.3.1. Homology modelling	222
9.3.2. Protein model refinement and preparation	222
9.3.3. Ligand preparation and grid generation	226
9.3.4. Validation of docking protocol and scoring function	226
9.3.5. Development and validation of the scoring function	231
10. Development of web application for identification of anti-Alzheimer's ligands	244
10.1. Introduction	244
10.2. Materials and methods	244
10.2.1. Dataset preparation	244
10.2.2. Molecular descriptor calculation	244
10.2.3. Classification of the datasets	245
10.2.4. Feature selection	245
10.2.5. Division of dataset	245
10.2.6. Training of machine learning models	245
10.2.7. Validation of machine learning model	246
10.3. Results and discussion	247
10.3.1. Model development for identification of acetylcholinesterase inhibitors	247
10.3.2. Model development for identification of butyrylcholinesterase inhibitors	248
10.3.3. Model development for identification of β -secretase 1 inhibitors	249
10.3.4. Model development for identification of glycogen synthase kinase 3 β inhibitors	249
10.3.5. Model development for identification of monoamine oxidase B inhibitors	250
10.3.6. Model development for identification of N ₂ B inhibitors of NMDA receptor	250
10.3.7. Alzleads	250
11. Summary and conclusions	258
12. References	262
13. Appendix	278

List of tables

Table 2.1 Selected machine learning-based scoring functions.....	34
Table 4.1 Performance of the final models on training, validation, and test sets.....	73
Table 4.2 Inhibitory potencies of 4-(phenylsulfonamido) benzoic acid derivatives (30 – 55) against cholinesterase enzymes.....	80
Table 4.3 Inhibition of 4-(phenylsulfonamido) benzoic acid derivatives (3, 63 – 69, 73 – 75) against cholinesterase enzymes.....	80
Table 4.4 Permeability data for selected compounds obtained from PAMPA-BBB assay.....	83
Table 4.5 Molecular and selected ADME properties of synthesised compounds.....	85
Table 4.6 Performance of the final QSAR models.....	93
Table 4.7 Energy contributions of protein-ligand complexes obtained from MM-GBSA calculation.....	99
Table 5.1 Inhibition of BChE and AChE by compounds (26 – 47) at a concentration of 50 µM.....	128
Table 5.2 IC ₅₀ of selected compounds against BChE.....	128
Table 5.3 Permeability data for selected compounds from the PAMPA-BBB assay.....	131
Table 5.4 Molecular and selected ADME properties of synthesised compounds.....	132
Table 5.5 Binding energies of the selected compounds.....	135
Table 6.1 Ramachandran parameter for developed homology model using PROCHECK and RAMPAGE.....	156
Table 6.2 Docking results displaying mean binding energy, cluster size and lowest binding energy of ligands against AChE (PDB id 4EY7).....	160
Table 6.3 Predicted pharmacokinetic properties and toxicity of selected compounds.....	162
Table 6.4 Energy contributions of protein-ligand complexes in MM-GBSA assay.....	163
Table 6.5 Energy contributions of protein-ligand complexes in MM-PBSA assay.....	163
Table 7.1 Mean binding energies, cluster size, lowest binding energy and free binding energies of ligands against α7-nAChR (PDB id - 5AFH).....	176
Table 8.1 Validation data of the developed homology models of eeAChE from crystal structures of various organisms.....	197
Table 8.2 Comparison of validation parameters of selected homology models obtained from energy minimisation.....	198
Table 8.3 Validation scores of various machine learning algorithms used to develop binary scoring function.....	209
Table 8.4 Validation scores of various machine learning algorithms used to develop multiclass scoring function.....	211
Table 8.5 Validation scores of various machine learning algorithms used to develop regression-based scoring function.....	214
Table 9.1 Homology models developed for ecBChE from crystal structures of various organisms.....	224
Table 9.2 Homology models developed for ecBChE from crystal structures of various organisms.....	224
Table 9.3 Validation scores of machine learning algorithms employed in the development of binary classification models.....	234
Table 9.4 Validation scores of machine learning algorithms employed in the regression-based models for prediction of IC ₅₀ below 10000 nM.....	237
Table 9.5 Validation scores of machine learning algorithms employed in the regression-based model for prediction of IC ₅₀ above 10000 nM.....	240
Table 10.1 RDKit descriptors used for the development of ML models.....	248
Table 10.2 Performance of the models trained on AChE inhibitor dataset.....	251
Table 10.3 Performance of the models trained on BChE inhibitor dataset.....	252

Table 10.4 Performance of the models trained on BACE1 inhibitor dataset.	253
Table 10.5 Performance of the models trained on GSK-3 β inhibitor dataset.	254
Table 10.6 Performance of the models trained on MAO-B inhibitor dataset.	255
Table 10.7 Performance of the models trained on N2B subunit inhibitor dataset.	256

List of figures

Figure 1.1 Chemical structure for FDA approved drugs for Alzheimer's disease	11
Figure 2.1 Structure of (a) AChE and (b) BChE.....	25
Figure 2.2 Chemical structure of BChE inhibitors.....	27
Figure 2.3 Chemical structure of BChE inhibitors.....	29
Figure 3.1 Objective of the research work.....	38
Figure 4.1 Sample distribution of compounds (a) among various datasets. (b) based on class labels across (c) Number of features before and after data processing	71
Figure 4.2 Accuracies of models on various datasets.	72
Figure 4.3 Screening of sulfonamide library through the developed ML model to obtain virtual hit.	74
Figure 4.4 Reaction scheme for synthesis of 4-(phenylsulfonamido)benzoic acid derivatives...75	
Figure 4.5 Lineweaver Burk double reciprocal plot of compounds (a) 34 and (b) 37 . Dixon plot of compound (c) 34 and (d) 37 for K_i calculation.	81
Figure 4.6 Cell viability assay of compounds (a) 34 (b) 37 and (c) 54 on SH-SY5Y cells.	83
Figure 4.7 Protein-ligand interaction profile of the compounds obtained from molecular docking.....	87
Figure 4.8 Orientation of the docking poses in the cavity of BChE enzyme.....	88
Figure 4.9 Results of the developed QSAR models.....	92
Figure 4.10 RMSD and RMSF analysis of BChE, BChE complexed with compounds 34 , 37 and donepezil: (a) RMSD of protein backbone. (b) RMSD of heavy atoms of ligands. (c) RMSF of protein backbone. (d) RMSF of heavy atoms of ligands	97
Figure 4.11 SASA and RoG of BChE, complexes of compounds 34 , 37 and donepezil with BChE: (a) SASA of complexes (b) SASA of ligands (c) RoG of complexes (d) RoG of ligands.	98
Figure 4.12 Hydrogen bonding analysis:(a, d, g) Hydrogen bond interactions of compounds 34 , 37 and donepezil with BChE during MD simulation. (b, e, h) Hydrogen bond interactions with respect to simulation time of compounds 34 , 37 and donepezil with BChE. (c, f, i) % Contact time of various interacting BChE residues with compounds 34 , 37 and donepezil.....	99
Figure 4.13 Effect of compounds 34 , 37 and donepezil on (a) scopolamine-induced impairment of % spontaneous alteration, (b) novel arm entries, (c) % arm entries and (d) % total arm entries.	101
Figure 4.14 Effect of compounds 34 , 37 and donepezil on (a) primary error, (b) primary latency time.	102
Figure 4.15 Effect of compounds 34 , 37 and donepezil on total cholinesterase activity with ATCI as substrate in (a) hippocampus, (b) PFC and with BTIC as substrate in (c) hippocampus, (d) PFC.....	105
Figure 4.16 Effect of compounds 34 , 37 and donepezil on CAT activity in (a) hippocampus, (b) PFC and SOD activity in (c) hippocampus, (d) PFC.	106
Figure 4.17 Effect of compounds 34 and 37 on (a) SGOT, (b) SGPT, (c) Serum creatinine and (d) Serum urea levels.	108
Figure 5.1 Schematic representation of the design of compound.	111
Figure 5.2 Reaction scheme for synthesis of (S)-(+)-N,2-diphenyl-2-(phenylsulfonamido)acetamide derivatives.....	125
Figure 5.3 Lineweaver Burk double reciprocal plot of compounds (a) 30 and (b) 33 . Dixon plot of compound (a) 30 and (b) 33 for K_i calculation.	129
Figure 5.4 Cell viability assay of compounds (a) 30 and (b) 33 on SH-SY5Y cells.	130
Figure 5.5 Interaction profile of ligands with BChE.....	133
Figure 5.6 3D interaction of compounds (a) 26 , (b) 27 , (c) 29 , (d) 30 , (e) 33 , (f) 37 and (g) 44 with BChE.....	134

Figure 5.7 RMSD and RMSF analysis of BChE, BChE complexed with compounds 30 , 33 and donepezil: (a) RMSD of protein backbone. (b) RMSD of heavy atoms of ligands. (c) RMSF of protein backbone. (d) RMSF of heavy atoms of ligands.	137
Figure 5.8 SASA and RoG analysis of BChE, BChE complexed with compounds 30 , 33 and donepezil: (a) SASA of protein-ligand complexes. (b) SASA of ligands. (c) RoG of protein-ligand complexes. (b) RoG of ligands.	138
Figure 5.9 (a, d, g) Hydrogen bond interaction of the compounds 30 , 33 and donepezil with residues, respectively. (b, e, h) Number of hydrogen bonds formed by compounds 30 , 33 and donepezil with enzyme over the period of 50 ns, respectively. (c, f, i) Contact time of crucial residues with compounds 30 , 33 and donepezil, respectively.	139
Figure 5.10 Effect of compounds 30 , 33 and donepezil on (a) scopolamine-induced impairment of % spontaneous alteration, (b) novel arm entries, (c) % arm entries and (d) % total arm entries.	141
Figure 5.11 Effect of compounds 30 , 33 and donepezil on (a) primary error, (b) primary latency time.	142
Figure 5.12 Effect of compounds 30 , 33 and donepezil on total ChE activity with ATCI as substrate in (a) hippocampus, (b) PFC and with BTIC as substrate in (c) hippocampus, (d) PFC.	145
Figure 5.13 Effect of compounds 30 , 33 and donepezil on CAT activity in (a) hippocampus, (b) PFC and SOD activity in (c) hippocampus, (d) PFC.	146
Figure 5.14 Effect of compounds 30 and 33 on (a) SGOT, (b) SGPT, (c) S. creatinine and (d) S. urea levels. Data are expressed in Mean \pm SEM (N =6).	148
Figure 6.1 Schematic representation of the workflow for <i>in silico</i> identification of potential AChE inhibitors.	151
Figure 6.2 (a) Pharmacophore model developed from co-crystallised donepezil (b) Distance between various pharmacophore features.	155
Figure 6.3 (a) Homology model of human AChE with the modelled loops in green. (b & c) Ramachandran plot of the homology model obtained from PROCHECK and RAMPAGE.	157
Figure 6.4 (a) Binding site of donepezil inhibiting human AChE (b & c) Superimposition of docked structure (green) with co-crystallised and molecular dynamics generated structure of donepezil (magenta and purple).	158
Figure 6.5 2D and 3D interaction diagrams of (a & b) ZINC000169753041, (c & d) ZINC000013719534, (e & f) ZINC000035551243, (g & h) ZINC000035596918, and (i & j) ZINC000014996252.	161
Figure 6.6 Virtual alanine scanning analysis of critical interacting residues of compounds (a) ZINC000013719534, (b) ZINC000035551243 and (c) ZINC000035596918.	164
Figure 6.7 (a) RMSD plots of AChE protein backbone of apo-form, and complexed with ZINC000013719534, ZINC000035551243, ZINC000035551243, and donepezil (b) RMSD plots of ZINC000013719534, ZINC000035551243, ZINC000035551243, and donepezil bound to AChE.	166
Figure 6.8 (a) RMSF plot of protein residue fluctuations of AChE apo-form and complexed with ZINC000013719534, ZINC000035551243, ZINC000035551243, and donepezil (b) RMSF plot of ZINC000013719534, ZINC000035551243, ZINC000035551243, and donepezil atoms.	167
Figure 6.9 Solvent Accessible Surface Area, the plot of AChE (apo-form) and AChE complexed with ZINC000013719534, ZINC000035551243, ZINC000035551243, and donepezil.	168
Figure 6.10 Hydrogen bond interaction analysis, critical residues and their contact time during hydrogen bond interactions of ZINC000013719534, ZINC000035551243, ZINC000035551243, and donepezil with protein during MD run.	169
Figure 7.1 Structural representation of $\alpha 7$ -subunit of $\alpha 7$ -nAChR showing top, vestibular, and agonist sub-pocket binding sites.	173

Figure 7.2 (a) Cluster distribution vs. mean binding energy of various poses of AVL-3288 in the top pocket of $\alpha 7$ -nAChR (PDB id – 5AFK), (b, c, d, & e) 2-D interaction diagram of t1, t2, t3, and t4 poses in the top pocket.	177
Figure 7.3 (a) Cluster distribution vs. mean binding energy of various poses of AVL-3288 in vestibule pocket of $\alpha 7$ -nAChR (PDB id – 5AFK), (b, c, d, & e) 2-D interaction diagram of v1, v2, v3, and v4 poses in vestibule pocket.	179
Figure 7.4 (a) Cluster distribution vs. mean binding energy of various poses of AVL-3288 in agonist sub-pocket of $\alpha 7$ -nAChR (PDB id – 5AFK), (b, c, d, & e) 2-D interaction diagram of a1, a2, and a3 poses in agonist sub-pocket.	180
Figure 7.5 Virtual alanine scanning of (a) t1, (b) v1 and (c) a1 poses.	181
Figure 7.6 (a) RMSD deviation of C α of protein backbone bound to poses of $\alpha 7$ -nAChR AVL-3288 complex (t1, v1, a1) and unbound protein, (b) RMSD deviation of heavy atoms of t1, v1, a1 poses of AVL-3288 bound to $\alpha 7$ -nAChR.	183
Figure 7.7 (a) Radius of gyration (Rg) of $\alpha 7$ -nAChR AVL-3288 complex (t1, v1, a1) and unbound protein, (b) Solvent accessible surface area (SASA) of $\alpha 7$ -nAChR AVL-3288 complex (t1, v1, a1) and unbound protein.	184
Figure 7.8 Fraction of time AVL3288 (t1, v1, a1) displayed hydrogen bonding with $\alpha 7$ -nAChR amino acid residues.	185
Figure 8.1 (a) Hits returned from a protein blast search using eeAChE as a query. (b) Comparison of eeAChE sequence with other organisms.	193
Figure 8.2 Sequence alignment of AChE <i>Electrophorus electricus</i> with <i>Tetronarce californica</i> , <i>Mus musculus</i> and <i>Homo sapiens</i>	194
Figure 8.3 (a) Sequence comparison (b) % homology, identity and similarity of active site and tunnel of eeAChE with other organisms.	195
Figure 8.4 (a) Homology model of eeAChE, (b) Ramachandran plot of homology model after energy minimisation, (c) Validation score of protein structure at various stages of energy minimisation, (d) Potential energy (Kcal/mol) of the protein model during energy minimisation.	199
Figure 8.5 (a, b, c, d, e) 3D interaction diagrams of 5GZ, A36, B3V, B3W and B32 with eeAChE, (f) Comparison of the interaction profile of 5GZ, A36, B3V, B3W and B32 with mmAChE and eeAChE.	201
Figure 8.6 (a, b, c, d, e) 3D interaction diagrams of C56, DUC, E5H, E5K and GC8 with eeAChE, (f) Comparison of the interaction profile of C56, DUC, E5H, E5K and GC8 with mmAChE and eeAChE.	202
Figure 8.7 (a, b, c, d, e) 3D interaction diagrams of N2K, Q4Q, SOF, Z5K and ZN4 with eeAChE, (f) Comparison of the interaction profile of N2K, Q4Q, SOF, Z5K and ZN4 with mmAChE and eeAChE.	203
Figure 8.8 (a, b, c) IC ₅₀ cut-off, area under the curve and receiver operating characteristic of the docking validation set for validation of the Autodock scoring function using binary classification. (d) Scatter plot showing the relationship between binding energy and pIC ₅₀ . (e) Scatter plot showing the relationship between pK _i and pIC ₅₀	206
Figure 8.9 (a) Proportion of compounds used in training and test sets for the ML, (b) Compounds labelled as active and inactive at IC ₅₀ cut-off value of 1000 nM for binary classification, (c) Compounds labelled as most active, active and moderately active at IC ₅₀ cut-off values of 1000 and 10000 nM for multi-class classification, (d) Distribution of the compounds in training and test sets on a log IC ₅₀ scale for regression modelling.	207
Figure 8.10 (a, b) Receiver operating characteristic of binary and multiclass scoring function, respectively for the bagging classifier. (c) Plot between log IC ₅₀ and predicted LogIC ₅₀ of the training and test sets for RF regressor. (d) Plot between log IC ₅₀ and predicted Log IC ₅₀ of the training and and 95% of the test set after removing outliers for RF regressor.	216
Figure 9.1 (a) Validation score of protein structures at various energy minimisation stages, (d) Potential energy (Kcal/mol) of the protein models during energy minimisation.	225

Figure 9.2 (a) Active site of human BChE (b) Sequence comparison (active site and tunnel residues) of ecBChE with human.	226
Figure 9.3 (a, b, c, d, e) 3D interaction diagrams of 40V, 3F9, 92H, 5HF and 9A5 with ecBChE, (f) Comparison of the interaction profile of 40V, 3F9, 92H, 5HF and 9A5 among human and horse BChE.	227
Figure 9.4 (a, b, c, d, e) 3D interaction diagrams of HUK, HUQ, HUN, BUW and THA with ecBChE, (f) Comparison of the interaction profile of HUK, HUQ, HUN, BUW and THA among human and horse BChE.	228
Figure 9.5 (a, b, c) IC ₅₀ cut-off, area under the curve and receiver operating characteristic of the docking validation set for validation of the Autodock scoring function using binary classification. (d and e) Scatter plot showing the relationship between binding energy with standardised IC ₅₀ and Log IC ₅₀ , respectively.	230
Figure 9.6 Data distribution for (a) classification models, (b and d) regression models dataset with IC ₅₀ below and above 10000 nM, (c and e) Log IC ₅₀ distribution of training and test sets for regression models dataset with IC ₅₀ below and above 10000 nM.	232
Figure 9.7 ROC of various algorithms used for the generation of binary classification models.	233
Figure 9.8 Scatter plot between predicted and experimental IC ₅₀ obtained from various algorithms for the development of regression-based models for the compounds with IC ₅₀ below 10000 nM (active).	236
Figure 9.9 Scatter plot between predicted and experimental IC ₅₀ obtained from various algorithms for developing regression-based models for compounds with IC ₅₀ above 10000 nM (moderately active).	238
Figure 9.10 (a) Comparison between the Autodock SF and selected ML models, (b) Schematic representation of the formulated scoring function.	242
Figure 10.1 Distribution of datasets into active and inactive classes.	246
Figure 10.2 Feature selection using filters.	247

List of abbreviations

Abbreviation	Full Form
3D	Three dimensional
ACh	Acetylcholine
AChE	Acetylcholinesterase
AD	Alzheimer's disease
AI	Artificial intelligence
Aβ	Amyloid- β
APP	Amyloid precursor protein
AS	Anionic site
ATP	Adenosine triphosphate
AUC	Area under the curve
BBB	Blood-brain permeability
BChE	Butyrylcholinesterase
BLAST	Basic local alignment search tool
CaMKII	Ca ²⁺ /calmodulin dependent protein kinase II
CAT	Catalase
ChAT	Choline acetyl transferase
ChE	Cholinesterase
CNN	Convolutional neural network
CNS	Central nervous system
CT	Computed tomography
DNP	Donepezil
EAAT2	Excitatory amino acid transporter 2
EC	Enzyme classification
FN	False negative
FP	False positive
GAFF	Generalised amber force field
GSK3	Glycogen synthase kinase-3
iGluRs	ligand-gated ionotropic glutamate receptors
JNK3	c-Jun N-terminal kinase 3
KNN	K-nearest neighbors
LBDD	Ligand based drug design
LDA	Linear discriminant analysis
LGA	Lamarckian Genetic Algorithm
LR	Logistic regression
MACCS	Molecular access system
MD	Molecular dynamics
MEKK	Mitogen-activated Protein/ERK Kinase Kinases
ML	Machine learning
MLP	Multi-layer perceptron
MRI	Magnetic resonance imaging
MSME	Mini-Mental state exam
Nct	Nicastrin
NFT	Neurofibrillary tangle

NMDA	N-methyl D-aspartate
PAC	Passive-aggressive classifier
PAINS	Pan-assay interference compounds
PAM	Positive allosteric modulator
PAMPA	Parallel artificial membrane assay
PAS	Peripheral anionic site
PBL	Porcine brain lipid
PLIP	Protein ligand interaction profiler
PS1	Presenilin 1
PS2	Presenilin 2
QDA	quadratic discriminant analysis
QSAR	Quantitative structure-activity relationship
R_f	Retention factor
RF	Random forest
R_g	Radius of gyration
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RO5	Lipinski rule of five
ROC	Receiver operation characteristic
ROS	Reactive oxygen species
RT	Room Temperature
SAR	Structure activity relationship
SASA	Solvent accessible surface area
SBDD	Structure-based drug design
SBVS	Structure based virtual screening
SCO	Scopolamine hydrobromide
SEM	Standard error of mean
SF	Scoring function
SOD	Superoxide dismutase
SVC	Support vector classifier
SVM	Support vector machine
SVR	Support vector regression
TI	Thermodynamic integration
TN	True negative
TP	True positive

Preface

Alzheimer's disease (AD) is the most common form of dementia causing memory, behaviour and thinking impairment. Eventually, the symptoms become severe and make it difficult for a patient to carry out daily activities. According to the World Health Organization (WHO), one in every 85 individuals will have AD by 2050. The therapeutic targets of the disease include acetylcholinesterase (AChE), butyrylcholinesterase (BChE), β -secretase-1, glycogen synthase kinase 3 β , monoamine oxidase B, matrix metalloproteases, N-methyl D-aspartate (NMDA) receptors, tau kinase etc.

Among the targets, inhibition of cholinesterase enzymes is still a major component of anti-AD therapy to provide symptomatic relief. The inhibition of AChE causes improvement in memory and cognition. However, AChE inhibitors produce cholinergic side effects and the therapeutic effect wear-off with the progression of the disease. Alternatively, the presence of a significant level of BChE in the latter stage of the disease and its inhibition causes improvement in memory and thus, makes it an attractive target. Machine learning (ML), structure-based drug design and ligand-based drug design are useful techniques in drug design. The research work presented in the thesis covers three-fold objectives. The first objective of the study is to design selective BChE inhibitors through ML/scaffold hopping. The ligands identified were synthesised, characterised and tested through various *in vitro* and *in vivo* tests. The second objective deals with the identification of the virtual hits and their binding modes. The third objective includes the development of *in silico* tools by using ML techniques for the identification of hits.

The work embodied in this thesis has been presented under the following chapters:

Chapter 1: The chapter provides an introduction to AD and deals with details regarding background, pathophysiology and available therapeutics for the treatment of AD. Further, the various approaches involved in drug design are also described.

Chapter 2: The chapter deals with the literature background related to targets involved in the cholinergic hypothesis. It also includes the field application of ML in drug discovery.

Chapter 3: In this chapter, the objectives of the study and plan of work are incorporated.

Chapter 4: The chapter deals with the development of selective BChE inhibitors using ML. It includes the methodology used for design, synthesis, characterisation, *in vitro*, *in silico* and *in vivo* evaluations of *sulfonamides of para-amino benzoic acid*, followed by a discussion.

Chapter 5: The development of selective BChE inhibitors using scaffold hopping is presented in the chapter. It describes the methodology used for design, synthesis, characterisation, *in vitro*, *in silico* and *in vivo* evaluations of *sulfonamides of phenylglycine*, followed by results and discussion.

Chapter 6: The chapter deals with *in silico* identification of the potential AChE inhibitors through computational techniques.

Chapter 7: The chapter includes the methodology and results obtained from *in silico* analysis of the binding mode of AVL-3288 with $\alpha 7$ -nicotinic acetylcholine receptor.

Chapter 8: The chapter includes the details of the procedure followed and results of the development of the homology model, docking protocol and ML-based scoring function for identification of electric eel's AChE inhibitors.

Chapter 9: In this chapter, the detailed procedure of the development of the homology model, docking protocol and ML-based scoring function for the identification of horse's BChE inhibitors is presented. The chapter also includes results and discussion on the above.

Chapter 10: The chapter deals with the development of ML models for the prediction of inhibitors for the important targets of AD. The ML models are deployed in the form of a web application – AlzLeads.

Chapter 11: This chapter outlines the summary and conclusions of the research work undertaken.

Chapter 12: The references, used to carry out the research work, are presented in the chapter.

An appendix consisting of the additional supporting information, spectral data of representative compounds and a list of publications during the course of the Ph.D. are included.

Acknowledgements

*At the very outset, I would like to thank the **ALMIGHTY GOD** for showering his continuous blessings on me in my daily life.*

*I express my sincere gratitude to the founder of the **Banaras Hindu University, Bharat Ratna Mahamana Pandit Madan Mohan Malviya ji**, who created this glorious temple of learning.*

*I would like to express my sincere gratitude and indebtedness to my venerated teacher and Supervisor **Prof. Sushil Kumar Singh**, Indian Institute of Technology (Banaras Hindu University), for his invaluable suggestions, constant encouragement & untiring endeavour. I am very thankful to my co-supervisor **Dr. Ashok Kumar** for his guidance, constructive suggestions and constant encouragement throughout my Ph.D.*

*I immensely express my profound gratitude to **Prof. P.K. Jain**, Director, Indian Institute of Technology (Banaras Hindu University), Varanasi and the Deans of the institute for providing necessary amenities and sophisticated instrumentation facilities for smooth conduct of research.*

*I would like to render my sincere thanks to **Prof. B. Mishra** and **Prof Sanjay Singh**, former Head of the department and **Prof. S. K. Shrivastava**, Head of the department, for providing me with the opportunity and necessary amenities to conduct research in the department.*

*I would like to thank my **Research Progress and Evaluation Committee (RPEC) members** for providing valuable suggestions during the evaluation of my research progress.*

*I would like to thank **Prof. Sairam Krishnamurthy** and his lab members for their support in performing in vivo experiments. I am thankful to **Prof. David A. Case** for providing licenses for Amber18, Amber20 and **Dr. Stefano Forli** for providing help in virtual screening. The support and the resources provided by PARAM Shivay Facility, Centre for Computing and Information Services (CCIS) and Central Instrument Facility (CIF), IIT(BHU), Varanasi are gratefully acknowledged.*

*I would like to keep on record my profound respect & sincere thanks to my respected teachers, **Prof.(Mrs.) S. Hemalatha, Dr. Senthil Raja A, Dr. Alakh N Sahu, Dr. S. K. Mishra, Mr. A. K. Srivastava, Dr.(Mrs.) Ruchi Chawla, Dr. M. S. Muthu, Dr. Prasanta Kumar Nayak, Dr. Gyan Prakash Modi, Dr. Sreyans K. Jain, Dr. Deepak Kumar, Dr. Rajnish Kumar, Dr. Vinod Tiwari and Dr. Ashish K. Agrawal** for their co-operation and valuable suggestions during my research work.*

*I pay sincere thanks to **Mr Yashwant Singh sir, Mr Atul Gupta sir, Mr Anand sir** and all the departmental non-teaching staff for their co-operation and help during my research work.*

*I am heartily thankful to **Devendra Sir, Dileep Sir, Satheesh Sir, Srabanti Mam, Gopi Sir, Avanish Sir, Rayala Swetha, Subhojit Makar** and all my M. Pharm. juniors of the lab for their valuable suggestions and co-operation.*

*My heartiest thanks are due to my friends **Qadir Alam, Ravi Singh, Kaushik Neogi and Pratigya Tripathi** for their care & support during my stay in IIT(BHU).*

Ministry of Education, New Delhi is gratefully acknowledged for providing fellowship during the tenure of my research work.

*Last but not least, words are incapable of expressing the depth of one's feelings. I would like to express my heartfelt and immense gratitude to my beloved **PARENTS, Aditya Dada, Prachi Vahini**, little **Advait** and all my relatives who had been the most significant pillars of my life. They always stood beside me and without them, I would never have been succeeded.*

Date:

Place – Varanasi

(Ankit Ganeshpurkar)

**Department of Pharmaceutical Engineering & Technology
Indian Institute of Technology
(Banaras Hindu University)
Varanasi-221005**

CERTIFICATE

It is certified that the work contained in the thesis titled “**Implementation of Computational and Machine learning techniques for the development of *in silico* tools and identification of novel leads for the treatment of Alzheimer's Disease**” by **Mr. Ankit Ganeshpurkar** has been carried out under our supervision and that this work has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all the requirements of Comprehensive Examination, Candidacy and SOTA for the award of Ph.D. Degree.

Prof. Sushil K. Singh
(Supervisor)

Dr. Ashok Kumar
(Co-supervisor)

Date:
Place: IIT (BHU), Varanasi

DECLARATION BY THE CANDIDATE

I, **Ankit Ganeshpurkar**, certify that the work embodied in this Ph.D. thesis is my own bonafide work and carried out by me under the supervision of **Prof. Sushil K. Singh** and co-supervision of **Dr. Ashok Kumar** from **July, 2016 to September, 2021** at the **Department of Pharmaceutical Engineering & Technology, Indian Institute of Technology (Banaras Hindu University), Varanasi**. The matter embodied in this Ph.D. thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credit to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, etc. reported in the journals, books, magazines, reports, dissertations, theses, etc., or available at websites and have not included them in this Ph.D. thesis and have not cited as my own work.

Date:

Place: IIT (BHU), Varanasi

Ankit Ganeshpurkar

CERTIFICATE BY THE SUPERVISOR(S) AND HEAD OF THE DEPARTMENT

It is certified that the above statement made by the student is correct to the best of our knowledge.

**Prof. Sushil K. Singh
(Supervisor)**

**Dr. Ashok Kumar
(Co-Supervisor)**

**Prof. S. K. Shrivastava
(Head of the Department)**

**Department of Pharmaceutical Engineering & Technology
Indian Institute of Technology
(Banaras Hindu University)
Varanasi-221005**

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: Implementation of Computational and Machine learning techniques for the development of *in silico* tools and identification of novel leads for the treatment of Alzheimer's Disease

Candidate's Name: Ankit Ganeshpurkar

Copyright Transfer

The undersigned hereby assigns to the Indian Institute of Technology (Banaras Hindu University), Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the Ph.D. degree.

Date: 4.10.2021

Place: IIT (BHU), Varanasi

Ankit Ganeshpurkar

Note: However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for author's personal use provided that the source and University's copyright notice are indicated.

