

**HIERARCHICAL AND FRAGMENT BASED
VIRTUAL SCREENING APPROACHES IN THE
DISCOVERY OF POTENT
ACETYLCHOLINESTERASE INHIBITOR**

ERMA FATIHA BINTI MUHAMMAD

UNIVERSITI SAINS MALAYSIA

2022

**HIERARCHICAL AND FRAGMENT BASED
VIRTUAL SCREENING APPROACHES IN THE
DISCOVERY OF POTENT
ACETYLCHOLINESTERASE INHIBITOR**

by

ERMA FATIHA BINTI MUHAMMAD

**Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy**

August 2022

ACKNOWLEDGEMENT

In the name of Allah, the Most Loving and the Most Merciful. All praises go to Allah for the strength, health, and passion He has given to me to complete this thesis. This thesis would not have been possible without the support of many people. I would like to express my deepest gratitude first and foremost, to my main supervisor, Prof Habibah A. Wahab for her tremendous motivation, support and opportunity in this academic journey. Her levels of patience, knowledge, and ingenuity are something I will always keep aspiring to. Thank you so much, Prof.

I am indebted to my field supervisor, Prof Kam Zhang for the extraordinary research experience, and opportunity to grow professionally while in RIKEN. Very special thanks to my teammates in Structural Bioinformatics team (Ashutosh, Kutu, Dileep, Aditya, Mateij, Johny, Rahul and Taya-san). It is an honour to learn and be part of the research team. I would also like to thank my co-supervisors, Assoc. Prof. Dr. Choi Sy Bing and Dr. Hassan Hadi who offered me guidance and support while I was in USM. I would also like to acknowledge a team in pharmaceutical and simulation laboratory in USM especially Mira, Nadhirah, Keseven, Selestin, Fadi, Adibah, Nasuha, Aisyah, Faisalina, Maram and Thiabat.

Further, I owe my appreciation to my father, Muhammad, my husband, Faizal, and my family for their strong support and inspiration for the whole of my postgraduate journey. Last but not the least, I owe my thanks to Universiti Sains Malaysia and RIKEN Institute Yokohama Japan, for providing me with the financial means to complete my projects. Alhamdulillah.

TABLE OF CONTENTS

ACKNOWLEDGEMENT	ii
TABLE OF CONTENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
LIST OF SYMBOLS	xi
LIST OF ABBREVIATIONS	xiii
LIST OF APPENDICES	xvi
ABSTRAK	xvii
ABSTRACT	xx
CHAPTER 1 INTRODUCTION	1
1.1 Problem Statement	1
1.2 Objectives.....	2
1.3 Content of Thesis.....	3
CHAPTER 2 LITERATURE REVIEW	4
2.1 Alzheimer's Disease.....	4
2.1.1 Etiology.....	5
2.1.2 Causes of AD.....	6
2.1.2(a) Cholinergic Hypothesis.....	6
2.1.2(a) (i) Acetylcholinesterase (AChE).....	7
2.1.2(a) (ii) Structural and Functional Sites.....	8
2.1.2(a) (iii) Function and Mechanism.....	9
2.1.3 Current Drugs Available and Research Targeting Cholinesterase for AD Treatment.....	12
2.2 Molecular Modelling.....	14

2.2.1	Quantum Mechanical (QM) Methods.....	15
2.2.2	Molecular Mechanical (MM) Methods.....	19
2.2.3	Hybrid Quantum Mechanical/Molecular Mechanical (QM/MM) Methods.....	21
2.3	Computer Aided Drug Design and Virtual Screening.....	23
2.3.1	Structure-Based Virtual Screening (SBVS) and Approaches.....	26
2.3.2	Ligand-Based Virtual Screening (LBVS) and Approaches.....	27
2.3.3	Fragment-Based Virtual Screening (FBVS).....	30
2.4	Molecular Dynamics Simulation.....	31
2.4.1	Force Field.....	35
2.5	Molecular Modelling for the Discovery of Acetylcholinesterase Inhibitors...	36
	CHAPTER 3 METHODOLOGY	38
3.1	Methodology.....	38
3.1.1	Hardware and Software Used in the Study.....	41
3.1.2	Source of Compounds for Virtual Screening.....	42
3.2	Virtual screening of Namiki-Shoji Compound Database for Identifying Scaffold with Potential Acetylcholinesterase Inhibitory Properties.....	43
3.2.1	Shape Similarity Calculation.....	43
3.2.2	Molecular Docking and $\Delta_{MM/GBSA}$ Scoring.....	44
3.3	Fragment-Based Approach Virtual Screening to Identify Potent AChE Inhibitor.....	46
3.4	<i>In Vitro</i> Acetylcholinesterase (AChE) Inhibition Assay.....	48
3.4.1	Materials and Chemicals.....	48
3.4.2	<i>In Vitro</i> Acetylcholinesterase (AChE) Inhibition Assay.....	49
3.5	Post Analysis.....	46
3.5.1	Molecular Orbital Calculation.....	50
3.5.2	Molecular Dynamics Simulation.....	51

3.5.3	Binding Free Energy Analysis Using MM/GBSA.....	52
3.5.4	Pan-Assay Interference (PAINS) and ADME Analysis.....	52
CHAPTER 4	RESULTS AND DISCUSSION.....	53
4.1	Scaffold Search Virtual Screening for AChE Inhibitors	53
4.1.1	Shape Similarity Screening.....	53
4.1.2	Molecular Docking and $\Delta_{MM/GBSA}$ Scoring.....	55
4.1.3	<i>In Vitro</i> AChE Inhibition Assay.....	58
4.1.4	Binding Mode and Structure Activity Relationship of the Selected Compounds.....	65
4.1.5	ADME Analysis.....	72
4.2	Fragment Based Approach to Identify Potent Nanomolar AChE Inhibitor....	75
4.2.1	Virtual Screening Protocol.....	75
4.2.2	Two-Round Molecular Docking and $\Delta_{MM/GBSA}$ Scoring.....	81
4.2.3	<i>In Vitro</i> AChE Inhibition Assay of Compound 1-19.....	82
4.2.4	Binding Mode of CF19 and Its Structure Activity Relationship.....	95
4.2.5	HOMO-LUMO Theory.....	98
4.2.6	Molecular Dynamics Simulation of Human AChE with CF19	100
4.2.7	Binding Free Energy.....	114
4.2.8	ADME Analysis.....	117
CHAPTER 5	CONCLUSION AND FUTURE RECOMMENDATIONS....	120
5.1	Acomplished of the Objectives	120
5.2	Recommendations for Future Research	121
REFERENCES.....		122
LIST OF PUBLICATIONS		

LIST OF TABLES

	Page
Table 2.1	Public and Commercial Databases Available28
Table 3.1	List of Software Used in This Study.....41
Table 4.1	Structure AChE Inhibitory Activities of Compounds with 1, 2, 4-Triazolylthioethanone Scaffold.....61
Table 4.2	Predicted Pharmacokinetic Parameters of Compound 1-13 73
Table 4.3	IC ₅₀ values of 19 Selected Compounds and Donepezil.....87
Table 4.4	XP-Score, Prime-Based Binding Free Energy and Ligand-Interacting Amino Acids.....97
Table 4.5	DFT Calculation for The Eight Potent Compounds in This Study 99
Table 4.6	MM/GBSA Binding Free Energy Change Profiles of Complex Structures of Donepezil/Human AChE and CF19/Human AChE.....116
Table 4.7	Predicted Pharmacokinetic Parameters of Compound 1-19.....118

LIST OF FIGURES

	Page
Figure 2.1	Schematic representation of AChE binding sites (Colovic et al. 2013) 9
Figure 2.2	Mechanism of AChE mechanism in neurotransmission (Colovic et al. 2013) 11
Figure 2.3	Schematic diagram of substrate binding to the enzyme AChE at the active site of the enzyme (Wiener, 2004) 12
Figure 2.4	AChE inhibiting drugs approved for the treatment of Alzheimer's disease 14
Figure 2.5	Representation of bonded and non-bonded interactions used in molecular mechanics..... 20
Figure 2.6	Modern drug design development. Each step contains characteristic actions, methods, and tools used. Computer-aided drug design can be applied to facilitate drug discovery in the phases denoted with a star (*). 24
Figure 2.7	Representative methods of SBVS and LBVS in hits discovery 26
Figure 2.8	Two compounds of cholinesterase inhibitors which have been discovered using the combined LBVS and SBVS protocol. Each inhibitor's potency is given for the specific enzyme human AChE (hAChE) and Electrophorus electricus (EeAChE) isoforms..... 37
Figure 3.1	Flowchart of the study in this thesis..... 40
Figure 3.2	Virtual screening workflow employed to identify 1, 2, 4-triazolylethanone scaffold compounds..... 43
Figure 3.3	The 2D and 3D structure of donepezil..... 44
Figure 3.4	Virtual screening protocol to identify AChE inhibitors..... 47

Figure 4.1	Distribution of ShapeTanimoto scores for the 1000 top-ranked compounds from the screening library (a). Overlay of Compounds 8 (magenta atom color) and 10 (yellow atom color) with donepezil (green atom color) (b). The shape of each compound is also shown as surface representation.....55
Figure 4.2	Glide docking energy scores of shortlisted 73 compounds from the first round of molecular docking screening (a). $\Delta_{MM/GBSA}$ energy of 73 compounds from the standard virtual screening methodology (b).....57
Figure 4.3	The binding interactions of the selected 58 compounds on AChE. All compounds show similar important binding interactions with TRP286 and TRP8658
Figure 4.4	Evaluation of AChE inhibitory activities of 58 compounds selected via virtual screening (a). Dose-independent inhibitory activities of the thirteen compounds against AChE (b). Dose-independent inhibitory activities of donepezil against AChE (c). Donepezil was used as a positive control for AChE inhibition.....60
Figure 4.5	Predicted binding modes of Compounds 1 - 13 in overlay with donepezil. Docking without binding site water (a) and docking with binding site water (b). Compounds are shown in cyan atom color while donepezil is shown in magenta sticks.....67
Figure 4.6	1, 2, 4-triazole ring scaffolds of compounds 1-13 which is predicted to occupy a region close to hydrophobic aromatic residues of PHE341, TYR337, and PHE338 and was found to make π - π stacking contacts with TYR337. Hydrophobic residues are represented in surface and compounds 1-13 are represented in wire representation.....69
Figure 4.7	Predicted binding modes of two most potent compounds, Compound 8 (a) and Compound 10 (b). Compounds are shown in cyan atom color while AChE is shown in green.....70
Figure 4.8	PIC ₅₀ value of 3840 AChE inhibitors from ChEMBL database in the range IC ₅₀ value between 0.006 to 1000 nm.....76
Figure 4.9	Number of AChE inhibitors by fragments from ChEMBL database in the range IC ₅₀ value between 0.006 to 1000 nm.....79

Figure 4.10	Number of Namiki Shoji compounds by selected fragment from previous study. Number of acridine containing compounds is 2530...	80
Figure 4.11	Percentage of inhibition for 19 compounds against AChE activity.....	84
Figure 4.12	Dose-independent inhibitory activity of donepezil (control) and CF19 against AChE. IC ₅₀ of donepezil and CF19 are 0.035 and 0.028 μM. The initial compound concentration was 50 μM and subsequently diluted to get the set concentration of 0.0064, 0.0032, 0.016, 0.08, 0.40, 2.00, 10.00 and 50.00 μM. Donepezil was used as a positive control for AChE inhibition.....	85
Figure 4.13	The binding mode of <i>N-benzylpiperidine</i> containing compounds (CF15 (IC ₅₀ : 0.65 μM), C18 (IC ₅₀ : 2.69 μM) and CF19 (IC ₅₀ : 28 nM) with AChE (a) the binding mode of <i>benzimidazole chloride</i> containing compound (CF6) (IC ₅₀ : 0.96 μM) with AChE (b).....	86
Figure 4.14	Predicted binding modes of CF19	96
Figure 4.15	Correlation between the experimental pIC ₅₀ (= -log IC ₅₀) and HOMO LUMO gap (HLG).....	100
Figure 4.16	Root mean square deviations (RMSDs) of backbone atoms of AChE protein upon binding to donepezil structure (black colour) and compound CF19 (red colour) during 100 ns simulation time.....	103
Figure 4.17	Root mean square deviations (RMSDs) of ligands, donepezil (black colour) and compound CF19 (red colour) upon the binding with AChE protein.....	104
Figure 4.18	Root mean square deviations fluctuation (RMSF) of backbone atoms of AChE protein in AChE-compound CF19 system (black colour) and AChE-donepezil system (red colour).....	105
Figure 4.19	Root mean square deviations fluctuation (RMSF) of ligands, donepezil (a) and compound CF19 (b) upon the binding with AChE.....	106
Figure 4.20	MD predicted binding mode of CF19 (a) and donepezil (b) which has been extracted at every 100 ps during 100 ns simulation time. H atom is removed for clarity.	107
Figure 4.21	Radius of gyration of backbone atoms of AChE protein upon binding with donepezil (black color) and compound CF19 (red color).....	110
Figure 4.22	Solvent accessible surface area (SASA) of AChE protein upon binding with donepezil (black colour) and compound CF19 (red colour)....	111

Figure 4.23	The interactions' fraction formed in AChE-compound CF19 complex system. Green, light violet, magenta, and blue represented hydrogen bonds, hydrophobic interactions, ionic and water bridge, respectively (a). The crucial AChE residues identified by percentage that were involved in interactions with compound CF19 for entire simulation time (b).....	113
Figure 4.24	AChE-compound CF19 complex structure during simulation time..	114

LIST OF SYMBOLS

a_i	acceleration of molecule i
T	actual temperature
\AA	angstrom
θ	bending of valence angles
$\Delta_{\text{MM/GBSA}}$	binding free energy which was computed using MM/GBSA
K	Boltzman's constant
T_0	desired temperature
r_{0ij}	distance at which LJ potential has its minimum
r_{ij}	distance between atom i and j
F_i	force exerted on molecule i
H	enthalpy
IC_{50}	half maximal inhibitory concentration
H	Hamiltonian
kcal/mol	kilocalorie per mol
E	kinetic energy
m_i	mass of the molecule i
μl	microliter
μM	micromolar
mM	millimolar
ml	mililiter
Ψ	molecular wave function
P	momentum
PIC_{50}	negative log of half maximal inhibitory concentration
nm	nanometer
nM	nanomolar
ns	nanosecond
N	numbers of particles
q_i	partial atomic charge
γ	phase
n	periodicity
V	potential energy

P	pressure
ω	rotation of torsional angles
l	stretching of bonds
T	temperature
τ_T	temperature time constant
t	time
N_{df}	total number of degree of freedom
ϵ_0	vacuum permittivity
v_i	velocities of molecule i
V	volume
ϵ_{ij}	well depth

LIST OF ABBREVIATIONS

ACh	acetylcholine
AChE	acetylcholinesterase
AD	Alzheimer's diseases
ADMET	absorption, distribution, metabolism, excretion, and toxicity
AM1	Austin Model 1
ARG	arginine
Asp	aspartic acid
ATCI	acetylthiocholine iodide
BTCI	butyrylcholine iodide
BuChE	butyrylcholinesterase
B3LYP	Becke 3-Parameter (Exchange), Lee, Yang and Parr
CADD	computer-aided drug design
CAMD	computer-aided molecular design
CAMM	computer-aided molecular modelling
CAS	cationic site
Caco-2	human colon adenocarcinoma
CI	configuration interaction
CC	coupled-cluster
CF	compound fragment
CG	coarse grain
CNS	central nervous system
COVID-19	corona virus disease
CYP-450	cytochrome P450
Cys	cysteine
DFT	density functional theory
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTNB	5,5'-dithio-bis-(2-nitrobenzoic acid)
EC	enzyme commission
<i>Ee</i> AChE	electric eel acetylcholinesterase
FDA	Food and Drug Administration (US)

GANs	Generative Auto-encoders (Gas) and Generative Adversarial Networks
GLU	glutamic acid
GLY	glycine
hAChE	human Acetylcholinesterase
HF	Hartree-Fock
His	histidine
HIV	Human Immunodeficiency Virus
HOA	human oral absorption
HOMO	highest occupied orbital
HTS	high-throughput screening
HTVS	hierarchical virtual screening
IUPAC	International Union of Pure and Applied Chemistry
LBVS	ligand-based virtual screening
LJ	Lennard-Jones
Log HERG	predicted IC ₅₀ value for blockage of HERG K ⁺ channels
Log Po/w	predicted octanol/water partition coefficient
Log S	predicted aqueous solubility
LUMO	lowest unoccupied orbital
LYS	lysine
MBPT	many-body perturbation theory
Mol wt	molecular weight
MC	Monte Carlo
MCSS	maximal common substructure
MD	molecular dynamics
MM	molecular mechanics
MM-GBSA	molecular mechanics-generalized Born surface area
NCI	National Cancer Institute
NMDA	N-methyl-D-aspartate
NMR	nuclear magnetic resonance
NPH	constant-enthalpy ensemble
NPT	isothermal isobaric ensemble
NSH	constant-stress ensemble
NVE	microcanonical ensemble

NVT	canonical ensemble
PAINS	pan-assay interference
PAS	peripheral anionic site
PBC	periodic boundary condition
PDB	Protein Data Bank
PES	potential energy surface
PHE	phenylalanine
QM/MM	quantum mechanic/molecular mechanic
QLogS	water solubility
QSAR	quantitative structure-activity relationships
RMSD	root mean square root deviation
RNA	ribonucleic acid
RCSB	Research Collaboratory for Structural Bioinformatics
RESPA	Reference System Propagator Algorithm
SARS	severe acute respiratory syndrome
SASA	solvent accessible surface area
SBVS	structure-based virtual screening
SCA	scaffold-based approach
Ser	serine
SP	standard precision
THR	threonine
TRP	tryptophan
TYR	tyrosine
VAL	valine
vdW	van der Waals
VS	virtual screening
XP	extra precision
μ VT	grand canonical ensemble
3D	three dimensional

LIST OF APPENDICES

- APPENDIX A THE 2D STRUCTURES AND THEIR ACHE INHIBITORY ACTIVITIES OF 58 HIT COMPOUNDS IDENTIFIED BY HIEARARCHICAL VIRTUAL SCREENING
- APPENDIX B THE 2D STRUCTURES OF CF1-19 COMPOUNDS AND THEIR ACHE INHIBITORY ACTIVITIES

PENDEKATAN SARINGAN MAYA BERASASKAN HIERARKI DAN FRAGMENT DALAM PENCARIAN PERENCAT ASETILKOLINESTERASE

ABSTRAK

Permodelan molekul telah terbukti dapat mempercepat dalam membantu reka bentuk ubat dengan memberikan pemahaman tentang sistem biomolekul pada tahap atom. Aplikasi pemodelan molekul yang melibatkan dua kaedah penyaringan maya untuk memudahkan reka bentuk ubat untuk mencari perencat Asetilkolinesterase (AChE) dilakukan dalam kajian ini. Terdapat dua kaedah penyaringan maya yang dibentangkan dalam kajian ini iaitu penyaringan maya berasaskan hierarki dan penyaringan maya berasaskan fragmen. Kaedah penyaringan maya hierarki (HVS) meliputi penyaringan kesamaan bentuk yang digabungkan dengan pendokkan molekul dan pemeriksaan visual membolehkan penemuan pada kelas sebatian perencat AChE yang baru iaitu, 1, 2, 4-triazoliltioetanon. Pada langkah pertama HVS, pemeriksaan kesamaan bentuk digunakan untuk mencari sebatian yang mempunyai bentuk yang sama dengan struktur donepezil. Sebanyak 920 sebatian telah disenarai pendek dan mempunyai bentuk yang sama dengan sebatian donepezil mengikut skor "ShapeTanimoto". Seterusnya, pendokkan molekul dilakukan dan menghasilkan kira-kira 73 sebatian yang mempunyai nilai tenaga pendokkan di antara -13.35 hingga -5.35 kkal/molar dan nilai tenaga $\Delta_{MM/GBSA}$ di antara -30.72 hingga -86.26 kkal/molar. Seterusnya, semua sebatian yang disenarai pendek diperiksa secara visual berdasarkan bentuk dan interaksi pada poket protein AChE. Sebanyak 58 sebatian telah di senarai pendek untuk dibeli dan diuji secara eksperimen. Tiga belas sebatian yang mempunyai bentuk memanjang seperti donepezil terdiri daripada 1, 2, 4- triazoliltioetanon dan dilaporkan mempunyai nilai IC_{50} dalam julat 0.15 ± 0.07 hingga $3.32 \pm 0.92 \mu\text{M}$. Tiga

belas sebatian ini adalah terdiri daripada dua struktur aromatik di kedua-dua hujung sebatian dengan kelas 1, 2, 4-triazol sebagai penghubung sebatian. Pendokkan molekul menunjukkan tiga penggantian sebatian, R1, R2 dan R3 terikat pada tiga poket protin AChE yang berbeza dan berinteraksi dengan amino asid TRP86 dan TRP286 AChE. Struktur kelas 1,2,4-triazol dalam sebatian 1-13 berada berdekatan dengan struktur aromatik amino asid yang bersifat hidrofobik seperti PHE341, TYR337, dan PHE338. Sebatian ini juga terikat pada amino asid TYR337 melalui ikatan interaksi π - π . Tiga belas sebatian tersebut juga belum dilaporkan sebagai perencat protin AChE sebelum ini. Pada bahagian seterusnya, pendekatan penyaringan maya berasaskan fragmen (FBVS) digunakan untuk mencari perencat protin AChE yang lebih aktif. Penyaringan maya melibatkan pencarian substruktur dan pendokkan molekul telah digunakan. Teknik ini membawa kepada penemuan sebatian aktif, **CF19** dengan nilai IC_{50} bernilai 28 nanomolar. Dengan menggunakan pencarian substruktur sebagai langkah awal, dua puluh satu fragmen dari kumpulan perencat AChE daripada pangkalan data ChEMBL telah diambil. Fragmen akridin (F3) telah didapati sebagai fragmen paling banyak ditemui (2530 sebatian). Fragmen ini seterusnya digunakan sebagai rujukan untuk menyaring sebatian daripada pangkalan data Namiki Shoji. Melalui kaedah pencarian substruktur yang kedua, penambahan tiga kumpulan fragmen sebatian penyaringan dilakukan untuk mencari sebatian yang berpotensi daripada pangkalan data Namiki Shoji. Kumpulan fragmen telah dibahagikan kepada (i) fragmen F3/F22, (ii) fragmen F3/F23, F24, F25 dan F26, dan (iii) fragmen F22 sahaja. Pada langkah seterusnya, dua pusingan pendokkan molekul dan pengiraan tenaga $\Delta_{MM/GBSA}$ untuk setiap kumpulan fragmen dilakukan. Pada setiap kategori fragmen, 50 sebatian disenarai pendek sebelum pendokkan kali kedua dijalankan. Berdasarkan nilai tenaga pendokkan dan pengiraan tenaga $\Delta_{MM/GBSA}$, sembilan belas sebatian disenarai pendek dan diuji secara

in vitro. Dua belas sebatian menunjukkan nilai peratusan perencat melebihi 50 %. Ini juga membuktikan kejayaan penyaringan maya sebanyak 63 % melalui kaedah ini. Akhirnya, sebanyak dua sebatian menunjukkan nilai IC₅₀ lebih daripada 30 µM, enam sebatian merekodkan nilai perencat AChE di antara 1 hingga 10 µM dan tiga sebatian merekodkan nilai IC₅₀ di bawah 1 µM. Fragmen *N-benzylpipredine* dan *benzimidazole* yang terdapat pada sebatian, **CF19** (IC₅₀: 28 nM) adalah merupakan sebatian AChE yang paling aktif, diikuti oleh **CF15** (mengandungi fragmen *benzylpipredine* dan *acridine*) dengan nilai IC₅₀, 0.65 µM. **CF19** juga adalah sebatian yang mematuhi peraturan lima Lipinski dan peraturan tiga Jorgensen. Ini menunjukkan bahawa sebatian **CF19** adalah berpotensi untuk dijadikan sebagai ubat. Simulasi dinamik molekul selama 100 nanosaat juga dilakukan untuk memberi informasi tentang kestabilan dan dinamik sebatian **CF19** pada protin AChE berbanding dengan donepezil. Kajian simulasi dinamik sebatian **CF19** juga menunjukkan interaksi ikatan hydrogen yang stabil di antara CF19 dengan amino asid PHE295 dan TYR34 sebanyak 99 % dan 96 % di sepanjang simulasi. Penemuan kelas baru sebatian 1, 2, 4-triazoliltioetanon dan sebatian **CF19** ini adalah sangat menarik kerana boleh menjadi titik permulaan yang berguna untuk pengembangan terapi baru untuk merawat penyakit Alzheimer.

**HIERARCHICAL AND FRAGMENT BASED VIRTUAL SCREENING
APPROACHES IN THE DISCOVERY OF POTENT
ACETYLCHOLINESTERASE INHIBITOR**

ABSTRACT

Molecular modelling has been proved to accelerate and guide drug design by providing an understanding about biomolecular system at the atomic level. A molecular modelling application involving two virtual screening methods to facilitate the drug design for finding Acetylcholinesterase (AChE) inhibitors was carried out in this study. There are two virtual screening methods were presented in this study which are the hierarchical and fragment based virtual screening. Hierarchical virtual screening (HVS) method involved a shape similarity screening that combined with molecular docking calculation and visual inspection allowed the identification of 1, 2, 4-triazolylthioethanone as a novel class of AChE inhibitors. In the first step of HVS, shape similarity screening was used to find compounds that having similar shape to donepezil. About 920 shortlisted compounds was discovered of having similar shape as donepezil according on “ShapeTanimoto” scores. In the next step, molecular docking was performed and resulted about 73 shortlisted compounds which having the docking and Δ_{MM}/G_{BSA} energies values ranging from -13.35 to -5.35 kcal/mol and -30.72 to -86.26 kcal/mol, respectively. Further, all shortlisted compounds were visually inspected based on their shape and binding interactions with AChE binding pocket. About 58 shortlisted compounds were selected to be purchased and experimentally tested based on their binding to AChE similarly with donepezil. Thirteen compounds which having an elongated shape similar to donepezil consists of 1, 2, 4-triazolylthioethanone core and reported having the IC_{50} values in the range of

0.15 ± 0.07 to 3.32 ± 0.92 μM. The thirteen compounds consists of two aromatic rings at either end of their structures with 1, 2, 4-triazole ring forming the middle core of the compounds. Molecular docking shows the three substitutions of compounds, R1, R2 and R3 which predicted to occupy three different sites of AChE and interacts with residues TRP86 and TRP286 of AChE. The 1,2,4-triazole ring in compounds **1-13** is predicted to occupy a region close to hydrophobic aromatic residues of PHE341, TYR337, and PHE338 and was found to make π - π stacking contacts with TYR337. The thirteen compounds have not been reported as AChE inhibitors in ChEMBL database. In the next part of the thesis, a fragment-based virtual screening (FBVS) approach was applied to discover more potent AChE inhibitors. The virtual screening involves substructure search and molecular docking calculation was carried out which leading to the discovery of a nanomolar potent compound, **CF19** with an IC₅₀ value of 28 nM. By using substructure searching as the initial step, twenty-one fragments from the collective of AChE inhibitors with below than 1 μM inhibition activity in ChEMBL database were retrieved. Out of all, acridine fragment (F3) was found to be the most collective fragments (2530 compounds). The F3 fragment was further used as query to screen over Namiki Shoji compounds database which resulting the addition of three fragment groups of screening compounds over Namiki Shoji Database. The resulting hit from the three shortlisted groups are (i) fragment F3/F22 , (ii) fragment F3/F23, F24, F25 and F26, and (iii) fragment F22 only, were then shortlisted into the next step of screening. In the next step, two rounds of molecular docking and $\Delta_{MM/GBSA}$ calculation for each fragments' group was performed. At each fragments' categories, 50 hits compounds were retained after first docking step and subjected to the second round molecular docking. Based on the the docking and $\Delta_{MM/GBSA}$ energies value, nineteen compounds were shortlisted and evaluated *in vitro* AChE assay. Twelve of

these demonstrated good inhibition potency ($\geq 50\%$), thus giving a VS success rate of 63 %. It was finally discovered that two, six and three compounds showed IC_{50} values more than 30 μM , between 1 to 10 μM and below 1 μM , respectively. *N*-benzylpiperidine and benzimidazole containing compound, **CF19** (IC_{50} : 28 nM) is found to be the most potent AChE compound, followed by **CF15** (*N*-benzylpiperidine and acridine containing compound) with the IC_{50} value of 0.65 μM . Compound **CF19** also obeys Lipinski's rule of five with the exception of one and obeys Jorgensen rule of three, signifying the fact that compound **CF19** is potentially to be a drug-like orally compound. A 100 ns molecular dynamics simulation was also conducted to understand the stability and conformational dynamics of **CF19** on AChE protein as compared to donepezil. Dynamics simulation study of **CF19** has also demonstrated the stable H-bond interactions of 99 % and 96 % occurrence of simulation time with residues PHE295 and TYR341, respectively. The scaffold and **CF19** were particularly interesting as it could be a useful starting point for the development of novel therapeutics to tackle Alzheimer's diseases.

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

Currently, more than 46 million people worldwide are living with Alzheimer's Disease (AD) and the number is expected to increase to 131.5 million by 2050 (Alzheimer's Disease International, 2021). Researches have presented various hypothesis about AD and how these hypothesis are used to guide the treatment of AD. Among the various ways to combat AD is targeting Acetylcholinesterase (AChE).

There are four FDA approved drugs available in the market to treat AD i.e donepezil, tacrine, galantamine and rivastigmine (Nguyen et al., 2021). However, these approved drugs have limitations and known to cause side effects such as gastrointestinal upset, nausea, vomiting, diarrhoea and muscle cramps (Berk & Sabbagh, 2012). In fact, tacrine was withdrawn due to its hepatotoxicity. The recent approval of aducanumab by the United States Food and Drug Administration approved for AD treatment since 2003 reflect the extremely high failure rate of drugs developed for AD. Therefore, more research is needed to discover more candidates that can treat AD such that patients will be presented with the best treatment option available.

On the related note, AChE has a unique protein structure that contains a deep and narrow shape of 20 Å long elongated gorge with the width dimension of ~45 x 60 x 65 Å, which leading to the active site of AChE (Cheung et al., 2012). The structural detail showed that AChE inhibitors are mainly involved with the

interactions at the two binding sites of AChE, i.e. the peripheral anionic site (PAS) and cationic site (CAS). The structure also indicates the different binding mechanisms of the substrate and inhibitors (Dvir et al., 2010; Silman & Sussman, 2008). Due to their well-characterized active sites and comprehensive understanding of the structure–activity relationships of existing inhibitors with AChE, virtual screening (VS) method is particularly well-suited to be used for drug discovery.

Over the last 5 years (2015–2020), at least 15 studies of VS over various database have been implemented and proved to discover potent new cholinesterase inhibitors (Miles & Roses, 2021). However, the VS methods applied were the stepwise screening method either with SBVS or LBVS separately. This increases the computing time for SBVS method to screen over the whole database as well as limit the structural diversity of the hits by using LBVS only. Therefore, the current study aims to applied VS in combination of SBVS and LBVS techniques as well as implemented a fragment based virtual screening (FBVS) in the discovery of potent AChE inhibitors.

1.2 Objectives

The present research aimed to utilize computational approaches in discovering acetylcholinesterase inhibitors. Therefore, these objectives were pursued:

1. To discover a potent AChE inhibitor that having the similar shape as donepezil structure using a stepwise combination of ligand based and structure based virtual screening (LBVS and SBVS) in hierarchircal manner.
2. To find the most potent AChE inhibitor using a fragment based virtual screening (FBVS).

3. To investigate the binding, conformational changes and dynamics of AChE with the most potent compound using molecular dynamics simulation.

1.3 Content of Thesis

The present chapter provides the problem statement and objectives of the study. The focus of this thesis is to showcase the application of molecular modelling in drug discovery for the treatment of AD. Thus, Chapter 2 provides the literature review related to the Alzheimer's disease and molecular modelling. The methodologies that were used throughout the projects are presented in Chapter 3. Finally, general summary for the significant discovery found in this research is presented in the last chapter. Some recommendations to improvise the study are also suggested in the context of drug discovery.

CHAPTER 2

LITERATURE REVIEW

2.1 Alzheimer's Disease

Alzheimer's disease (AD) is defined as a slowly progressive neurodegenerative disease that is characterized by a decline in cognitive function in patients (Burns & Iliffe, 2009). It is estimated that more than 46 million people worldwide are living with Alzheimer's disease (AD) and the number is expected to increase to 131.5 million by 2050. The majority of AD patients are from low and middle income countries (Alzheimer's Disease International, 2021). In Malaysia, the prevalence of AD is estimated around 0.126% and 0.454% in 2020 and 2050, respectively (Tey et al., 2016). Management of AD patient is complex, therefore, the cost of managing AD is quite expensive and is estimated at US\$818 billion. It is forecasted that this cost is to increase to a trillion dollar by 2030 (Alzheimer's Disease International, 2021).

AD is named after a German neuropathologist and clinician, Dr. Alois Alzheimer in 1906 (Ramirez-Bermudez, 2012; Thies & Bleiler, 2013). Dr. Alzheimer noticed changes in the brain tissue of a woman, Aguste Deter, who had died of an unusual mental illness. Her symptoms included memory loss, language problems, and unpredictable behavior. After she died, Dr. Alzheimer investigated her brain and spinal cord tissues. A symptomatic description of cognitive failures, as well as descriptions and illustrations of senile plaques (SP) and neurofibrillary tangles (NFT), were included in Alzheimer's initial report on her. SP and NFT are now widely acknowledged as pathological markers of Alzheimer's disease by researchers (Ramirez-Bermudez, 2012; Thies & Bleiler, 2013).

Commonly, AD patients develop various symptoms that can be divided into three groups which are: (1) cognitive failures, (2) disruptions in psychiatry and behaviour, and (3) problems in doing regular activities (Burns & Iliffe, 2009; Thies & Bleiler, 2013). The cognitive failures are the loss of memory, having difficulties in language, and challenges in planning or solving problems. AD patients with these symptoms have difficulties to remember new information. Agitation, sadness, and hallucinations are among the second set of symptoms. These symptoms prevent patients from being able to perform tasks such as eating, bathing, or dressing. Patients with Alzheimer's disease were also unable to identify or communicate with family members at this point. Individual instances vary in their evolution from moderate to severe Alzheimer's disease, although it usually takes years (Burns & Iliffe, 2009).

2.1.1 Etiology

AD develops from multiple factors. Age is the primary risk factor, since the condition is most prevalent in those aged 65 and beyond (Burns & Iliffe, 2009; Thies & Bleiler, 2013). Additionally, cerebrovascular illness, diabetes, obesity, hypertension, dyslipidemia, depression, smoking, traumatic brain injury, genetics, and family history are the contributing factors of AD (Alzheimer's Disease International, 2021). At present, there is no cure for AD. Current pharmacology treatments for AD patients are only to improve the cognitive symptoms (James et al., 2020; Yiannopoulou et al., 2020). To date, the etiology of AD is not known. However, many hypotheses such as cholinergic dysfunction (Scarpini et al., 2003), amyloid- β ($A\beta$) deposits (Terry et al., 1964), τ -protein aggregation (Grundke-Iqbal et al., 1986; Wilson et al., 2013), oxidative stress (Wilson et al., 2013), neuroinflammation (Linker et al., 2011), excitotoxicity (Kaidery et al., 2013) calcium impairment (Diaz et al., 2009), mitochondrial dysfunction (Aliev et al., 2014) and bacterial infection (Domini et al.,

2019) have been suggested and used to explain the pathogenesis of AD. Exploration of these hypotheses has led to the identification of several proteins that influence the generation and exacerbation of the disease. These include acetylcholine and butyrylcholine esterases (AChE and BuChE) which are involved in cholinergic hypothesis (Davies & Maloney, 1978), BACE-1 (amyloid hypothesis) (Hardy & Higgins, 1992), PDEs (non-amyloid hypothesis) (Cummings et al., 2019), GSK-3b (tau protein hypothesis) (Zhang et al., 2019; Panza et al., 2016), MAOs (monoaminergic hypothesis) (Panza et al., 2016; Massart et al., 2012), NMDA receptor (glutamatergic hypothesis) (Shimizu et al., 2000; Johnston et al., 1968), 5-HT receptors (serotonergic hypothesis) (Shimizu et al., 2000; Rodriguez et al., 2012) and H3 receptor (histaminergic system) (Panula et al., 2015). Targeting AChE involved in cholinergic hypothesis is one of many approaches to impair cholinergic function of AD brain (Weng et al., 2003).

2.1.2 Causes of AD

Of the various hypothesis, cholinergic dysfunction and amyloid-beta ($\alpha\beta$) accumulation are believed to be the main causes of AD. The cholinergic hypothesis was initially postulated in 1982, when researchers discovered that AD patients' brains lacked activity at cholinergic neurons (Davies & Maloney, 1976; Perry et al., 1981). According to cholinergic hypothesis, the cognitive deficits of AD are related to the decline of neurotransmitter acetylcholine (Terry & Buccafusco, 2003). Although the cholinergic theory has been addressed with considerable doubt, it remains critical for understanding AD today. Most of treatments for AD has been discovered aimed at restoring the cholinergic deficit.

2.1.2(a) Cholinergic Hypothesis

In the brain, Acetylcholine (ACh) acts as a neurotransmitter and involved in many physiological processes. There are sensory information, memory, learning, focus and many critical functions (Breijyeh & Karaman, 2020). Two enzymes are involved in the regulation of ACh levels in the brain: choline acetyltransferase (ChAT) and acetylcholinesterase (AChE). While ChAT catalyses the production of ACh from choline and acetyl-CoA, AChE catalyses ACh's hydrolysis to choline and acetate. ChAT enzyme levels have been shown to be reduced by up to 90% in patients with severe AD when compared to normal (Giacobini, 2003). AChE on the other hand acts to terminate synaptic transmission which was mediated by ACh. These are the two factors that cause ACh degeneration in AD patients (Falco et al., 2016; Green et al., 2005).

2.1.2(a)(i) Acetylcholinesterase (AChE)

Acetylcholinesterase (AChE, acetylcholine acetylhydrolase, E.C. 3.1.1.7) is a serine hydrolase that belongs to the cholinesterase family of proteins (Sussman et al., 1991). AChE is discovered in many types of conducting tissue. There are nerve and muscle, central and peripheral tissues, motor and sensory fibers, and cholinergic and noncholinergic fibers. The molecular weight of AChE is 76 kD and is made up of a large central β -sheets surrounded by 14 α -helices (Weinstock, 1999). In cholinergic hypothesis, it is an enzyme that involved in the breakdown of the neurotransmitter acetylcholine (ACh) into acetate acid and choline (McHardy et al., 2017). As AChE catalyzes the breakdown of acetylcholine, inhibition of AChE activity has been reported

to increase synaptic acetylcholine levels and improve cholinergic function in the brain (Bartus et al., 1982; Tabet, 2006).

2.1.2(a) (ii) Structural and Functional Sites

The first AChE crystal structure was purified from *Torpedo californica* (TcAChE). It was reported in 1991 (Sussman et al., 1991). The crystal structure of AChE from human is later successfully purified by Kryger et al. in 1999. Structurally, AChE contains a long and narrow active-site gorge, about 5 Å wide and 20 Å deep, that extends from the surface of the enzyme down to the catalytic site. The gorge has two distinct ligand binding sites which are the anionic sub-site of the catalytic site (CAS), and the peripheral anionic site (PAS). The CAS site which contains a catalytic triad (H440-E334-S203) and a key aromatic residue, W86 is posited close the bottom of the gorge. The PAS which consists among others of W286 residue is located close to the entrance of the active-site gorge. In the middle of the gorge, a bottleneck formed mainly by the side-chains of F330 and Y124. The middle of the gorge is so narrow and only admits passage of a water molecule. Generally, this intrinsic property is shared by AChEs from various species.

The active-site AChE gorge has two additional remarkable features (Sussman et al., 1991; Colletier et al., 2006; Kryger et al., 1999). One is that the walls and base of the gorge are comprised of aromatic residues. The other characteristic is a substantial electrostatic dipole that is properly oriented along the gorge axis and is generated by the overall distribution of charge residues across the catalytic region. This properties attract positively charged ligands to penetrate into the active site. Research has suggested that these two features may work together (Ripoll et al., 1993). The dipole acts as a propellant to pull ligands by electrostatic interactions, while the many

aromatic groups in the residues act as a cascade of low-affinity binding sites for positively charged ligands. Schematic representation of AChE binding sites is shown in Figure 2.1.

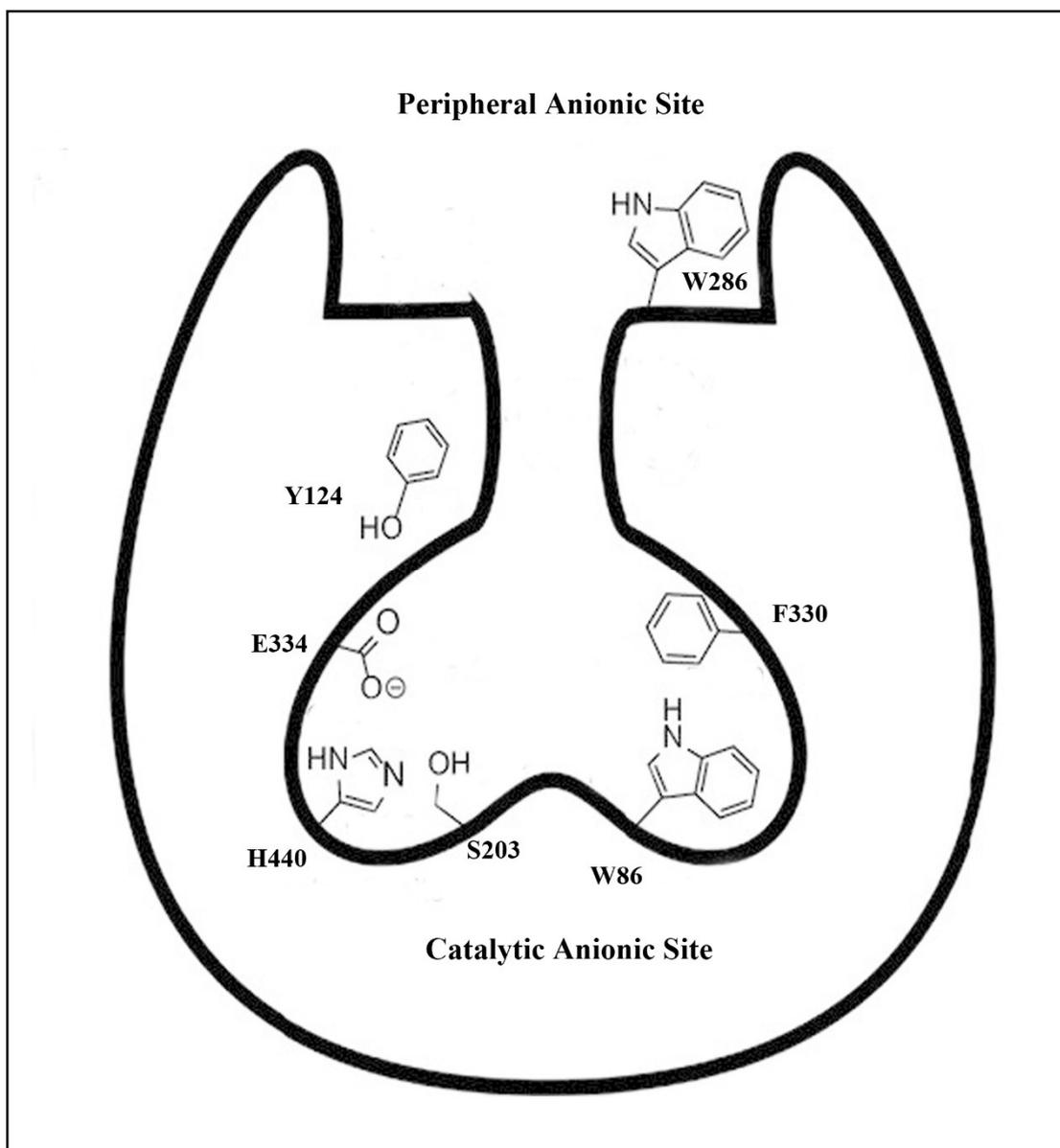


Figure 2.1 Schematic representation of AChE binding sites.

2.1.2(a) (iii) Function and Mechanism

Acetylcholinesterase (AChE) is the enzyme that degrades acetylcholine in the synaptic cleft at the neuromuscular junction and cholinergic synapses in the central

nervous system. The events occurring at the synapse are shown in the Figure 2.2. Initially, a nerve impulse from a pre-synaptic neuron will cause the release of the neurotransmitter, acetylcholine (ACh) at the synaptic cleft. Acetylcholinesterase at the synaptic cleft hydrolyzes and breaks down the neurotransmitter ACh. The breakdown product choline is then absorbed by the pre-synaptic neuron to re-synthesize more neurotransmitter. When a nerve impulse reaches a cholinergic synapse, vesicles on the presynaptic neuron undergo exocytosis thereby releasing ACh into the synaptic cleft. On the postsynaptic cholinergic neuron, ACh binds to AChRs (ACh receptors, some of which may be ligand-gated ion channels). This will cause the ion channel to open and allowing sodium ions to pass. This results in a depolarization of the postsynaptic neuron and the continuation of the nerve impulse. Also, AChE is present on the postsynaptic neuron, anchored to the cell membrane via a covalently attached lipophilic moiety (i.e. phosphatidylinositol). AChEs scavenge and hydrolyze the ACh from cholinergic synapses after neurotransmission (Colovic et al., 2013).

AChE initiates a nucleophilic attack on the ACh carbonyl carbon, thereby acylating the enzyme and releasing four choline as depicted in Figure 2.3. This is followed by the hydrolysis of the acyl-enzyme yielding acetic acid and a regenerated enzyme (Turgeon, 1998). AChEs are well known to be extremely efficient enzymes (Wiener, 2004), for which the speed with which AChE hydrolyzes ACh is well suited to the known function of the enzyme.

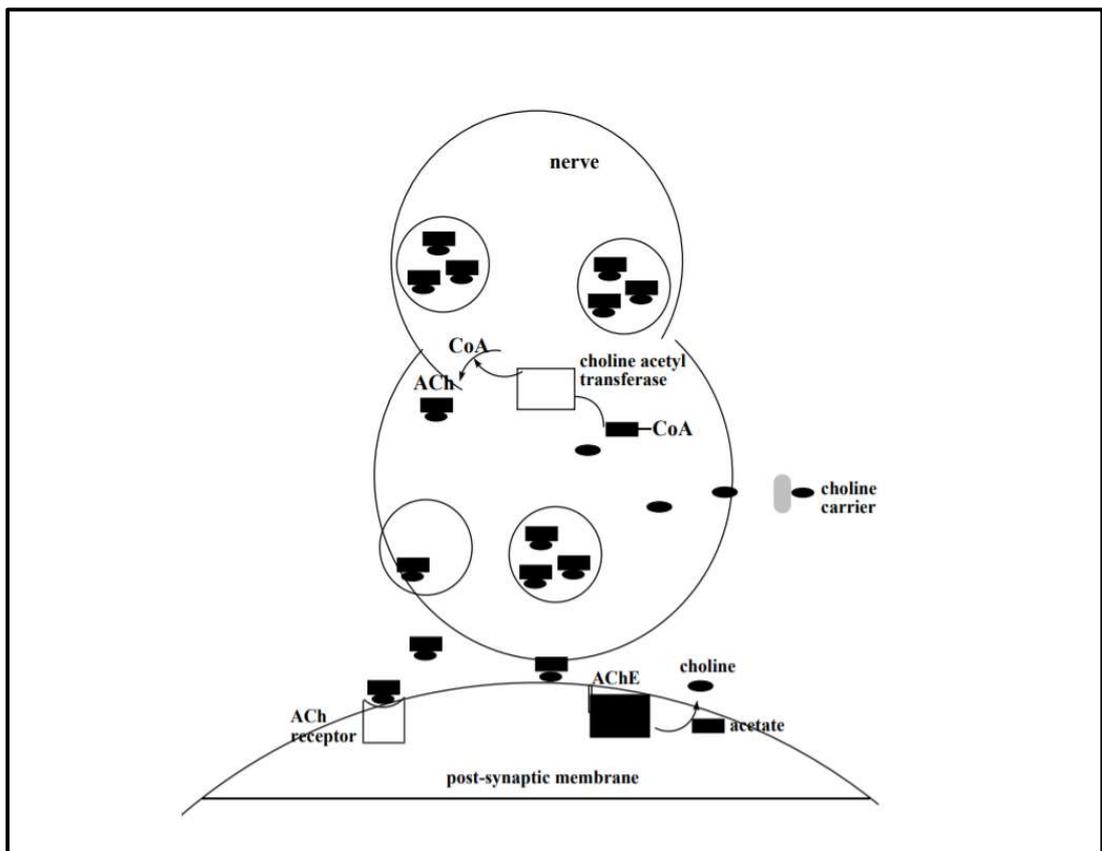


Figure 2.2 Mechanism of AChE mechanism in neurotransmission.

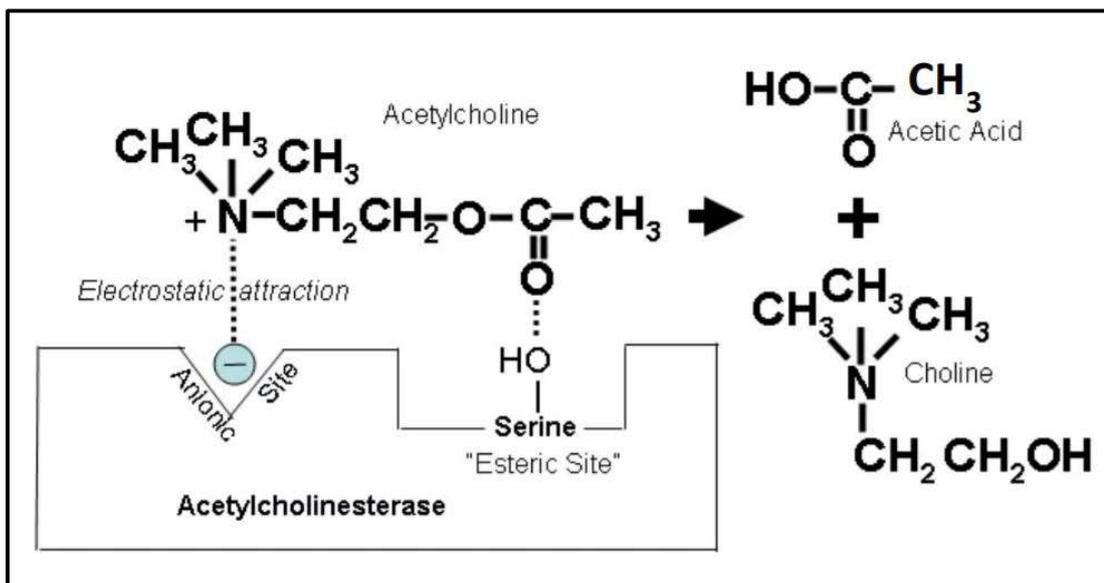


Figure 2.3 Schematic diagram of substrate binding to the enzyme AChE at the active site of the enzyme (Wiener, 2004).

2.1.3 Current Drugs Targeting Cholinesterase for AD Treatment

For more than two decades, the development of potent acetylcholinesterase (AChE) inhibitors has been an ongoing task to treat Alzheimer's diseases. The first cholinesterase inhibitor licensed was tacrine. Tacrine was the first molecule to enter clinical trials for AD treatment and demonstrated positive effects on memory function in young and aged patients. The research on tacrine started in 1984, and the drug was approved by the US Food and Drug Administration (FDA) for AD treatment and related dementias in 1993. However, tacrine was then withdrawn due to its hepatotoxicity (Watkins et al., 1994).

In 1970, physostigmine was developed. Some studies demonstrated that this drug provided temporarily moderate relieve in symptoms of AD patients. Physostigmine has also proved to stabilize the decline of cognitive function and functional ability (van Dyck et al., 2000). Donepezil on the other hand is a new AChE

inhibitor which was the second drug approved by FDA in 1996. It is structurally different from other cholinesterase inhibitors. It was derived from indanone group of compounds and was developed in 1983 by Sugimoto and co-workers at the Eisai Research Laboratory in Japan (Sugimoto et al., 2002). It is also a selective reversible inhibitor of AChE acting centrally by increasing the bioavailability of acetylcholine at the synaptic cleft (Jennifer & Simon, 2016).

Rivastigmine, known as Exelon® is approved by FDA in 2000. Rivastigmine is a pseudo irreversible carbamate-selective inhibitor of AChE and BuChE for the treatment of mild-to moderate AD. It was also used for the treatment of mild-moderate Parkinson's dementia (Grossberg & Desai, 2003). In 2006, a selective reversible inhibitor of AChE and allosteric modulator of nicotinic cholinergic receptors known as galantamine was introduced (Loy & Schneider, 2006).

Many research have also highlighted a potent, reversible, and selective inhibitor of AChE and NMDA receptor antagonist such as Huperzine A. Huperzine A is a new alkaloid derived from the Chinese herb *Huperzia serrata* (Yang et al., 2013). Until now, Huperzine A has not yet been recommended for clinical use despite of many clinical trials have shown the potential effect of this compound (Li et al., 2008). Four approved drugs targeting cholinesterase to treat AD are shown in Figure 2.4.

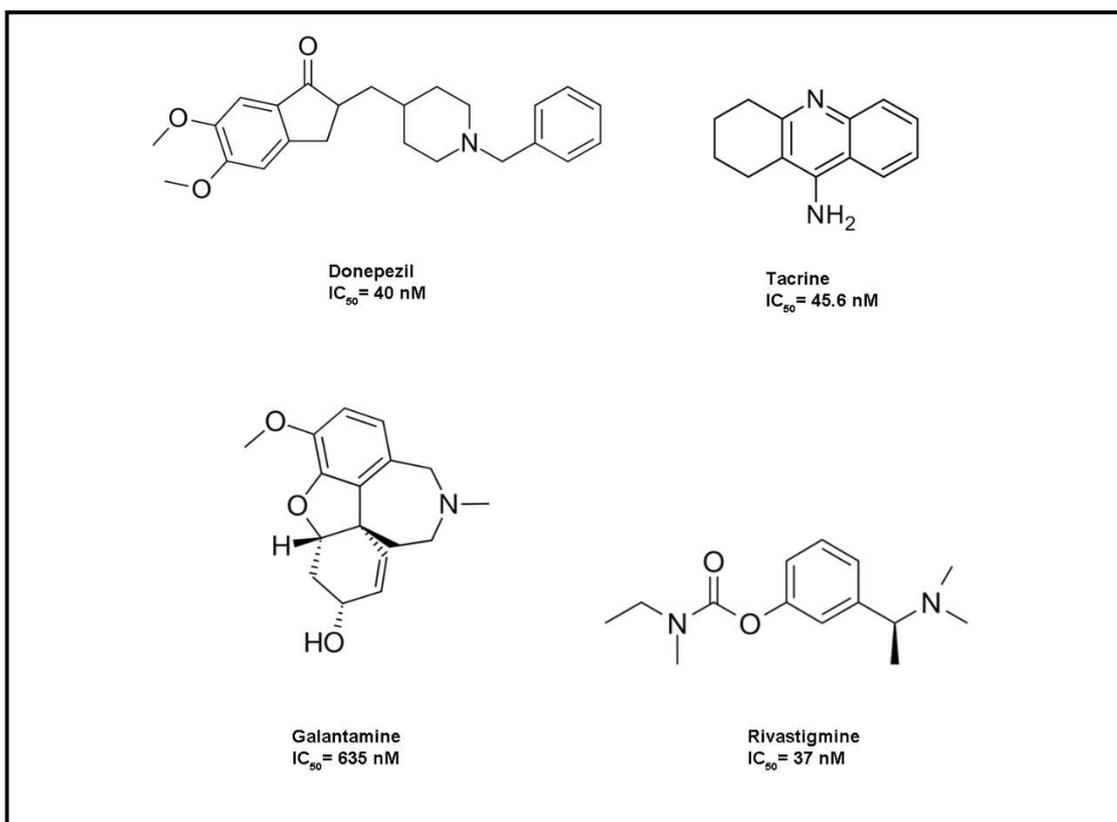


Figure 2.4 AChE inhibiting drugs approved for the treatment of Alzheimer's disease.

2.2 Molecular Modelling

Molecular modelling has been called the fourth axis of chemistry where it lies somewhere between theory, observation, and experiment. The core of molecular modelling is to describe the state and behaviour of molecules through computer simulations. Molecular modelling has been applied to various research studies, including drug design, computational biology, nanostructures, and material science (Pimentel et al., 2013). The advancement of molecular modelling in drug discovery is due to these main reasons: explosive growth in the available structures of proteins and ligands in the database extracted from X-ray crystallography, nuclear magnetic resonance (NMR) and electron microscopy studies, and secondly, substantial advances in methodology and software, as well as the availability of massive supercomputers (Saxena et al., 2009). In the late 90s, molecular modelling has been introduced as an

application of a "*computational chemistry*" method where International Union of Pure and Applied Chemistry (IUPAC) describe it as a molecular element of study made practicable via the use of computers (Wermuth et al., 1998). Later, IUPAC defined molecular modelling as the study of properties and structure of molecule using computational chemistry and graphical visualisation techniques with the goal of providing a three-dimensional representation under a given set of conditions (Barreiro et al., 2002). Ever since, the growth in software and hardware, along with a continued decline in expenditure, have made molecular modelling one of the most promising fields of the twenty-first century.

Briefly, molecular modelling relies on two branches of theories; one is based on quantum mechanics (Griffiths, 2005), and the other is based on classical mechanics (Burkert & Allinger, 1982). Any sufficiently small molecules must be described by the laws of quantum mechanics. Nonetheless, under the right conditions, it is still beneficial to approximate the molecule using classical mechanics. This approach is called the "molecular mechanics" (MM) or "force-field" method (Burkert & Allinger, 1982). All molecular mechanics methods are *empirical* where the parameters in the model are obtained by fitting to known experimental data. Considering these two central cores of molecular modelling, molecular mechanics and quantum mechanics are briefly introduced below.

2.2.1 Quantum Mechanical (QM) Methods

In general, quantum mechanics methods are aimed at solving Schrodinger equation which represents the electrons in energy calculation (Szabo & Ostlund, 1989). To describe the state of a system in quantum mechanics, it was postulated the

existence of a coordinate function called a molecular wave function, which is the solution of Schrodinger equation as below;

$$H\Psi = E\Psi \quad 2.1$$

where H is the Hamiltonian, a mathematical expression of the energy terms of the molecule comprising the potential to kinetic energy of the electron system is particles and Ψ molecular wave function described in terms of spatial coordinates of the particles constituting the system in a certain state (Foresman & Frisch, 2015).

There are several computational methods used in the quantum mechanics calculation, among them are the density functional theory, *ab initio* and semi-empirical calculation (Johansson et al., 2013). *Ab initio* calculations implies an approach from fundamental physical constants which includes a summation electronic population of molecules. This computation is not based on empirical data. This category includes Hartree-Fock (HF), configuration interaction (CI), many-body perturbation theory (MBPT), coupled-cluster (CC) theory, and other approaches (Szabo & Ostlund, 1989). These techniques, most notably Hartree-Fock (HF), utilise the whole Schrodinger equation to handle all electrons in a chemical system (Barreiro et al., 2002) using sets of basic functions. Additionally, partly polarised basis functions have been devised, such as 3-21G*, which is the same minimal 3-21G basis function with partial polarisation (Leach, 2001). *Ab initio* method is more accurate; however, it is time consuming and incurs high computation cost. In contrast to *ab initio* procedures, semi-empirical methods involve assessing empirical parameters, such as those derived from experimental data such as geometry of equilibrium, heat of formation, molecular dipole moment and ionisation potentials. (Barreiro et al., 2002; Hehre, 2003). The most often used semi-empirical methods are AM1 (Austin Model

1) (Dewar et al., 1985) and PM3 (Parametric Method 3) (Stewart, 1989). Both methods employ fairly similar concepts but differ in their parameterization.

Density Functional Theory (DFT) is the most often employed approach in drug design (Szabo & Ostlund, 1989). The primary goal is to substitute the wave function for the electron density in methods such as the Hartree-Fock. The HF calculations take an average electron density into account, whereas the DFT calculations take into account the instant interactions between pairs of electrons with opposing spins (Foresman & Frisch, 1996). It is a method based on Hohenberg and Kohn's hypothesis, according to which all properties of a system are functions of the charge density. The density functional's molecular orbital computations are often expressed as a linear expansion of atomic orbitals (basis functions), which can be expressed as Gaussian type functions, Slater orbitals, or orbital numeric functions (Leach, 2001). These models are also applicable to molecules with 50–100 atoms (Hehre, 2003). Because the precise function is unknown, a diverse variety of available functions can produce different outcomes for the same issue. B3LYP (Becke, Lee, Yang, and Parr) is a frequently used hybrid approach in which a portion of the function is derived using quantum mechanics (HF combines energy exchange and DFT exchange term) and the remaining parts is parameterized (adds functional correlation) (Foresman & Frisch, 1996).

A frontier molecular orbital (FMO) theory is introduced by Kenichi Fukui in 1952. He published paper in the *Journal of Chemical Physics* entitled, “A molecular theory of reactivity in aromatic hydrocarbons (Fukui et al., 1952). He later shared the Nobel Prize in Chemistry with Roald Hoffman for his work on reaction mechanism. The FMO theory was initially used to explain the electrophilic substitution in naphthalene, but it became gradually clear that the scope of this theory is much

broader. For instance, the concept of frontier orbital symmetries was successfully applied to rationalize the outcomes of cycloaddition reactions and other pericyclic reactions (Houk, 1975). This theory led to the important aspect of the theory which are the effects of the highest occupied and lowest unoccupied molecular orbitals (HOMO and LUMO) on reaction mechanisms. According to the theory, the perturbation energy of interaction between two molecules mainly comes from the interaction between HOMO and LUMO. The energy of the HOMO is linked to the ionization potential and characterizes the tendency of the molecule toward attack by electrophiles, while the energy of the LUMO is linked to the electron affinity and characterizes the susceptibility of the molecule toward attack by nucleophiles. A high HOMO level also represents a good nucleophile, alternatively, a lower LUMO level represents a good electrophilic compound. If the energy of one molecular is different from another molecule, the effect of the interaction gives an energy split. As a result, the smaller the energy difference between the two orbitals, the stronger the interaction (HOMO-LUMO gap) (Heifetz, 2020).

In the drug design (protein-ligand), the reaction between protein and ligand can only happen between the lowest unoccupied molecular orbitals (LUMOs) of a protein and the highest occupied molecular orbital (HOMO) of its ligand. The second rule of the protein-ligand interactions is that only those residues located in both the LUMOs of a protein and a surface pocket of a protein are active residues of the protein and the corresponding pocket is the ligand binding site (Pang et al., 2008). Other definitions relating LUMO and HOMO energies are the hardness and softness. Hard nucleophiles have a low-energy HOMO; soft nucleophiles have a high-energy HOMO; similarly, hard electrophiles have a high-energy LUMO; and soft electrophiles have a low-energy LUMO (Heifetz, 2020).

2.2.2 Molecular Mechanical (MM) Methods

Molecular mechanics (MM) method can be applied if the systems are too large to be effectively treated using quantum mechanical methods. MM method can easily handle several hundred thousand atoms, and in case of a coarse-grained approach, several million atoms (Genheden et al., 2018). Compared with quantum mechanics, molecular mechanics ignores electrons in the energy calculation (Leach, 2001).

Molecular mechanics (MM) (Tsibouklis, 1998) is based on a mathematical model of a molecule that consider atoms as sphere and bonds as springs. The concept behind MM is to express the energy of a molecule as a function of its resistance toward the ability of bonds to stretch, bend and twist as shown in Figure 2.5. The non-bonded atoms on the other hand, interact through van der Waals attraction, steric repulsion, and electrostatic attraction/repulsion. These properties are easiest to explain mathematically when atoms are considered as spheres of characteristic radii. This energy equation is also used to achieve the potential energy surface minima (Hinchliffe, 2008).

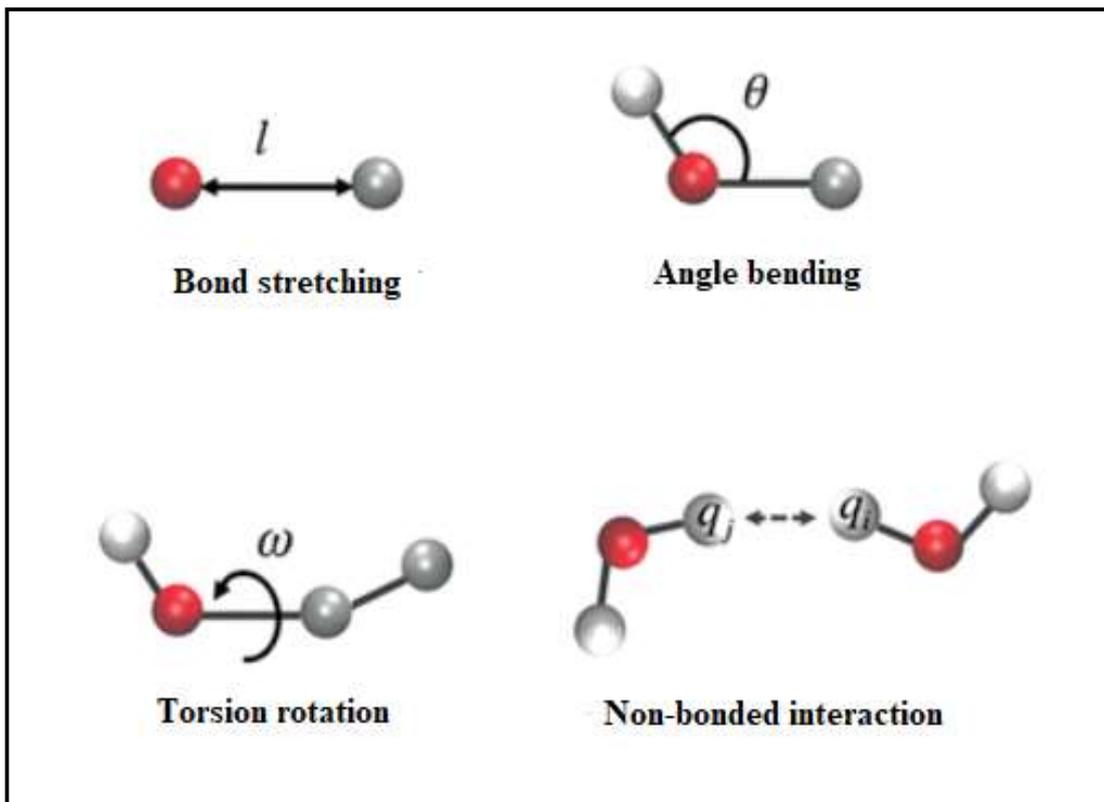


Figure 2.5 Representation of bonded and non-bonded interactions used in molecular mechanics.

In MM, the potential energy is calculated by the summation of the energy terms that describe interactions between bonded atoms which are bonds, angles, and torsions. The calculation also include the non-bonded interactions, such as van der Waals and electrostatic interactions using the following equation,

$$\begin{aligned}
 V(r^N) = & \sum_{\text{all bonds}} k_1(l - l_0)^2 + \sum_{\text{all angles}} k_\theta(\theta - \theta_0)^2 & 2.2 \\
 & + \sum_{\text{all torsions}} \frac{1}{2} V_n[1 + \cos(n\omega - \gamma)] \\
 & + \sum_{j=1}^N \sum_{i=j+1}^N \left\{ \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^6 \right] + \frac{q_j q_i}{4\pi\epsilon_0 r_{ij}} \right\}
 \end{aligned}$$

where the bonded terms represent the stretching of bonds (l), bending of valence angles (θ) and rotation of torsional angles (ω). Three force constants: k_l , k_θ and V_n characterize the energetic cost relative to the equilibrium value, needed to increase the value of a bond length (l_0), angle (θ_0) or rotation around a torsion angle. The torsion term represents a periodic rotation of a dihedral angle with periodicity n and phase γ . The non-bonded energy is the sum of repulsion, attraction, and electrostatics between non-bonded atoms. The parameter ϵ_{ij} is related to the well-depth of Lennard-Jones (LJ) potential, r_{0ij} is the distance at which the LJ potential has its minimum. q_i is the partial atomic charge, ϵ_0 is the vacuum permittivity, and r_{ij} is the distance between atom i and atom j . The Lennard Jones and Coulomb potentials describe the short-range non-bonded interactions (Ivanov, 1996; Wang et al., 2020). Ewald summation and particle mesh ewald (PME) method were introduced for long range electrostatic calculation (Sagui & Darden, 1999). The calculations of potential energy via Eq. 2.1 are used to search for local energy minima, to construct and analyze multidimensional potential energy surfaces (PES), to follow trajectory of movement (in MD, molecular dynamics simulations), or to study averaged thermodynamic and geometry characteristics (via MC, Monte Carlo sampling) of the systems (Leszczynski et al., 2017).

2.2.3 Hybrid Quantum Mechanical/Molecular Mechanical (QM/MM) Methods

Molecular mechanics methods have become a primary tool for computational biochemistry nowadays. It allows the modelling of enormous molecules, such as protein and DNA, but it is simplified and lacks inaccuracy. Quantum mechanics approaches aim to solve Schrodinger equations and be very efficient; however, it is also very computationally expensive. As computers powers continue to improve, the

optimal method would be one that combined the speed of molecular mechanics and quantum mechanics' accuracy by developing QM/MM in solving many biological systems (Sabin & Canuto, 2010). A massive step towards this was achieved in 1976 when two scientists, Arieh Warshel and Michael Levitt, published a paper about enzymatic reactions that outlined a new, powerful hybrid tool of QM/MM (Warshel & Levitt, 1976). Warshel and Levitt were able to simulate a much larger system in a computationally effective way. This was also a groundbreaking achievement in molecular modelling as they were awarded the Nobel prize in chemistry in 2013 along with Martin Karplus (Fersht, 2013).

Hybrid QM/MM simulations can be implemented to a large system such as protein ligand system by portioning the target problem into two parts. One part of the protein/enzyme which are involved in a chemical reaction (catalysis) can be treated with QM level simulation. The remaining part of the enzyme which encompasses a much larger number of atoms is simulated using MM. QM/MM methodologies are different in different ways; (1) the type of scheme used to calculate the QM/MM energy; (2) the different boundary regions chosen; (3) how the interaction between the QM and MM region is investigated; (4) how an appropriate computational method is selected; and (5) how the enzymatic reaction and the associated conformational flexibility are tackled (Aminpour et al., 2019). There are advantages and disadvantages of QM/MM methods. The most common QM/MM methods are Car–Parrinello/Molecular Mechanics MD (Car & Parrinello, 1985), empirical valence bond (EVB) Method (Warshel & Levitt, 1976), the cluster model (Sousa et al., 2012), and QM/MM MD methods.

2.3 Computer Aided Drug Design and Virtual Screening

The process of drug discovery and development is challenging, time-consuming, and expensive (Leelananda & Lindert, 2016). Traditionally, drug discovery approaches relied on stepwise synthesis and screening of large number of compounds to identify the correct solution for a particular disease. Nowadays, scientists are combining computer-aided drug design efforts as an initial approach in combining chemical and biological space to streamline drug discovery, design, and optimization. It is estimated that the time and cost of currently bringing a new drug to market from drug discovery vary, but 7-12 years and tens to hundreds of millions of U.S dollars are often cited (Sertkaya et al., 2016). In 2020, it was reported that the median cost of getting new drug into the market was \$985 million, and the average cost was \$1.3 billion (Wouters et al., 2020). It has been estimated that in general five out of 40,000 compounds tested in animals reach human testing, and only one of five compounds reaching clinical studies is approved (Wouters et al., 2020).

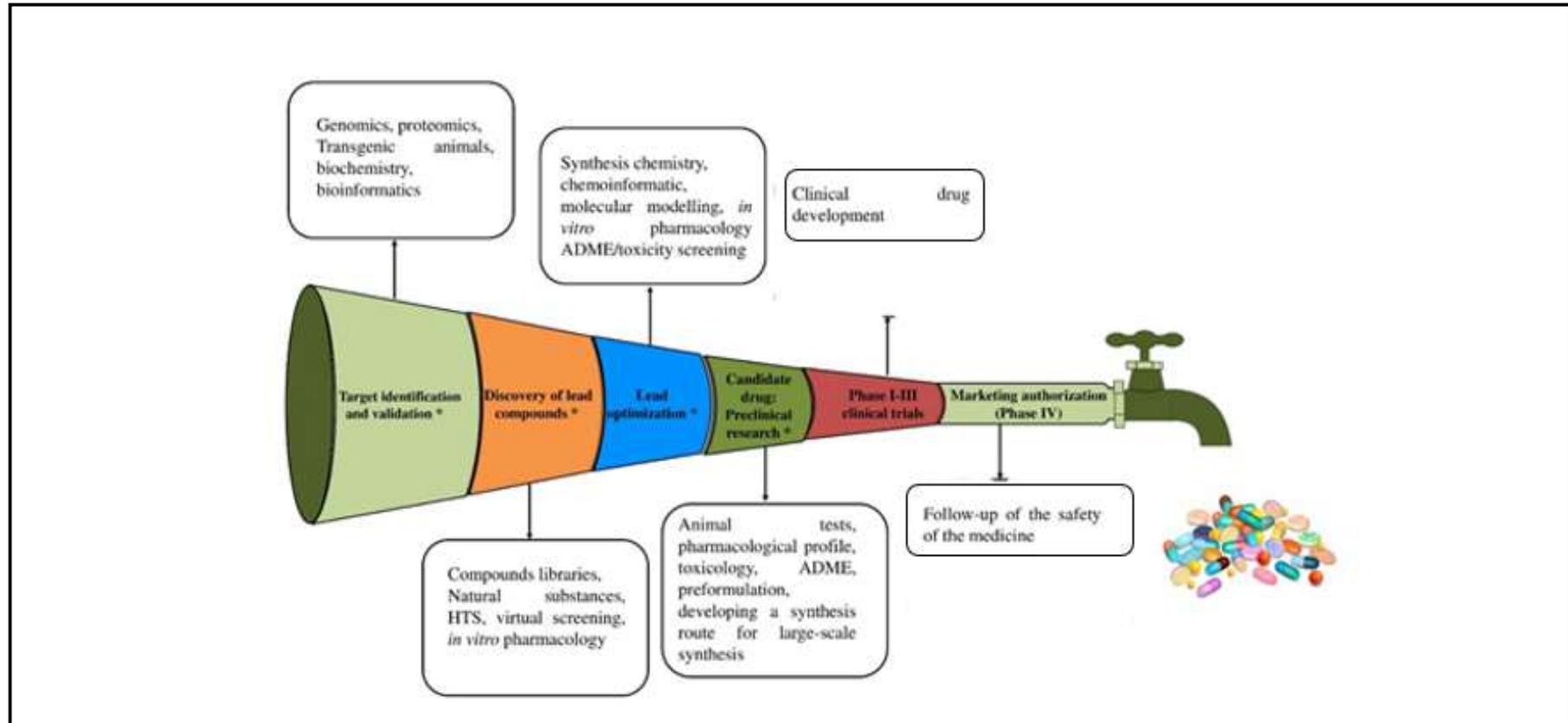


Figure 2.6 Modern drug design development. Each step contains characteristic actions, methods, and tools used. Computer-aided drug design can be applied to facilitate drug discovery in the phases denoted with a star (*).