

**INTERACTION OF ISONIAZID WITH *MYCOBACTERIUM TUBERCULOSIS*
ENOYL-ACYL CARRIER PROTEIN REDUCTASE (INHA):
FROM MOLECULAR PERSPECTIVES**

by

CHOONG YEE SIEW

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

June 2008

ACKNOWLEDGEMENTS

I would like to express my truthful appreciation and gratitude to my family who guided me along the correct path of education with much sacrifice and patience. I thank them for their unconditional love and continuous moral support.

I am most grateful to Assoc. Prof. Dr. Habibah Wahab, Assoc. Prof. Dr. Pazilah Ibrahim and Assoc. Prof. Dr. Amirin Sadikun, whom I am fortunate to have as my supervisors, for their excellent guidance, encouragement and persistence for making this thesis a reality. Their keen advice not only sparked my interest in the field of molecular modelling but also continuously providing great insight and motivation for me as well.

I am also thankful to MOSTI for NSF, NBBnet for the work stations, PPKT and PhDs Laboratory for the working space; and Assoc. Prof. Dr. Shukri Sulaiman and Assoc. Prof. Dr. Mohamed Ismail Mohamed Ibrahim for the technical support and advices.

I would like to express my warmest gratitude to my past and present laboratory-mates: Achsan, Belal, Ezatul, Dr. Hassan, Imtiaz, Nani, Nurul, Rashid, Sek Peng, Sohail, Suhaini, Sy Bing and Wai Keat for creating sweet memories to cherish and making life more fun and interesting in PhDs Laboratory. Not forgetting Adli, Fitriah and Suhaini for their kind technical assistance. Special thanks to Assoc. Prof. Dr. Tan Soo Choon for providing learning experience in VFL and also to my wonderful and faithful lunch-mate, Theam Soon. I offer my best wishes in their future undertakings.

To the most special and important person in my life, my husband, Peng Lee, thank you for your enthusiasm, encouragement and tolerance with me all these years, no words can ever express my gratefulness and love for you. I hope you would find forgiveness for me for having to compromise the time I should be spending with you for this project.

To all of you, I thank you.

	Page
ACKNOWLEDGEMENTS	ii
TABLE OF CONTENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	viii
LIST OF ABBREVIATION	xiii
LIST OF APPENDICES	xiv
LIST OF PUBLICATIONS	xv
ABSTRAK	xvi
ABSTRACT	xviii
CHAPTER ONE: GENERAL OVERVIEW	1
1.1 Statement of the problem	1
1.2 Tuberculosis (TB)	2
1.3 The importance of molecular modelling in TB research	4
1.4 Molecular modelling	5
1.5 The fundamental of molecular modelling	7
1.5.1 Energy minimization	7
1.5.1.1 Minimization algorithms	9
1.5.2 Molecular mechanics force field	11
1.5.3 Principles of MD simulations	17
1.5.3.1 MD simulation ensembles	18
1.5.3.2 Boundaries and cutoff	19
1.5.3.3 Free energy calculation in MD simulation	21
1.5.4 Molecular docking	23
1.5.4.1 Free energy calculation in molecular docking (AutoDock)	25
1.5.4.2 Docking problem	28
1.5.4.3 Lamarckian Genetic Algorithm in AutoDock docking	28
1.5.4.4 Ligand flexibility in AutoDock docking	29
1.6 Application of molecular modelling simulations	29
1.6.1 Application of MD simulation	29
1.6.2 Application of molecular docking	33
1.6.3 Limitations/challenges of molecular modelling methods	36
1.7 Content of the thesis	38

CHAPTER TWO: ELUCIDATING ISONIAZID RESISTANCE	39
2.1 Isoniazid (INH)	39
2.1.1 INH and mycolic acid biosynthesis	39
2.1.2 InhA enzyme	40
2.2 Action mechanism of INH	42
2.3 INH resistance	46
2.3.1 INH resistance and <i>katG</i> gene mutation	47
2.3.2 INH resistance and <i>inhA</i> gene mutation	48
2.4 Specific aim/objective	50
2.5 Methodology	51
2.5.1 Molecular docking	51
2.5.1.1 Overview of the method	51
2.5.1.2 Protein setup	51
2.5.1.3 Calculation of grid map	54
2.5.1.4 Ligand setup	54
2.5.1.5 Docking simulation	57
2.5.1.6 Data analysis	57
2.5.2 Molecular dynamics (MD) simulation	59
2.5.2.1 Overview of the method	59
2.5.2.2 System setup	59
2.5.2.3 Minimization	61
2.5.2.4 MD simulation	64
2.5.2.5 System equilibration and production	64
2.5.2.6 Binding free energies calculation in MD simulation	64
2.6 Results	65
2.7 Discussion	125
2.7.1 The interactions of INADH with WT InhA	150
2.7.1.1 Mechanism of INH resistance in S94A InhA	153
2.7.1.2 Mechanism of INH resistance in I16T InhA	155
2.7.1.3 Mechanism of INH resistance in I21T InhA	157
2.7.1.4 Mechanism of INH resistance in I21V InhA	159
2.7.1.5 Mechanism of INH resistance in I47T InhA	161
2.7.1.6 Mechanism of INH resistance in V78A InhA	163
2.7.1.7 Mechanism of INH resistance in I95P InhA	165
2.8 Conclusions	167

CHAPTER THREE: INVESTIGATION OF THE BINDING OF ISONIAZID DERIVATIVES ON INHA	170
3.1 The problem of tuberculosis (TB) therapy	170
3.2 The mycobacteria envelope	171
3.2.1 The mycobacterial cell wall	171
3.2.1.1 Mycobacterial mycolic acids	173
3.2.1.2 Mycobacterial porin	173
3.3 Multi-drug resistant TB	174
3.4 Current status of drug discovery research in TB	176
3.5 Specific aim/objective	177
3.6 Methodology	178
3.6.1 Overview of the method	178
3.6.2 Protein setup and calculation of grid map	178
3.6.3 Ligand setup	178
3.6.4 Docking simulations and data analysis	178
3.7 Results	180
3.8 Discussion	190
3.9 Conclusions	195
CHAPTER FOUR: GENERAL CONCLUSIONS	197
4.1 Summary of the present study	197
4.2 Possible future studies	201
4.3 Concluding remarks	202
REFERENCES	203
APPENDICES	232

LIST OF TABLES

Page

Table 2.1	The RMSD, energies and inhibition constants of ligands (INH, NADH and INADH) in WT and S94A InhA from docking simulation.	66
Table 2.2	Summary of hydrogen bonds formed by INH, NADH and INADH in WT and S94A InhA.	68
Table 2.3	Summary of hydrophobic, van der Waals, electrostatic and π - π interactions between (a) INH, (b) NADH, and (c) INADH with WT and S94A InhA.	74
Table 2.4	The estimated INH and INADH binding free energies from MD simulations using LIE method.	78
Table 2.5	The average RMS structural fluctuation, the average and last (at 1000 ps) MD structures of INADH, WT and MT InhA backbone after 200 ps of energy equilibration in MD simulation compared to the x-ray crystal structure.	82
Table 2.6	Cluster analysis for INADH-InhA and INADH in complex with InhA.	91
Table 2.7	The distance of INADH from residues S94, I16, I21, I47, V78 and I95 of WT InhA compared with the corresponding mutated residues (S94A, I16T, I21T, I21V, I47T, V78A, I95P) in MT InhA during MD simulations.	113
Table 2.8	Comparison of amino acids side chain properties in WT InhA (S94, I16, I21, I47, V78 and I95) with MT InhA (S94A, I16T, I21T, I21V, I47T, V78A and I95P).	113
Table 2.9	The diffusion coefficient calculated from the slope of mean square displacement plot (Figure 2.20) at production run (201-1000 ps) for water molecules as the function of time around the solute molecules for INADH-S94A, INADH-I16T, INADH-I21T and INADH-I21V, INADH-I47T, INADH-V78A, and INADH-I95P compared to INADH-WT.	116
Table 2.10	The results from radial distribution functions analysis. Only the pair distribution between INADH receptor atoms and oxygen atom of water molecule with the pronounced sharp peak within 3.20 Å showed here.	121
Table 2.11	The hydrogen bond analysis in INADH-InhA systems from MD simulations (at production run: 201-1000ps). Only the hydrogen bonds that occupied at least 80% were stated here. Drawing tool: ACD/ChemSketch Freeware (http://www.acdlabs.com).	123
Table 2.12	Chemical structures and length comparison of amino acids side chain in WT InhA (S94, I16, I21, I47, V78 and I95) with MT InhA (S94A, I16T, I21T, I21V, I47T, V78A and I95P). Drawing tool: ACD/ChemSketch Freeware (http://www.acdlabs.com).	136

Table 3.1	Summary of hydrogen bonds, hydrophobic contacts, van der Waals, electrostatic and π - π interactions of INH derivatives with WT and S94A MT InhA.	185
Table 3.2	Calculated binding free energies and inhibition constants of INH derivatives in WT and S94A MT InhA from docking simulation compared with experimental MIC values (Mohamad, 2002).	188
Table 3.3	The properties of the potential “druglikeness” compound (INH17) compared with INH and INADH.	194

LIST OF FIGURES		Page
Figure 1.1	Chemical structure of INH ($C_6N_3OH_7$). Drawing tool: ACD/ChemSketch Freeware (http://www.acdlabs.com).	1
Figure 1.2	A schematic one-dimensional energy surface. Minimization methods move downhill to the nearest minimum.	8
Figure 1.3	The steepest descents method in the search of minimum energy.	10
Figure 1.4	The conjugate gradients minimisation in the search of minimum energy.	10
Figure 1.5	Schematic representations of the key contributors to a molecular mechanics force field: bond stretching, angle bending and torsional terms as bonded interactions while van der Waals and electrostatic forces as non-bonded interactions.	13
Figure 1.6	The Lennard-Jones (LJ) potential. $V(r)$ is the pairwise potential energy; r is the internuclear separation in between two non-bonded atom; r_{eqm} is the equilibrium internuclear separation; ϵ is the well depth at r_{eqm} .	16
Figure 1.7	A two dimensional periodic boundary conditions.	20
Figure 1.8	The minimum image convention and spherical cutoff (circle).	20
Figure 1.9	The thermodynamic cycle for the binding of enzyme (E) and inhibitor (I) in both solvated and <i>in vacuo</i> phase. Circles indicate the solvent molecules.	27
Figure 1.10	Illustration of grid map.	27
Figure 2.1	Chemical structures of <i>M. tuberculosis</i> mycolic acids adapted from Sacchettini and Blanchard (1995). Drawing tool: ACD/ChemSketch Freeware (http://www.acdlabs.com).	41
Figure 2.2	The chemical structure of nicotinamide adenine dinucleotide (NAD^+). Drawing tool: ACD/ChemSketch Freeware (http://www.acdlabs.com).	44
Figure 2.3	The chemical structure of isonicotinic acyl-NADH (INADH) adapted from Chen <i>et al.</i> (2001). Drawing tool: ACD/ChemSketch Freeware (http://www.acdlabs.com).	44
Figure 2.4	The pathways for formation of InhA inhibitor, isonicotinic acyl-NADH (INADH), as proposed by Rozwarski <i>et al.</i> (1998). Drawing tool: ACD/ChemSketch Freeware (http://www.acdlabs.com).	45
Figure 2.5	The ribbon model of x-ray crystallographic structure of WT InhA from H37Rv <i>M. tuberculosis</i> (PDB ID: 1ZID).	52

Figure 2.6	The ribbon model of x-ray crystallographic structure of <i>M. tuberculosis</i> S94A MT InhA (PDB ID: 1ENZ). The green CPK model represents the mutation of serine to alanine (green) at residue 94.	52
Figure 2.7	The line model of superimposed WT (blue line presentation) and S94A MT (red line presentation) InhA with root mean square 0.6 Å difference in backbone atoms. Serine 94 (in WT InhA) is shown in green while alanine 94 (in S94A MT InhA) is yellow.	53
Figure 2.8	The molecular surface presentation of WT InhA (PDB ID: 1ZID) with the inhibitor- INADH (blue stick presentation). The black line represents the coverage of the grid map calculation.	55
Figure 2.9	The chemical structures of INH, NADH and INADH. Drawing tool: ACD/ChemSketch Freeware (http://www.acdlabs.com).	56
Figure 2.10	Line models of superimposed before (blue) and after (red) minimisation of a) INADH-WT, b) INADH-S94A, c) INADH-I16T, d) INADH-I21T, e) INADH-I21V, f) INADH-I47T, g) INADH-V78A, and h) INADH-I95P InhA with root mean square of 0.8, 0.8, 0.7, 0.8, 0.7, 0.7, 0.7 and 0.7 Å difference (all atom).	62
Figure 2.11	Docked conformations of INH, NADH, INADH (stick presentation) compared with the crystal structure of NADH and INADH (stick presentation) in the active site (surface and line presentation) of a) WT InhA and b) S94A MT InhA. All ligands bound at the same binding site.	67
Figure 2.12	Ball and stick presentation of the docked conformations of a) INH, b) INADH, and c) NADH in the active site (grey line presentation; grey stick is Ser94) of WT InhA (grey ribbon presentation). Blue dotted line shows the hydrogen bond between ligand-InhA. Ser94 of WT InhA is within the binding site of the ligands.	69
Figure 2.13	Ball and stick presentation of the docked conformations of a) INH, b) NADH, and c) INADH in the active site (grey line presentation) of S94A (grey stick presentation) MT InhA (grey ribbon presentation). Blue dotted line shows the hydrogen bond between ligand-InhA. S94A of MT InhA is within the binding site of the ligands.	71
Figure 2.14	WT and a) S94A, b) I16T, c) I21T and I21V, d) I47T, e) V78A, and f) I95P MT InhA RMS fluctuation per residue about the time average structure.	80
Figure 2.15	WT InhA radius of gyration as a function of time compared with a) S94A, b) I16T, c) I21T and I21V, d) I47T, e) V78A and f) I95P MT InhA.	83

Figure 2.16	Average and last (1000 ps) structures of a) INADH-WT, b) INADH-S94A, c) INADH-I16T, d) INADH-I21T, e) INADH-I21V, f) INADH-I47T, g) INADH-V78A, and h) INADH-I95P from MD simulation compared to the crystal structure. While (i) is the docked, average MD and last MD INADH (stick presentation) structures in WT compared to that of in S94A (j). The CPK represents INADH while ribbon represents InhA.	84
Figure 2.17	The χ_1 dihedral (torsional) angle of residues S94, I16, I21, I47, V78 and I95 in WT InhA compared to a) S94A, b) I16T, c) I21T and I21V, d) I47T, e) V78A, and f) I95P in MT InhA plotted as a function of time.	90
Figure 2.18	The cluster analysis for a) INADH-WT, b) INADH-S94A, c) INADH-I16T, d) INADH-I21T, e) INADH-I21V, f) INADH-I47T, g) INADH-V78A, and h) INADH-I95P complex with the function of time.	92
Figure 2.19	The cluster analysis for INADH in a) WT, b) S94A, c) I16T, d) I21T, e) I21V, f) I47T, g) V78A, and h) I95P system with the function of time.	94
Figure 2.20	The member with the closest RMSD value from the centroid of the most populated cluster for a) INADH-InhA complex, and b) INADH in InhA system.	97
Figure 2.21	The dihedral angle (D1 to D12) of INADH backbone.	98
Figure 2.22	The dihedral angle of INADH backbone in a) S94A, b) I16T, c) I21T and I21V, d) I47T, e) V78A, and f) I95P InhA system compared with that of WT InhA system as a function of time.	99
Figure 2.23	The distance of residues S94, I16, I21, I47, V78 and I95 (in WT InhA) from INADH compared to a) S94A, b) I16T, c) I21T and I21V, d) I47T, e) V78A, and f) I95P mutant type residue in MT InhA plotted as a function of time.	112
Figure 2.24	The average mean square displacement plot for water molecules as the function of time around the solute molecules a) INADH-S94A, b) INADH-I16T, c) INADH-I21T and INADH-I21V, d) INADH-I47T, e) INADH-V78A, and f) INADH-I95P compared to INADH-WT.	115
Figure 2.25	The average mean square displacement plot for total distance travel by water molecules as the function of time around the solute molecules a) INADH-S94A, b) INADH-I16T, c) INADH-I21T and INADH-I21V, d) INADH-I47T, e) INADH-V78A, and f) INADH-I95P compared to INADH-WT.	117
Figure 2.26	The number of water molecule within the distance of 3.20 Å from INADH in a) S94A, b) I16T, c) I21T and I21V, d) I47T, e) V78A, and f) I95P system compared to WT system.	118

Figure 2.27	The radial distribution function (solid line) and the corresponding integration number (dotted line) from INADH receptor atoms (phosphorus atom, nitrogen atom and oxygen atom) to oxygen atom of water molecule in different InhA systems.	120
Figure 2.28	Water mediated hydrogen bonding in INADH-InhA from x-ray crystallography (PDB ID: 1ZID). Drawing tool: ACD/ChemSketch Freeware (http://www.acdlabs.com).	145
Figure 2.29	Schematic drawing of INADH, the position of WT/MT InhA residues, water molecule and some InhA residues which occur commonly in all INADH-InhA systems during MD simulations. Drawing tool: ACD/ChemSketch Freeware (http://www.acdlabs.com).	151
Figure 2.30	Average MD simulation structure at WT InhA binding pocket (ribbon model represents WT InhA backbone, line model represents residues in the binding pocket of WT InhA, stick model represents INADH and CPK model represents water molecule).	152
Figure 2.31	INADH at S94A MT InhA binding pocket (ribbon model represents S94A InhA backbone, line model represents residues in the binding pocket of S94A InhA, ball and stick model represents Ala94, stick model represents INADH and CPK model represents water molecule).	154
Figure 2.32	INADH at I16T MT InhA binding pocket (ribbon model represents I16T InhA backbone, line model represents residues in the binding pocket of I16T InhA, ball and stick model represents Thr16, stick model represents INADH and CPK model represents water molecule).	156
Figure 2.33	INADH at I21T MT InhA binding pocket (ribbon model represents I21T InhA backbone, line model represents residues in the binding pocket of I21T InhA, ball and stick model represents Thr21, stick model represents INADH and CPK model represents water molecule).	158
Figure 2.34	INADH at I21V MT InhA binding pocket (ribbon model represents I21V InhA backbone, line model represents residues in the binding pocket of I21V InhA, ball and stick model represents Val21, stick model represents INADH and CPK model represents water molecule).	160
Figure 2.35	INADH at I47T MT InhA binding pocket (ribbon model represents I47T InhA backbone, line model represents residues in the binding pocket of I47T InhA, ball and stick model represents Thr47, stick model represents INADH and CPK model represents water molecule).	162
Figure 2.36	INADH at V78A MT InhA binding pocket (ribbon model represents V78A InhA backbone, line model represents residues in the binding pocket of V78A InhA, ball and stick model represents Ala78, stick model represents INADH and CPK model represents water molecule).	164

Figure 2.37	INADH at I95P MT InhA binding pocket (ribbon model represents I95P InhA backbone, line model represents residues in the binding pocket of I95P InhA, ball and stick model represents Pro95, stick model represents INADH and CPK model represents water molecule).	166
Figure 3.1	The mycobacteria cell envelope. This figure is adapted in part from Somoskovi <i>et al.</i> (2001).	172
Figure 3.2	The chemical structures of INH derivatives (INH6 to INH19). The number after INH represents the number of carbon atom in the aliphatic side chain. Drawing tool: ACD/ChemSketch Freeware (http://www.acdlabs.com).	179
Figure 3.3	Docked conformations of 14 INH derivatives (stick presentation) in WT InhA (surface and line presentation; Ser94 is presented as ball and stick) compared with the crystal structure of INADH (stick presentation).	181
Figure 3.4	Docked conformations of 14 INH derivatives (stick presentation) in the S94A (ball and stick presentation) MT InhA (surface and line presentation) compared with the crystal structure of INADH (stick presentation).	183
Figure 3.5	Calculated free energies of INH derivatives in WT and S94A InhA from docking simulation compared to the number of carbon atom in aliphatic side chains of the INH derivatives.	189
Figure 3.6	Calculated inhibition constant (K_i) values of INH derivatives in <i>M. tuberculosis</i> WT InhA from docking simulation compared with experimental MIC values (Mohamad, 2002) of INH derivatives in <i>M. tuberculosis</i> H37Rv (MIC values for INH10, INH11, INH12, INH13, INH15, INH17 and INH19 are not available).	189

LIST OF ABBREVIATION

1ZID	PDB identity for WT InhA
1ENZ	PDB identity for S94A MT InhA
AHAs	<i>N</i> -aryl hydroxamic acids
AHUs	<i>N</i> -aryl- <i>N</i> -hydroxy urethanes
AIDS	Acquired immune deficiency syndrome
AMBER	Assisted Model Building and Energy Refinement
ARP	<i>Arythromyces ramosus</i> peroxidase
ATP	Adenosine triphosphate
CAII	Carbonic anhydrase II
CK2	Casein kinase 2
CoA	Coenzyme A
FEP	Free energy perturbation
GA	Genetic Algorithm
GroES	Heat shock 10 kDa protein 1
GroEL	Heat shock 60 kDa protein 1
HIV	Human immunodeficiency virus
I16T	Isoleusine to threonine mutation at amino acid 16
I21T	Isoleusine to threonine mutation at amino acid 21
I21V	Isoleusine to valine mutation at amino acid 21
I47T	Isoleusine to threonine mutation at amino acid 47
I95P	Isoleusine to proline mutation at amino acid 95
ID	Identity
IgE	Immunoglobulin E
INADH	Isonicotinic acyl-NADH
InhA	Enoyl-acyl carrier protein (ACP) reductase
INH	Isoniazid
LGA	Lamarckian Genetic Algorithm
LIE	Linear interaction energy
LJ	Lennard-Jones
MAC	<i>Mycobacterium avium</i> complex
MD	Molecular dynamics
MIC	Minimum inhibitory constant
MMTSB	Multiscale Modeling Tool for Structural Biology
MT	Mutant type
NAD	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide (oxidised form of NAD)
NMR	Nuclear magnetic resonance
NPT	Isothermal-isobaric ensembles
NVE	Microcanonical ensembles
NVT	Canonical ensembles
PME	Particle mesh Ewald
QSAR	Quantitative structure-activity relationship
RMS	Root mean square
RMSD	Root mean square deviation
S94A	Serine to alanine mutation at amino acid 94
SH2/SH3	Sarcoma homology 2/3
Src	Sarcoma
TB	Tuberculosis
V78A	Valine to alanine mutation at amino acid 78
WHO	World Health Organisation
WT	Wild type

LIST OF APPENDICES		Page
Appendix 2.1	Grid parameter file (.gpf) for INADH.	232
Appendix 2.2	Dock parameter file (.dpf) for INH10.	233
Appendix 2.3	System fluctuation of MD simulation in a) INH, b) INH-WT, c) INADH, d) INADH-WT, e) INADH-S94A, f) INADH-I16T, g) INADH-I21T, h) INADH-I21V, i) INADH-I47T, j) INADH-V78A, and k) INADH-I95P systems.	235
Appendix 4.1	Computational time for docking and MD simulations.	237

LIST OF PUBLICATIONS

Page

- 1.1 Wahab, H., Choong, Y. S., Ibrahim, P. and Sadikun, A. (2008). Elucidating isoniazid resistance using molecular modelling. *J. Chem. Inf. Model.* **In Press.** 238

**INTERAKSI ISONIAZID DENGAN PROTEIN PEMBAWA ENOIL-ASIL
REDUKTASE (INH) *MYCOBACTERIUM TUBERCULOSIS*:
DARI PERSPEKTIF MOLEKUL**

ABSTRAK

Masalah kerintangan ubat dan peningkatan kes-kes tibi yang berterusan telah menggalakkan penyelidikan pembangunan ubat baru dan meningkatkan pemahaman mekanisme kerintangan ubat. Pendokkan dan simulasi dinamik molekul telah dijalankan untuk mengkaji aktiviti pengikatan isoniazid (INH; ubat terutama tibi) pada tapak aktif protein pembawa enoil-asil reduktase (InhA) *Mycobacterium tuberculosis* untuk meningkatkan pemahaman mekanisme tindakan and kerintangan INH terhadap bakteria ini. Hasil kajian menyokong teori yang pengaktifan INH kepada isonikotinic-asil-NADH (INADH) oleh enzim KatG menyumbang terhadap aktiviti terbaiknya. INADH didapati mempunyai tarikan kuat terhadap InhA dengan pembentukan lebih banyak ikatan hidrogen, interaksi-interaksi hidrofobik, elektrostatik, π - π dan daya van der Waals berbanding ubat induk (INH). Mutasi S94A pada InhA menyebabkan INADH berbeza daripada konformasi hablurnya akibat daripada kurang interaksi antara bahagian berketub pada NADH dengan Ala94 yang hidrofobik. Disebabkan oleh kehidrofobikan dan pengurangan isipadu rantai sisi asid amino (S94A, I16T, I21T, I21V, I47T, V78A and I95P) pada InhA mutan (MT InhA), maka INADH dan MT InhA terpaksa mengaturkan semula struktur mereka supaya dapat menyesuaikan dan menstabilkan kesan sterik dan elektrostatik. Akan tetapi perubahan atomik yang lebih besar dan ketidakstabilan struktur pada MT InhA berbanding dengan InhA liar (WT InhA) tidak ketara, maka hanya kekuatan ikatan dengan INADH pada MT InhA hanya dikurangkan sedikit. Gly14, Ile15 dan Ala22 didapati kerap menjana ikatan hidrogen dengan INADH atau molekul air semasa simulasi dinamik molekul mencadangkan bahawa asid-asid amino ini mungkin penting dalam pengikatan INADH. Bagi jenis *M. tuberculosis* dengan S94A MT InhA yang rintang terhadap INH, INH12 and INH17 daripada terbitan INH adalah alternatif yang lebih baik kerana tenaga pengikatan bebas mereka lebih rendah daripada INH.

Kajian ini telah membuktikan teknik permodelan molekul adalah cara yang dapat digunakan dalam menyelidik sifat-sifat pengikatan antara molekul penerima dan receptor pada peringkat molekul sebab ia berupaya menjana semula data-data eksperimen. Pendekatan pemodelan molekul juga merupakan cara yang berguna untuk mengkaji sesuatu molekul novel dengan sasarannya, khasnya jika tiada maklumat eksperimen mengenai cara pengikatannya.

**INTERACTION OF ISONIAZID WITH *MYCOBACTERIUM TUBERCULOSIS*
ENOYL-ACYL CARRIER PROTEIN REDUCTASE (INH):
FROM MOLECULAR PERSPECTIVES**

ABSTRACT

The problem of tuberculosis (TB) drug resistance and the continuing rise in the disease incidence has prompted the research on new drug development as well as on increasing the understanding of the mechanisms of drug resistance. Molecular docking and molecular dynamics (MD) simulations were performed to study the binding activity of isoniazid (INH; a first line anti-TB drug) onto the active site of *Mycobacterium tuberculosis* enoyl-acyl carrier protein reductase (InhA) in an effort to increase the understanding of the action and resistance of INH in this bacteria. The results support the theory that the activation of INH to isonicotinic acyl-NADH (INADH) by KatG enzyme is required for its ultimate activity. It is shown that INADH has tremendously high binding affinity towards InhA by forming more hydrogen bonds, hydrophobic, van der Waals, electrostatic and π - π interactions, compared to the parent drug (INH). S94A mutation of InhA causes INADH to deviate from its crystal structure probably due to the poor contact between its highly polar moiety of NADH region with the hydrophobic Ala94. It was found that due to the hydrophobicity and reduction in side chain volume of mutated residue (S94A, I16T, I21T, I21V, I47T, V78A and I95P) in mutant type (MT) InhA, INADH and MT InhA have to rearrange their structures in order to accommodate and stabilize the steric and electrostatic effects. However, the greater atomic fluctuation and structural instabilities by INADH-MT InhA compared with that of INADH-wild type (WT) InhA is not prominent, thus, resulted only slightly lower affinity of INADH in MT InhA. Gly14, Ile15 and Ala22 emerged quite frequent in forming hydrogen bonding with INADH or water molecules during MD simulations suggested that these molecules might be an essential part in INADH binding. In INH-resistant strains of *M. tuberculosis* with S94A MT InhA, derivatives INH12 and INH17 are better alternatives to INH as calculated binding free energy were more negative than INH. These studies show that

molecular modelling is a reliable method to investigate the binding properties of an acceptor and a receptor at the molecular level as it can reliably reproduce experimental data. The molecular modelling approach is also useful for exploring novel compound classes directed against a given target, especially when the experimental information about the binding mode is not available.

CHAPTER ONE

GENERAL OVERVIEW

1.1 Statement of the problem

The continuing rise in tuberculosis (TB) incidence especially in the present of human immunodeficiency virus (HIV) infection and the problem of drug resistance strains towards isoniazid (a first line anti-TB drug; also known as isonicotic acid hydrazide; INH; Figure 1.1) have prompted the research on new drug candidates and targets, as well as the mechanism of drug resistance (Styblo, 1991; Kirschner, 1999; Smith & Sacchetti, 2003; Oliveira *et al.*, 2006; Wells *et al.*, 2007). Thus, the objective of this study is to utilise molecular docking and molecular dynamics (MD) simulations to investigate the binding of INH onto the active site of *Mycobacterium tuberculosis* enoyl-acyl carrier protein (ACP) reductase (InhA) in an attempt to address the mycobacterial resistance against INH.

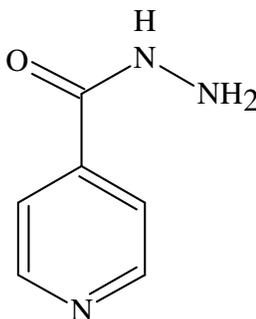


Figure 1.1 Chemical structure of INH (C₆N₃O₂H₇). Drawing tool: ACD/ChemSketch Freeware (<http://www.acdlabs.com>).

The formulation of a good drug candidate is much depend on the interaction of the candidate with its target enzyme, therefore there is an urgent need to elucidate the mechanism of INH resistance in InhA mutants at molecular level. So far, experimental studies (Dessen *et al.*, 1995; Rozwarski *et al.*, 1998, 1999; Oliveira *et al.*, 2006) were unable to address the resistance mechanism of INH with its target enzyme, InhA. Computational studies (Chen *et al.*, 2001; Pantano *et al.*, 2002; Schroeder *et al.*, 2005) on the resistance mechanism of INH were looking at interaction between the cofactor, oxidised form of nicotinamide adenine dinucleotide (NADH), with InhA. There were none that examine the mechanism of resistance of INH itself towards InhA mutants. Thus, the present study will allow the insights of molecular event that leads to INH resistant to be explored. This study also models the interaction between INH derivatives and InhA which indeed might lead to some clues for the strategies of new drug development.

1.2 Tuberculosis (TB)

TB is a microbial disease, commonly infect the human lower respiratory system, which has affected humans for several millennia (Tortora *et al.*, 1995). *M. tuberculosis*, which was isolated 125 years ago by Koch (Koch, 1882; cited from Youmans, 1979), is the aetiological agent of TB. However, this ancient scourge still remains as a major pathogen of human and a global tragedy with enormous public health and economic implications despite the advances in the prevention and treatment of the disease that it brings. Migliori *et al.* (2007) reported that World Health Organisation (WHO) identifies a need of US\$47 billion to implement countrywide programmes to stop TB while another US\$9 billion for the research and development of new diagnostics and treatments for TB.

TB accounts for 7% of all deaths in developing countries and 26% of avoidable adult deaths worldwide (Gomez & McKinney, 2004). Of all the infectious bacteria, it is currently the leading killer of adults in the world, with an estimated 3 million people dying from TB every year (Stokes & Doxsee, 1999; Morgan *et al.*, 2003). Manca *et al.* (1997) also reported

that WHO has estimated a staggering 8 million new cases globally and has projected about 30 million deaths from TB in this decade.

For over the past 10 years, there has been a steady increase notification of TB in Malaysia. A shocking 15,057 cases of TB was reported where the incidence rate is 64.7 per 100,000 populations in year 2000. The number of both TB and HIV co-infections has also escalated from 6 cases in 1990 to 734 cases in 2000. Most patients with TB-HIV co-infections are seen with advanced TB, thus the number of deaths due to TB-HIV has also increased. The worries are further compounded by the fast growing numbers of immigrant workers from high TB burden neighbouring countries which might add to the problem of multi-drug resistant TB (Iyawoo, 2004).

Since 1952, INH is the frontline drug for the treatment of TB (Middlebrook, 1952). However its use has been restricted by as much as 30% increase of INH resistant strains (Miesel *et al.*, 1998). The problem has further been complicated by the increase of multi drug resistant *M. tuberculosis* strains especially amongst HIV infected individuals (Eltringham *et al.*, 1999; Rodríguez *et al.*, 2002; Wells *et al.*, 2007). A significant number of non-TB mycobacteria have been isolated from acquired immune deficiency syndrome (AIDS) patients (Srivastava *et al.*, 1997). Such opportunistic mycobacteria include the member of *Mycobacterium avium* complex (MAC) caused the prevalent “TB-like” infection in the AIDS patients. These pathogens are generally intrinsically resistant to INH (Milano *et al.*, 1996; Espinal, 2004). The survival rates of AIDS patients with MAC infection are much lower if compared to those who are not infected (Musser, 1995; Pratt, 2007).

The above statements have prompted the search for alternatives to INH. However, in order to develop strategies for the design of novel and potent drugs against *M. tuberculosis*, the understanding of drug-receptor interactions is required. Thus, there is an urgent need to

understand resistance development at the molecular level, which until today remains an enigma.

1.3 The importance of molecular modelling in TB research

Today, the effectiveness of TB treatment, prevention and control programs have been complicated by the rise of multi-drug resistant strains and the limitation number of efficacious therapeutic agents to treat patients infected with MAC. The continuing worldwide threat of TB emphasises the urgent need of more effective diagnosis therapies, which is currently at a very slow process (Musser, 1995; Bardou *et al.*, 1996; Biava *et al.*, 2004) due to the lack of detailed structural features regarding the drug-receptor interactions. Therefore, the understanding and insights of the molecular events that lead to drug action or resistant in *M. tuberculosis* are important to facilitate the rational development and improvement of anti-TB medications (Musser, 1995; Rouse *et al.*, 1995; Torres *et al.*, 2000; Biava *et al.*, 2004).

Drug design mostly depends on the pharmacophore hypotheses which derived from those inhibitors with known structures (Carlson *et al.*, 1999, 2000; Masukawa *et al.*, 2000). Although these hypotheses were able to discover some new inhibitors, the pharmacophore model did not provide the details of the drug-receptor interactions. Therefore, molecular modelling method can predict the binding modes of INH as well as its derivatives onto InhA. With the understanding of binding modes at the molecular level between INH and InhA, the search for lead/potential inhibitor(s) and the strategies for the design of new anti-TB compounds can be formulated.

With impressive improvements in the accuracy and speed of molecular docking, acceptor-receptor binding mode predictions have led to a faster discoveries of new lead compounds. However, there are still many difficulties to overcome (Carlson & McCammon, 2000). First, the binding modes accuracy relies on the correct estimation of acceptor-receptor

interaction energies and scoring functions which are simplified for computational efficiency (Totrov & Abagyan, 1994; Zacharias *et al.*, 1994). Second, the sampling of acceptor in flexible binding pocket has not been achieved (Rosenfeld *et al.*, 1995; Carlson & McCammon, 2000). Third, the involvement of solvent molecules is yet to be addressed in molecular docking method (Ota & Agard, 2001). Fourth, the molecular docking method is not able to provide the dynamics data of acceptor within the receptor binding site as the acceptor might be more mobile in the bound state (Ota *et al.*, 1999).

Therefore, MD simulation is also another alternative to elucidate a more refined and complete understanding of the binding properties of INH within the InhA binding pocket. MD simulation technique is time consuming and expensive (regarding large computational storage compared to molecular docking method). However, MD simulation is able to provide detailed insights of intermolecular relationships as well as the dynamics of INH-InhA complex. MD simulation also able to tackle the docking problems mentioned above, as it uses more accurate force field (an all atom approach instead of the united atom method in molecular docking simulation). In most cases, water molecules play a critical role in determining the conformation of an acceptor in a binding pocket by mediating interactions with the protein (Levitt & Park, 1993; Ota *et al.*, 1999). Therefore, unlike molecular docking, MD simulation allows the involvement of explicit waters (instead of the simplified solvation parameter used in docking simulation). Finally, MD simulation methodology will also permit flexibility of both INH and InhA where their flexibilities might be critical for recognition between INH and InhA (Nagai & Mattaj, 1994; Janin, 1995; Westhof, 1997).

1.4 Molecular modelling

Advances in molecular biology, x-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy have led to high resolution at the molecular level of three dimensional structures for a numerous proteins (Lambert, 1997). However, not all proteins are easy to be crystallized (Peters & Pinto, 1996), while NMR experiment does not give

complete atomic structures but only overviews for proteins with amino acids beyond the number of 200 residues (Campbell, 1995; Asensio *et al.*, 2000). Therefore, computer-based molecular modelling techniques such as molecular docking and MD simulation are obviously attractive methods to predict and derive a molecular insight of the acceptor-protein complex structure (Hetényi & van der Spoel, 2002).

Molecular modelling is a term used to describe the study of molecules and molecular systems. It is also a technique to investigate and predict chemical entities and processes using theoretical methods. This technique has been used in many fields such as chemistry, biology, or materials science to study systems ranging from small chemical molecules to large biological molecules and material assemblies. Molecular modelling is able to give information at the atomistic level description of the studied systems. The simplest system can be performed by hand using theoretical calculations, but inevitably computers are required to study reasonably sized systems. Today's molecular modelling is also known as computational chemistry as it is invariably associated with computer modelling. Advances in speed and memory availability of computational power have allowed broader range of models and up to millions of atoms to be included in the calculation (Leach, 1996).

In general, computer-based molecular modelling simulates the chemical structures and reactions numerically based on the fundamental laws of physics. This method allows researchers to study chemical phenomena by running calculations instead of wet-lab experiment. Some simulations can be used to model not only stable molecules, but also unstable intermediates and transition states of chemical structures and reactions. In this way, these theoretical methods can provide information about molecules and reactions, which is almost impossible to observe through wet-lab experiment. Therefore, molecular modelling is an independent research area which could be vital adjunct and alternative to experimental studies (Foresman & Frisch, 1993).

1.5 The fundamental of molecular modelling

1.5.1 Energy minimization

In real life, molecules vibrate around their equilibrium structures with minimum energy. The change of energy can be regarded as movements on a three dimensional energy surface. At a minimum point, the forces on all atoms are zero and correspond to stable structures and any movement away from the minimum will give a higher energy structure. Figure 1.2 shows an energy surface and the minimum that would be gained by starting from point X, Y and Z. The minimum is corresponding to the positions where a ball rolling down the energy surface with the gravity influences and come to a stop. Global energy minimum is the minimum with the lowest energy. Minimization algorithm is thus, a method in the calculation and identification of the geometries of a system that correspond to minimum points on the energy surface (Cramer, 2002).

Energy minimization is an integral part of molecular modelling techniques. It is recommended to be part of the protocol for the simulation of complex systems such as macromolecules or large molecular assemblies, e.g. drug-protein complex in order to relieve any unfavourable interactions and stereo overlaps in the starting configuration of a system. For example, during a molecular simulation, the missing hydrogen atoms are usually added based on the geometry specified in the residue database and may not correspond to the actual minimum in the force field. Thus, the placement does not consider any conflicts, electrostatic inconsistencies or overlaps with other atoms or residues. Large stereo overlaps will lead to large forces initially, which will also form large velocities and may cause the distortion of the structure or system. For this reason, prior to simulation, it is necessary to minimise the initial system, thus the energy.

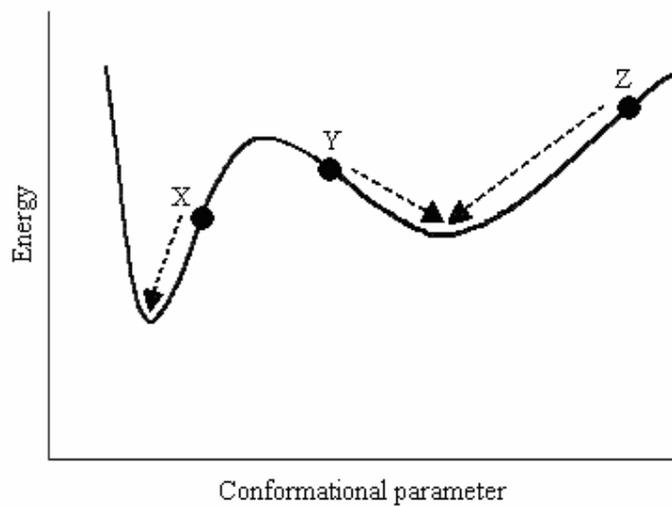


Figure 1.2 A schematic one-dimensional energy surface. Minimization methods move downhill to the nearest minimum.

1.5.1.1 Minimization algorithms

Minimization algorithms are classified into two types: those that use derivatives of energy with respect to the coordinates and those that do not. The choice of minimization algorithm depends on a number of factors. These include the speed and storage of the workstation, the availability of analytical derivatives and the robustness of the algorithm. Calculations on systems containing thousands of atoms such as biological systems usually use the first derivative minimization methods, e.g. steepest descents method and conjugate gradients minimization.

Steepest descents method moves to the minimum force in a parallel direction, which corresponds to walking straight down hill in the geographical analogy (Figure 1.3). The direction of this method is determined by the highest inter-atomic forces and thus, it is a good method to relieve the initial highest energy features. However, steepest descents method faces the problem of too many small steps have to be performed when going down a long narrow valley and, might oscillate and continuously over-correct itself (Cramer, 2002).

The conjugate gradients method provides directions which do not show the steepest descents oscillatory behaviour in narrow valleys. In this method, the gradients at each point are orthogonal but the directions are “conjugated” (Figure 1.4). It is less superior to steepest descents method when the initial starting configuration is still too far from the minimum (Leach, 1996).

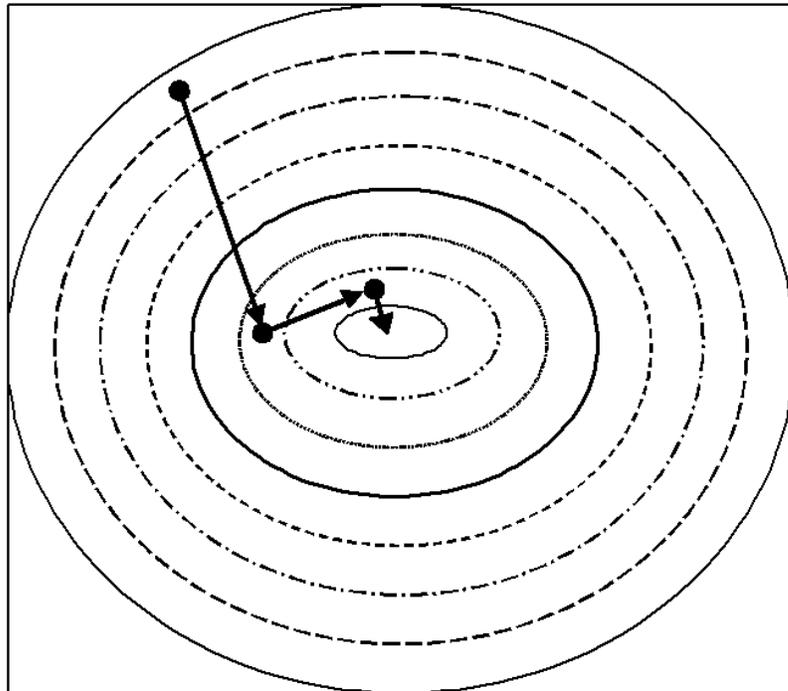


Figure 1.3 The steepest descents method in the search of minimum energy.

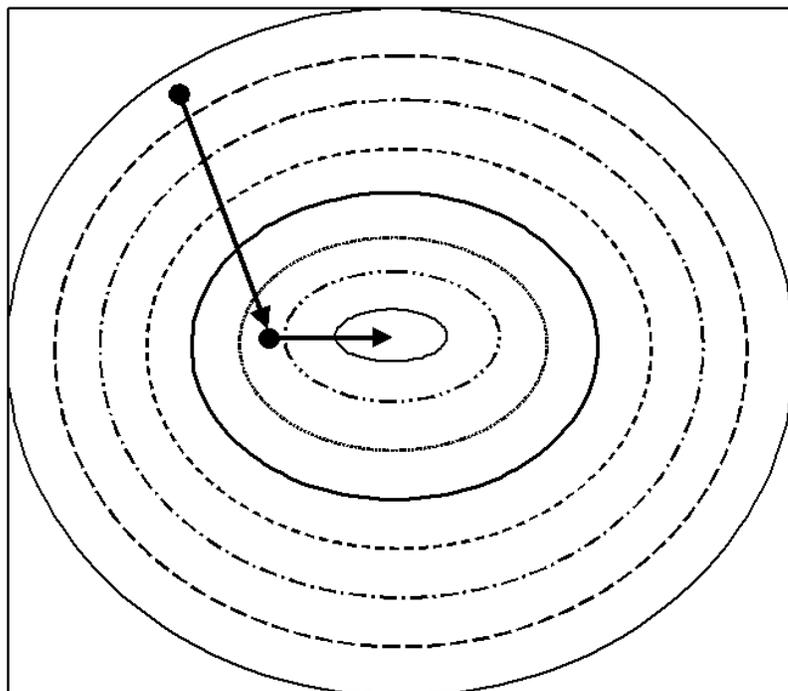


Figure 1.4 The conjugate gradients minimisation in the search of minimum energy.

It is believed that the best way of minimization is to relieve the initial stress of a system by the steepest descents prior to the conjugate gradients. Specifically, in a system with large number of atoms, frequently two stages of minimization will be used. In order to get a structure without significant high-energy interactions, the steepest descents method will be used as a first stage of minimization. Conjugate gradients minimization will be the second stage minimization after the steepest descents method to give the structure closer to the minimum energy. However, too few steps of minimization will lead to instabilities while too many minimization steps will over minimized and modified the model away from the control structure (the x-ray crystallography structure). Thus, the complete structure of a system is best to be minimized briefly with respect to the atomic coordinates, usually with a mix of both steepest descend and conjugate gradient methods (e.g. 1000 steps for each method) without any constrains on the atoms (Beck & Daggett, 2004).

1.5.2 Molecular mechanics force field

The potential energy in a relatively large system such as biological system is derived using molecular mechanics (MM; also known as empirical force field method or force field method) method. A force field has three components (Foresman & Frisch, 1993):

- a. A set of equations to define how the potential energy of an atom varies with the position.
- b. A series of atom types to define the characteristics of the elements. The atom types prescribe different characteristic depending on the environment, and depend on hybridisation, charge and the types of other atoms to which an atom is bonded. For example, a carbon atom in carbonyl group is treated differently than the one bonded to a methyl group.
- c. Parameter sets that fit the equations and atom types to experimental data. The parameter sets define force constant values used in the equations to relate atomic characteristics to energy components, and structural data (such as bond lengths and angles).

In MM, electrons effects in a system are implicitly included in the force fields via its parameterisation. This approximation makes MM computationally relatively inexpensive and allows the used of large systems containing millions of atoms, but it carries some limitations (Foresman & Frisch, 1993):

- a. Certain force fields achieve good results only for a particular class of molecules which have been parameterised. No force field can be generally used for all molecular systems.
- b. Implicit treatment of electrons means that MM cannot simulate chemical problems where electronic effects predominate. For example, bond formation or bond breaking, or proton transfer, or predict molecular properties which depend on molecular orbital interaction, cannot be described.

In a molecular system, the intra-molecular forces, generated by the bonded interaction are contributed by bond stretching, angle bending and bond torsions. While the inter-molecular forces, generated by the non-bonded interaction, include van der Waals and electrostatic interactions (Figure 1.5). Equation 1.1 generally represents the MM force field.

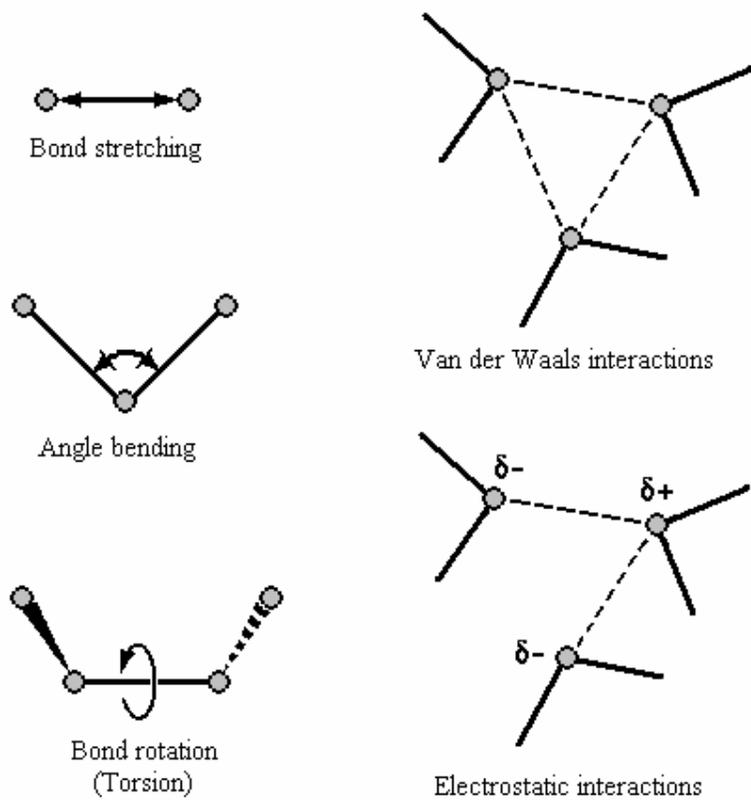


Figure 1.5 Schematic representations of the key contributors to a molecular mechanics force field: bond stretching, angle bending and torsional terms as bonded interactions while van der Waals and electrostatic forces as non-bonded interactions.

$$\begin{aligned}
E(\mathbf{r}^N) = & \sum_{\text{bonds}} \frac{k_\lambda}{2} (l_i - l_{i,0})^2 \\
& + \sum_{\text{angles}} \frac{k_\theta}{2} (\theta_i - \theta_{i,0})^2 \\
& + \sum_{\text{torsions}} \frac{k_n}{2} (1 + \cos[n\omega - \gamma]) \\
& + \sum_{\text{improper}} k_\tau (1 - \cos 2\omega) \\
& + \sum_{i=1}^N \sum_{j=i+1}^N 4\mathcal{E}_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \\
& + \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad \dots \quad 1.1
\end{aligned}$$

$E(\mathbf{r}^N)$ denotes the potential energy with the function of the positions (\mathbf{r}) of N atoms. The first to the fourth terms are the bonded terms while the fifth and the sixth terms are the non-bonded terms. This potential energy function is the summation of all the bonded (i.e. bond, angle, torsion and improper torsion energies) and the non-bonded terms (i.e. van der Waals and electrostatic interactions).

The first term in Equation 1.1 calculates the interaction between pairs of bonded atoms using harmonic potential (Hooke's law) that will increase in energy as the bond length l_i deviates from the reference value $l_{i,0}$; k_ℓ is the force constant. The second term summates all valence angles in the molecule, also modelled using Hooke's law (the valence angle, θ_i , is the angle between atom A, B and C, which A and C are both bonded to B, to a reference value θ_0); k_θ is the force constant. The third term is to model torsional potential and to calculate the energy changes when a torsional bond rotates; k_n is the force constant, n is the multiplicity (the number of minimum points when the bond is rotated through 360°), ω is the torsion angle, γ is the phase factor (determines the torsion angle that passes through its

minimum). The k_τ in improper torsion potential, the fourth term, is the force constant and ω is the improper dihedral angle.

The last two terms (the fifth and the sixth terms) are the calculation between pairs of atoms (atom i and atom j) that are in different molecules, or in the same molecule but are separated by at least three bonds. The fifth term models the van der Waals interactions using Lennard-Jones (LJ) 12-6 potential, where ϵ_{ij} is the dispersion well depth, σ_{ij} is the LJ distance and r is the distance. The last term models the electrostatic interactions using Coulomb potential, where q_i and q_j are the atomic partial charges for atom i and atom j , ϵ is the dielectric constant, r_{ij} is the distance between atom i and atom j (Figure 1.6).

Electrostatic interaction energy calculation can be improved by Particle Mesh Ewald (PME) method. This method is for fast implementation of Ewald summation with B -splines interpolation to calculate the electrostatic energy of a given unit cell in a macroscopic lattice consists of repeating images for a better analytic gradients and accuracy (Essmann *et al.*, 1995). PME method in Assisted Model Building and Energy Refinement (AMBER) simulation package, which is used in this study for MD simulation, not only provides a better treatment of long-range electrostatics at a modest computational cost, but can also be applied in periodic boundary simulation (see below).

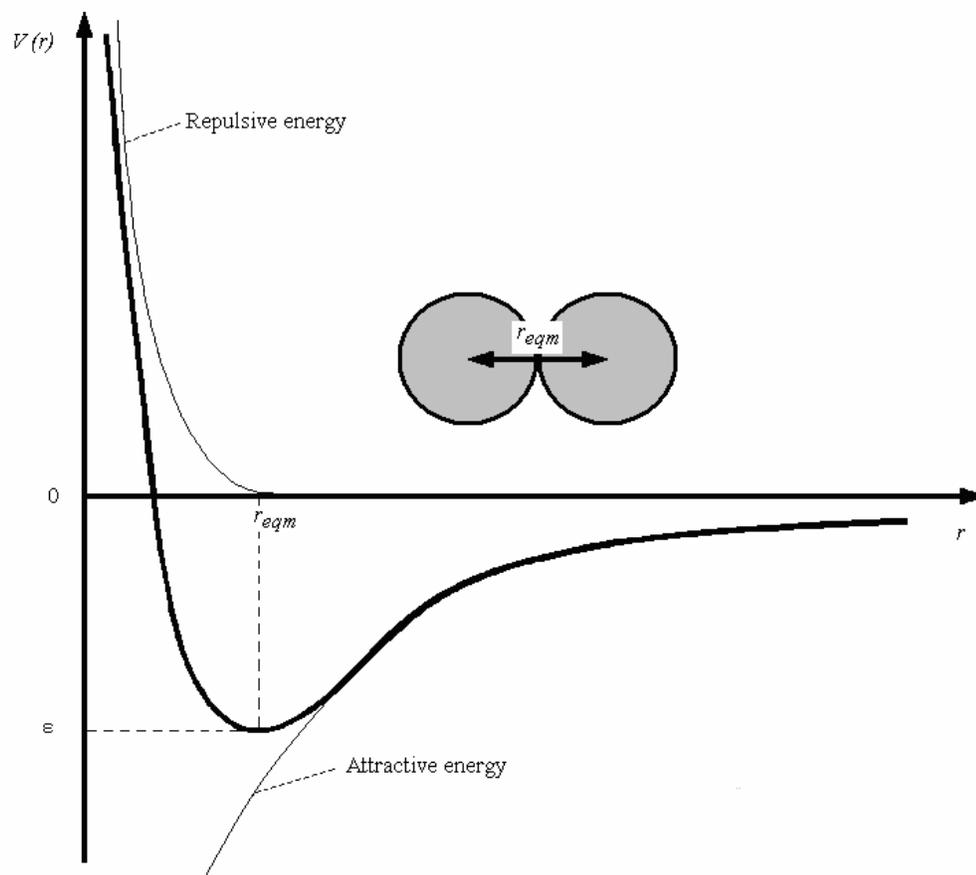


Figure 1.6 The Lennard-Jones (LJ) potential. $V(r)$ is the pairwise potential energy; r is the internuclear separation in between two non-bonded atom; r_{eqm} is the equilibrium internuclear separation; ϵ is the well depth at r_{eqm} .

1.5.3 Principles of MD simulations

MD simulation is the science of simulating the motions of particles in a system and gives the fluctuation in the relative positions of atoms as a function of time (Karplus & Petsko, 1990). In other words, MD simulation is a method/technique of studying time evolution of the atomic degrees of freedom by solving the Newtonian equations of motion numerically (Allen & Tidesley, 1987). However, MD cannot be solved simultaneously. Thus, numerical methods are applied to calculate the dynamics of the system. A large number of MD programs have been developed to provide an insight of the dynamics of molecules, for example, ENCAD (Levitt *et al.*, 1995), GROMOS96 (van Gunsteren *et al.*, 1996), Q (Marelius *et al.*, 1998c), MAXIMOBY (Jödicke *et al.*, 2001) and AMBER (Case *et al.*, 2004).

In MD, integrating Newton's law of motion will generate successive configurations of a system and give the trajectory that specifies how the positions and velocities of the particles in the system vary with time. These trajectories and the motions of the system are obtained by solving the equation embodied in Newton's second law (Equation 1.2).

$$F_i = m_i a_i \quad \dots 1.2$$

The forces exerted on the atoms, F_i , is determined by the masses, m_i , and the accelerations, a_i of the system with i atoms. The forces acting on the atoms can be obtained from the derivatives of the potential energy function, U , with respect to the atoms positions, r_i (Equation 1.3).

$$\frac{dU}{dr_i} = F_i \quad \dots 1.3$$

The accelerations of the atoms (a_i) can be determined (Equation 1.4).

$$a_i = \frac{F_i}{m_i} \quad \dots 1.4$$

Thus, the changes in atoms velocities and positions can be calculated with the knowledge of the accelerations of all atoms and the trajectories of the atoms as a function of time can be obtained.

Equation 1.5 is the Verlet algorithm (Verlet, 1967) for the atomic trajectories integration in MD simulation. The atoms new positions at $t + \delta t$, $r_i(t + \delta t)$ can be calculated from the previous step, $r_i(t - \delta t)$, uses the positions (r_i) and accelerations (a_i) at time.

$$r_i(t + \delta t) = 2r_i(t) - r_i(t - \delta t) + \delta t^2 a_i(t) \quad \dots 1.5$$

Thus, the velocities (v_i) can be calculated similarly by dividing the difference in positions at time $t + \delta t$ and $t - \delta t$ by $2\delta t$ (Equation 1.6).

$$v_i(t) = [r_i(t + \delta t) - r_i(t - \delta t)]/2\delta t \quad \dots 1.6$$

The leap-frog algorithm from Hockney (1970) is a variation of Verlet algorithm. It is used in AMBER, a program that is used throughout this study. Other methods to integrate the equation include the predictor-corrector methods (Gear, 1971), Beeman's algorithm (Beeman, 1976) and velocity Verlet method (Swope *et al.*, 1982).

1.5.3.1 MD simulation ensembles

Traditionally, MD is performed with the microcanonical ensembles (NVE)- constant number of atoms, volume and energy; canonical ensembles (NVT)- constant number of atoms, volume and temperature; or isothermal-isobaric ensembles (NPT)- constant number of atoms, pressure and temperature (Mihailescu & Smith, 2000). Most experimental measurements are made under the conditions of constant temperature and pressure, thus simulations in NPT ensembles, which is used in this study, are most directly relevant to experimental data.

1.5.3.2 Boundaries and cutoff

Computer simulation enables the calculation of macroscopic properties using relatively small numbers of particles. Thus, the boundaries and boundary effects are crucial and have to be treated carefully in any MD simulation. The periodic boundary conditions application will enable the simulation of a relatively small number of particles to experience the forces as if they are in bulk fluid. In a two-dimensional example, each box is surrounded by eight image neighbours (Figure 1.7); thus, in three-dimensions, each box will be surrounded by 26 image neighbours. Should a particle leave the box during the simulation, it will be replaced by an image particle from its image neighbour (Essmann & Darden, 1997).

In MD simulation, the number of bonded terms in a force field is proportional to the number of atoms (N) but the number of non-bonded terms increases as the square of the number of atoms (N^2). Thus, minimum image convention and non-bonded cutoff are applied to handle this time consuming calculation. In minimum image convention, each atom only sees one image of every other atom at the most in the system (which the system is repeated via periodic boundary conditions for macroscopic effects). The non-bonded energy is thus calculated with the closest atom (Figure 1.8).

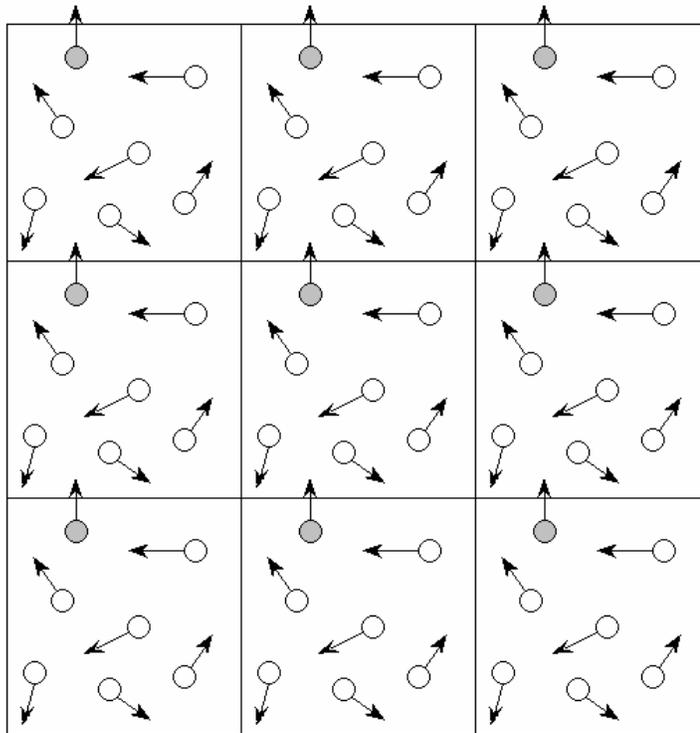


Figure 1.7 A two dimensional periodic boundary conditions.

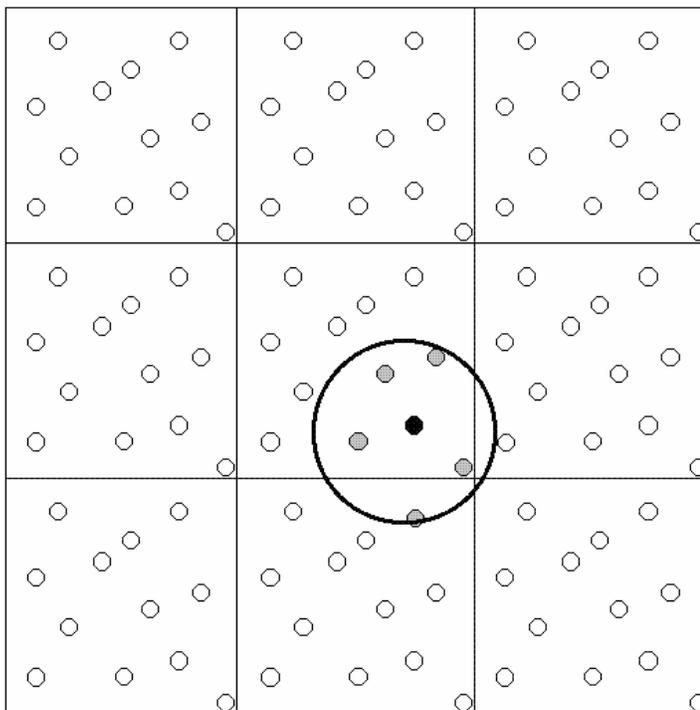


Figure 1.8 The minimum image convention and spherical cutoff (circle).

While in non-bonded cutoff, the interactions between all pairs of atoms among the closest image will be considered, but those further away from the cutoff value will be set to zero. However, during the application of periodic boundary conditions, the cutoff should not be so large until the particle sees its own image or the same molecule twice. Therefore, the cutoff is limited to less than half the length of the cell (Essmann & Darden, 1997). A very short cutoff (4 Å) usually will not adequately model the electrostatic properties of a system. Therefore, slightly longer cutoff ranges (such as 8 and 10 Å) have been shown to behave quite similar to much longer cutoff such as 12, 14 and 16 Å (Levitt *et al.*, 1995; Levitt *et al.*, 1997; Beck *et al.*, 2005).

1.5.3.3 Free energy calculation in MD simulation

Binding affinity calculation for an acceptor based on MD simulation provides an important and useful link between the structure and function of a biomolecular system. Therefore, characterizing the structures and energetics of molecular complexes are the keys for the understanding of many biological systems/functions (Ajay *et al.*, 1997; Marelus *et al.*, 1998c; Åqvist & Marelus, 2001; Åqvist *et al.*, 2002). Computational methods that could be used for molecular complexes binding affinity calculations including empirical scoring method (Böhm, 1992, 1998; Wang *et al.*, 1998b; Wang *et al.*, 2002b), free energy perturbation (FEP) approach (Beveridge & DiCapua, 1989; Kollman, 1993; Kollman *et al.*, 2000) and linear interactions energy (LIE) method (Åqvist *et al.*, 1994, 2002; Jones-Hertzog & Jorgensen; 1997; Hansson *et al.*, 1998; Åqvist & Marelus, 2001).

In empirical scoring method, the accuracy is limited due to the simplicity of the scoring function. Water molecules are usually not considered, which could be a problem especially in biological complexes as these water molecules are an integral part of the system, e.g. bridging interactions between the acceptor and the receptor. FEP method is a more accurate method for the acceptor-receptor binding free energy calculation. It provides a rigorous way of free energies and pair-wise interactions evaluation, but it is however

computationally expensive and time consuming (Åqvist & Marelus, 2001; Asi *et al.*, 2004). Between the two extremes of empirical scoring functions and FEP methods, LIE method (based on the estimation of force field energy from acceptor-receptor interactions and thermal conformational sampling) has emerged as a useful method to calculate the binding free energy from MD simulation. LIE approach has the advantages of a) the thermal averaging over conformations could be calculated, b) in contrast to FEP method, it does not require the transformation process involving unphysical states, c) it is based on estimating the absolute binding energies, and d) it is slower than scoring of single conformations, but it is faster than FEP calculations (Hansson *et al.*, 1998).

In the present study, the ligand free energy of binding from MD simulation is calculated using LIE method. This method is based on the evaluation of ligand interaction energies in both bound (complex with receptor) and free states (without receptor). The idea deriving from LIE method is that only averages of the interaction energies between the ligand and its surroundings need to be evaluated. Thus, accordingly, two MD simulations have to be carried out, one with the ligand free in solution and one where it is bound to solvated receptor. The binding free energy (ΔG_{bind}) is thus, obtained from the averages (Equation 1.7).

$$\Delta G_{\text{bind}} = \alpha (\langle V_{l-s}^{LJ} \rangle_{\text{bound}} - \langle V_{l-s}^{LJ} \rangle_{\text{free}}) + \beta (\langle V_{l-s}^{el} \rangle_{\text{bound}} - \langle V_{l-s}^{el} \rangle_{\text{free}}) \quad \dots 1.7$$

The α is scaling factor for LJ interaction energy and β is scaling factor for electrostatics interaction energy. The $\langle V_{l-s}^{LJ} \rangle_{\text{bound}}$ denotes LJ interaction energy between the bound ligand and its surrounding environment (e.g. water molecules, protein atoms and ions) and $\langle V_{l-s}^{LJ} \rangle_{\text{free}}$ is LJ interaction energy between free-state ligand and its surrounding environment. The $\langle V_{l-s}^{el} \rangle_{\text{bound}}$ denotes electrostatic interaction energy between the bound ligand and its surrounding environment (e.g. water molecules, protein atoms and ions) and

$\langle V_{l-s}^{el} \rangle_{\text{free}}$ is electrostatic interaction energy between free-state ligand and its surrounding environment.

Åqvist and Hansson (1996) observed that the β value varies with the number of hydroxyl groups in a ligand (compounds with higher number of hydroxyl groups are associated with lower values of β). These considerations have led Hansson *et al.* (1998) to a model where the chemical group deviations for each ligand were taken into account. They divided ligand into four groups: a) charged, b) dipolar with no hydroxyl, c) dipolar with one hydroxyl and, d) dipolar with two or more hydroxyl, and assigned different electrostatic coefficient β parameter value to each group. These values of β (0.50, 0.43, 0.37 and 0.33, respectively) were obtained from the simulations of compounds from different groups. Same β values were used for both bound and free states for the same compound. The results obtained with the set of FEP-derived model yielded a good agreement with experimental values for the 18 complexes in the calibrations set. The α value (0.181) was also obtained from the same simulations. Thus, Hansson *et al.* (1998) summarized that an optimal equation for binding free energy calculation can be obtained with only one free parameter, while β depends on the class/group of ligand.

1.5.4 Molecular docking

Generally, biologically active molecules exert their effect by binding to a target or receptor. The ability of a “ligand” or “acceptor” to interact with a “target” or “receptor” is determined by the functional group(s) and the three-dimensional (3D) features (DesJarlais, 1997). A detailed understanding of the interactions between ligands and their receptors is essential for interpreting many biochemical phenomena (Sotriffer *et al.*, 2000a). “Ligand” comes from “ligare” (Latin) meaning a “band” or “tie”. It is used to define a molecule of any size that interacts with another molecule through non-covalent forces which does not involve any formation of covalent bond. The “target” or “receptor” - is typically larger molecule

(macromolecule e.g. enzyme or protein). The interaction between a ligand and receptor depends on the chemical or physical forces on each of them, between them and also their environment (Krumrine *et al.*, 2003).

The interactions between a ligand and a receptor have been widely investigated using molecular modelling techniques reviewed by Taylor *et al.* (2002). Hetényi and van der Spoel (2002) have listed steps for successful prediction of a ligand-target complex: a) the definition of the protein structure, b) the location of binding site, and c) the determination of the binding mode. Molecular docking is one of the extensively used molecular modelling techniques to predict the conformation of ligand-receptor complex. It is a method that seeks to find ways of fitting two molecules together in the favourable configuration and suggest the binding modes of the complex. A large numbers of docking programs have been developed, e.g. DOCK (Shoichet & Kuntz, 1993), GOLD (Jones *et al.*, 1995, 1997), FlexX (Rarey *et al.*, 1996), AutoDock (Morris *et al.*, 1996) and EUDOCK (Pang *et al.*, 2001).

The docking programs provide an automated procedure to predict the interaction of ligand with its receptor, fall into two broad categories: matching and docking simulated methods (Goodsell & Olson, 1990). The matching method creates a model of the active site from the receptor and attempts to dock the ligand into the modelled site as a rigid body by matching the ligand's geometry to that of the modelled active site and it is efficient to screen an entire chemical database rapidly for lead compounds. The second type of docking techniques: the docking simulated method models the docking of a ligand to a receptor in details where the ligand explores the translations, orientations, and conformations until an ideal site is found. The docking simulated technique are relatively slower than the matching method, but allows the flexibility of the ligand to be modelled and able to utilize detailed MM to calculate the energy of the ligand in complex with the active site. AutoDock that is used throughout this research is an example of the latter technique which allows flexible ligand to be docked.