

**COMPUTER-AIDED DRUG DESIGN OF
NEURAMINIDASE INHIBITORS AND MCL-1
SPECIFIC DRUGS**

NITIN SHARMA

(M.Sc. (Bioinformatics), BIT, Mesra)

**A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
DEPARTMENT OF PHARMACY
NATIONAL UNIVERSITY OF SINGAPORE**

2014

Declaration

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Nitin Sharma
2 December 2014

Acknowledgements

I would like to dedicate this thesis to the two most important people of my life my mother and my wife, who have supported me in good and bad times. In addition I would like to thank my brother and my friends who have been with me throughout the journey.

I wish to express my heartfelt appreciation to my supervisor, Assistant Professor YAP Chun Wei, who has provided me with excellent guidance and gave enough support and freedom to perform scientific research.

I would like to thank to Dr. CHAI Li Lin, Christina for allowing me to be a part of MCL-1 project which gave me valuable experience.

Finally, I wish to thank all members of the Pharmaceutical Data Exploration Laboratory (especially Sreemane) for their suggestions and help in one way or another.

Table of Contents

<i>Declaration</i>	<i>ii</i>
<i>Acknowledgements</i>	<i>iii</i>
<i>Table of Contents</i>	<i>iv</i>
<i>Summary</i>	<i>ix</i>
<i>List of Tables</i>	<i>xiii</i>
<i>List of Figures</i>	<i>xiv</i>
<i>List of Abbreviations</i>	<i>xvi</i>
<i>List of Publications</i>	<i>xviii</i>
<i>List of oral and poster presentations</i>	<i>xix</i>
<i>Thesis structure</i>	<i>xx</i>
Chapter 1	1
Introduction	1
1.1 Drug discovery process	1
1.2 Computer Aided Drug Design	3
1.2.1 Target identification	4
1.2.1.1 Homology Modeling.....	4
1.2.2 Lead Discovery.....	5
1.3 Ligand and Structure based drug design	6
1.3.1 Ligand-based drug design.....	6
1.3.1.1 Quantitative structure–activity relationship (QSAR).....	8
1.3.2 Structure-based drug design	10
1.3.2.1 Docking	10
1.3.2.2 Molecular dynamics.....	15
1.4 Lead optimization	18
1.5 Objective	19
Chapter 2	22
Methods	22
2.1 QSAR	22
2.1.1 Data selection and curation.....	25
2.1.2 Descriptor calculation.....	25
2.1.3 Descriptor selection	26

2.1.3.1	Pre-processing.....	26
2.1.3.2	Selection	27
2.1.3.2.1	Genetic Algorithm	28
2.1.4	Model development	29
2.1.4.1	k nearest neighbor.....	30
2.1.4.2	Support Vector Machine	31
2.1.4.3	Applicability domain (AD).....	31
2.1.5	Validation	33
2.1.5.1	Internal validation.....	34
2.1.5.2	External validation.....	34
2.1.5.3	Predictive performance	35
2.1.6	Consensus model.....	37
2.2	Docking.....	38
2.2.1	Receptor Preparation	38
2.2.2	Identification of active site	38
2.2.3	Ligand preparation.....	39
2.2.4	Docking	39
2.3	Molecular Dynamics.....	39
2.3.1	System Preparation.....	40
2.3.2	Minimization	40
2.3.3	Heating up the system and equilibration	41
2.3.4	Production run	41
Chapter 3	42
Neuraminidase	42
3.1	Influenza virus.....	42
3.1.1	Influenza A	43
3.1.2	Structure of Influenza A virus	43
3.1.3	Virus life cycle	44
3.1.4	Antigenic variation	47
3.1.4.1	Antigenic Drift.....	47
3.1.5	Antigenic Shift	48
3.1.6	Characteristic function of Neuraminidase	48
3.1.7	Neuraminidase as a drug target	51
3.1.8	Structure of neuraminidase	51
3.1.9	Active site of neuraminidase	52
3.1.10	Neuraminidase inhibitors	54
3.1.11	Drug resistance.....	55
Chapter 4	57
Neuraminidase Methods	57
4.1	QSAR	57
4.1.1	Dataset curation.....	57

4.1.2	Descriptor calculation.....	59
4.1.3	Development of QSAR model and screening.....	60
4.2	Docking.....	60
4.2.1	Structure preparation	60
4.2.2	Active site.....	62
4.2.3	Dataset for virtual screening.....	62
4.2.4	Molecular docking.....	62
4.2.5	Energy minimization and rescoring	63
Chapter 5		66
Neuraminidase Results and Discussion		66
5.1	QSAR	66
5.1.1	Base Models	69
5.1.2	Performance of consensus model	69
5.1.3	Compounds outside AD	70
5.2	Docking.....	75
5.2.1	Energy Minimization and Rescoring	80
5.2.1.1	Standard Deviation of the docking scores.....	80
5.2.1.2	Correlation between IC50 and average binding free energy (ABFE).....	82
5.2.2	Conformations of Glutamic276 in non-mutant strains	84
5.2.3	Conformation of Glutamic276 leading to resistance	84
5.2.3.1	N294S and H274Y mutations	84
5.2.3.2	R292K mutation.....	87
5.2.4	Comparison of the poses of potential inhibitors with wild strains.....	88
5.2.5	Comparison of the poses of potential inhibitors with mutant strains.....	91
Chapter 6		97
MCL-1		97
6.1	Apoptosis.....	97
6.1.1	Apoptosis and Cancer.....	98
6.1.2	Apoptotic Pathways.....	98
6.2	BCL-2 Protein Family	101
6.2.1	BCL-2 family protein-protein interactions	102
6.2.2	BCL-2 family proteins as therapeutic targets	102
6.2.3	BH3 mimetic as potential drugs	104
6.2.4	MCL-1 as a drug target.....	105
6.3	MCL-1	106
6.3.1	MCL-1 function.....	108
6.3.2	MCL-1 versus BCL-2 family member's specificity	108
6.3.3	BH3 and interaction with MCL-1	109
6.3.3.1	Position 2d.....	110
6.3.3.2	Position 3a	111

6.3.3.3	Positions 3d.....	111
6.3.3.4	Position 4a	111
6.3.3.5	Positions 3g.....	112
6.3.4	Targeting MCL-1.....	112
6.3.4.1	ABT-737.....	113
Chapter 7		114
MCL-1 Methods		114
7.1	Docking.....	114
7.1.1	Structure preparation	114
7.1.2	Active site.....	115
7.1.3	Dataset for docking.....	115
7.1.3.1	Fluorescence polarization assay.....	116
7.1.4	Molecular Docking	118
7.2	Molecular Dynamics.....	118
7.2.1	System preparation	118
7.2.2	Minimization, heating up and equilibration of system	119
7.2.3	Production run	120
7.2.4	Binding free energy	121
Chapter 8		123
MCL-1 Results and Discussion.....		123
8.1	MCL-1 versus BCL-XL	123
8.2	Docking.....	123
8.3	Molecular Dynamics.....	124
8.3.1	Clustering	124
8.3.2	Binding free energy calculation.....	127
8.3.3	Interactions	127
8.3.3.1	ST_1_046.....	127
8.3.3.2	ST_1_109.....	128
8.3.3.3	ST_1_R1N.....	128
8.3.3.4	ST_1_208.....	131
8.3.3.5	ST_1_247.....	131
8.3.3.6	ST_1_202.....	131
8.3.3.7	ST_1_159.....	132
8.3.3.8	ST_1_249.....	132
8.3.3.9	ST_1_162.....	132
8.3.3.10	ST_1_227 and ST_1_222	134
8.3.3.11	ST_1_261.....	134
8.3.4	Conformation of the residues	134
8.3.5	Comparison between different scaffolds	135
8.3.5.1	Rhodanine	135
8.3.5.2	Thiohydantoin.....	136

8.3.5.3	Hydantoin	137
8.3.5.4	Thiazolidinedione	137
<i>Chapter 9</i>		138
<i>Conclusions</i>		138
9.1	Contributions	138
9.2	Limitations	144
9.3	Future work	145
<i>Bibliography</i>		147

Summary

Drug discovery is a lengthy and complicated process. In order to reduce the time to market, computational methods such as molecular modeling, chemoinformatics and chemometrics have been incorporated successfully in many drug discovery projects. The aim of the study is to contribute to the achievement of Pharmaceutical Data Exploration Laboratory in the field of drug discovery by developing novel drugs against two targets i.e. neuraminidase and MCL-1 and in process learn different methodologies used in computer aided drug design such as QSAR, docking and molecular dynamics. The two targets were selected due to the difference in the nature of the proteins. While neuraminidase has small buried hydrophobic pocket, MCL-1 has long narrow binding site on the surface of the protein. The difference in the active site has its own challenges and can lead to different approaches in computer aided drug design.

Influenza is a contagious viral disease of respiratory tract. The primary drug target for treatment influenza is neuraminidase due to its conserved nature and important role in virus life cycle. Neuraminidase can be divided into two groups i.e. group I and group II. Oseltamivir and zanamivir are two FDA approved drugs for treatment of influenza. Mutations like H274Y, N294S and R292K have already resulted in resistance against oseltamivir and zanamivir. These mutations are group specific e.g. H274Y and N294S belong to group I while R292K is found in group II neuraminidase. Hence, pan neuraminidase inhibitor effective against both groups and as well as wild and mutant strains is required.

To achieve this, consensus QSAR model with applicability domain was developed to screen potential neuraminidase inhibitors. The compounds screened by model were later used in docking study against group I and group II neuraminidase strains along with major mutations i.e. H274Y, N294S and R292K to discover novel pan neuraminidase inhibitors.

The results show that the probable inhibitors had similar orientations as zanamivir and oseltamivir in wild type i.e. N1_closed and N9_closed. As a result of H274Y, the side chain was found to be pushed back thus negating the inward movement of Glu276. The longer side chain was found to be facing away from Glu276 and closer to Ile222, Arg224, Ala246 (N1)/Ser246 (N9). R292K mutation resulted in the constriction of the hydrophobic cavity thereby resulting in rotation of side chain. ZN88 was able to form hydrogen bond between amino group of the side chain and Glu276, Glu277, Asp151 in both wild and mutant strains. The extra flexibility of the side chain in ZN88, ZN33 and ZN35 was due to bifurcation at 1st atom. Thus, it can be concluded that inhibitors having guanidino group, flexible side chain with an amino group can be pan neuraminidase inhibitors. Low SD observed for of ZN43, ZN88, ZN35 and ZN46 indicates less deviation in binding against mutant strains as well as different groups of neuraminidase.

Anti-apoptotic proteins, like BCL-XL, play important roles in apoptosis and have been a target of number of anti-cancer efforts. However, MCL-1 overexpression has been one of the reasons behind the resistance against anti-cancer drugs targeting BCL-XL. In a recent study rhodanine based compounds have shown promise as MCL-1 specific inhibitor. However, compounds with

rhodanine scaffold are known as pan assay interference compounds (PAINS). Hence, the second objective is to analyze the role of rhodanine scaffold in selective inhibition of MCL-1 to guide the development of more potent and selective MCL-1 inhibitors. In order to achieve second objective, our collaborator Miss Tang Shi Qing graduate student Dr. Christina CHAI synthesized compounds belonging to four different classes i.e. rhodanine, thiazolidinedione, thiohydantoin and hydantoin by scaffold hopping.

Molecular dynamics was performed to analyze the interactions of MCL-1 with compounds of different scaffolds in order to improve potency and selectivity of MCL-1 inhibitors. Crystal structure of MCL-1 inhibitors reported in previous studies utilizes mostly one or sometimes two pockets in MCL-1 binding groove. On the other hand, most active compound ST_1_046, belonging to rhodanine scaffold, was found to be aligned with the hydrophobic groove and interacted with pockets P1, P2 and P3. This alignment was supported by non-polar rhodanine ring flanked with electronegative groups. More polar central ring of other scaffolds led to decrease in activity. Thus it was concluded that increase in occupancy of the binding groove, which depends on the electrostatics of ligand, increases the activity.

On the basis of the computational results, five compounds with rhodanine scaffold were synthesized by our collaborators. Analysis of these compounds indicates that further increase in length of the inhibitor does not lead to better activity. Thus in future, compounds with bulkier non-polar central group can be developed which can help to improve the activity to a greater extent.

Both studies have been successful in predicting the probable inhibitors for neuraminidase and MCL-1. Predicted probable neuraminidase inhibitors will be subjected to molecular dynamics study against different mutant strains. ZN43, ZN88, ZN35 and ZN46 will be used to develop pharmacophore model for screening potent pan neuraminidase inhibitors. Recent discovered neuraminidase 10 and 11 will be included for the screening and testing. The effect of the compounds on the human sialidase also needs to be tested in the future.

The knowledge gained from the interaction of the ligands with MCL-1 will be utilized to develop novel selective inhibitors against MCL-1. In-vitro studies will be performed against both MCL-1 and BCL-2 to establish the selectivity of the ligands. As poor results were obtained in docking studies therefore novel algorithms should be developed to target such binding grooves. Despite the fact that molecular dynamics improved the results, there is a need to establish a relation between number and duration of trajectories required for a molecular dynamics experiment to attain a good correlation between predicted binding energy and experimental activity.

List of Tables

Table1.1 Brief overview of some of the common docking software	15
Table2.1 Confusion matrix showing the predictions made by QSAR model	35
Table3.1 Binding cavity residues.....	53
Table4.1 Neuraminidase strains used for docking study	61
Table 5.1 Performance of the base models selected to form consensus model.....	68
Table5.2 Performance of the consensus model.....	70
Table 5.3 Compounds outside the AD of the consensus model.....	71
Table 5.4 Functional group present in the compounds outside of AD.....	74
Table5.5 The final 10 PNI and their ZINC codes	77
Table5.6 Tanimoto coefficient of the PNI against established inhibitors	78
Table5.7 Information related to PNI.....	79
Table 5.8 Binding free energy (kcal/mol) of 10 PNI along with oseltamivir, zanamivir and laninamivir.....	81
Table 5.9 Average binding free energy (kcal/mol) and IC50 (nM) oseltamivir, zanamivir and laninamivir.....	82
Table 5.10 Correlation between IC50 and calculated binding free energy	83
Table 6.1 Physiological role of BCL-2 protein families	103
Table7.1 Dataset for Mcl-1 studies.....	117
Table 8.1 The cluster size of top three clusters is shown.....	124
Table 8.2 Average binding free energy	127

List of Figures

Figure1.1 Drug Discovery Pipeline	7
Figure1.2 Workflow Of Homology Modeling.....	7
Figure1.3 Computer Aided Drug Design.....	7
Figure2.1 General Workflow Of Qsar	23
Figure2.2 K-Nearest Neighbor	24
Figure2.3 Support Vector Machine	24
Figure2.4 Five-Fold Cross Validation.....	24
Figure3.1 General Symptoms Of Influenza (Häggström, 2014).....	45
Figure3.2 Structure Of Influenza Virus (Mackay).....	45
Figure3.3 Overview Of Influenza Virus Life Cycle (Times, 2007).....	46
Figure3.4 Role Of Neuraminidase In Influenza Life Cycle (Can005, 2011).....	46
Figure3.5 Schematic Representation Of Different Ways Causing Virus Mutation (Niaid, 2011).	49
Figure3.6 Neuraminidase Tetramer [2hty]	50
Figure3.7 Neuraminidase Group 1 Monomer Depicting Putative Active Site, 430 Loop And (A) Closed 150 Loop [2hu4] And Open 150 Loop [2hty].....	50
Figure3.8 The First Two Neuraminidase Inhibitors	55
Figure4.1 Overview Of Docking Process.....	64
Figure5.1 Structures Of Oseltamivir, Zanamivir, Laninamivir And Top 5 Pni Accoring To Abfe	85
Figure5.2 Conformation Of Glu276 With Osetlamivir, Zanamivir And Laninamivir As Inhibitors	85
Figure5.3 Comparision Of Oseltamivir And Zanamivir Poses In N1_Closed And N1_N294s.....	85
Figure5.4 Comparison Of Pose Of Oseltamivir, Zanamivir In N1_Closed And N1_H274y.	86
Figure5.5 Comparison Of Poses Of Oseltamivir, Zanamivir In N9_Closed And N9_R29k.	86
Figure5.6 Comparision Of Zn88 And Oseltamivr Pose In N1_Closed And N9_Closed.....	86
Figure5.7 Comparision Of The Interactions Of Zn88 In N1_Closed And N9_Closed.....	89
Figure5.8 Comparision Of Zn33 And Oseltamivr Pose In N1_Closed And N9_Closed.....	89
Figure5.9 Comparision Of Zn35 And Oseltamivr Pose In N1_Closed And N9_Closed.....	89
Figure5.10 Comparision Of Zn21 And Oseltamivr Pose In N1_Closed And N9_Closed.....	90
Figure5.11 Comparision Of Zn46 And Oseltamivr Pose In N1_Closed And N9_Closed.....	90
Figure5.12 Comparision Of The Poses Of Zn88 In Different Strains	90
Figure5.13 A) Comparision Of The Poses Of Zn88 In N1_H274y And N9_R292k B) Interaction Of Zn88 In R292k.....	93
Figure5.14 Comparision Of The Poses Of Zn88 In N1_N294s And N1_H274y	93
Figure5.15 Comparison Of The Poses Of Zn33 In Different Strains	93
Figure5.16 Comparison Of The Poses Of Zn35 In Different Strains	94
Figure5.17 Comparison Of The Poses Of Zn21 In Different Strains	94
Figure5.18 Comparison Of The Poses Of Zn46 In Different Strain.....	94
Figure6.1 The Intrinsic And Extrinsic Apoptotic Pathways (Adapted From (Peter E. Czabotar, Lessene, Strasser, & Adams, 2014; Youle & Strasser, 2008)).	99
Figure6.2 Classification Of Core B-Cell Lymphoma-2 (Bcl-2) Family Proteins On The Basis Of Bcl-2 Homology (Bh) Domains (Adapted From (L. W. Thomas, Lam, & Edwards, 2010))	100
Figure6.3 The Selective Interactions Within Bcl-2 Family Members. (Adapted From (Peter E. Czabotar Et Al., 2014)).....	100
Figure6.4 Bh3 Mimetic Abt-737 And Abt-263	105

Figure6.4 Structure Of Mcl-1	107
Figure8.1 Rmsd Comparison Of The Backbone Atoms Between Five Trajectories	125
Figure8.2 Rmsd Comparison Of The Backbone Atoms Between Five Trajectories Continued.	126
Figure8.3 Orientation Of St_1_046, St_1_109, St_1_R1n, St_1_261, St_1_208.....	129
Figure8.4 Orientation Of St_1_202, St_1_227, St_1_159, St_1_162, St_1_222 And St_1_227	130
Figure8.5 Orientation Of St_1_249 And The Distance Between The Pocket Residues And Closet Atom Of The Pose St_1_046 In 45nst	133
Figure8.6 Comparison Of The Residues Of A3 And A4 And Loop A2-A4 Loop For St_1_046 25nst, St_1_046 45nst, St_1_109 25nst And St_1_R1n 25nst	133

List of Abbreviations

AD	Applicability domain
ADME	Absorption Distribution Metabolism and Excretion
ANN	Artificial neural network
BCL-2	B-cell lymphoma-2
CADD	Computer aided drug design
DANA	2-deoxy-2,3-didehydro-N-acetylneuraminic acid
DOF	Degrees of freedom
FDR	False discovery rate
FN	False negatives
FP	False positives
FPR	False positive rate
GA	Genetic algorithm
gaff	General AMBER force field
GPU	Graphics processor unit
HA	Hemagglutinin
HTS	High-throughput screening
IMS	Inter-membrane space
kNN	k nearest neighbor
M1	Matrix protein
M2	Membrane ion channel protein
MC	Monte Carlo
MCC	Matthew's correlation coefficient
MCL-1	Myeloid cell leukemia-1
MLR	Multiple linear regression
N1	Neuraminidase group I
Neu5Ac2en	N-acetylneuraminic acid
NP	Nucleoprotein
NS1	Nonstructural protein 1
NS2	Nonstructural protein 2
PA	Polymerase acidic protein
PB1	Polymerase basic protein 1
PB2	Polymerase basic protein 2
PDB	Protein data bank
PME	Particle-mesh Ewald
PNI	Probable neuraminidase inhibitors
QSAR	Quantitative structure activity relationship
RPC	RNA dependent RNA polymerase complex
RMSD	Root mean square deviation

SA	Sialic acid
SE	Sensitivity
SP	Specificity
SVM	Support vector machine
TN	True negatives
TP	True positives

List of Publications

1. **Sharma N** and Yap CW* (2012). Consensus QSAR model for identifying novel H5N1 inhibitors. *Molecular Diversity*. 16 (3): 513-524.
2. He YY, Liew CY, **Sharma N**, Woo SK, Chau YT and Yap CW* (2013). PaDEL-DDPredictor: Open-source software for PD-PK-T prediction. *Journal of Computational Chemistry*. 34 (7): 604-610.

List of oral and poster presentations

1. 8th Annual Pharmacy Research Symposium "Integrating clinical practice with advances in biomedical research"
2. Identification of novel inhibitors against neuraminidase using computer aided drug design; 8th PharmSci@Asia Symposium, NUS, June 2012
3. Discovery of Novel Neuraminidase Inhibitor by In-silico Screening Approach; ITB-NUS Pharmacy Scientific Symposium 2013
4. Investigating the Feasibility of Scaffold Hopping Strategy in the Design of Pro-survival Mcl-1 Protein Inhibitors; Annual Pharmacy Research Symposium 2013, NUS
5. Discovery of novel broad range neuraminidase inhibitor by in-silico screening approach; YLLSoM 4th Annual Graduate Scientific Congress 2014
6. Discovery of novel broad range neuraminidase inhibitors: A ligand-based and structured based drug designing approach; Annual Pharmacy Research Symposium 2014, NUS
7. Scaffold Hopping Strategy in the Design of Pro-survival Mcl-1 Protein Inhibitor, 9th PharmSci@Asia2014 (China) Symposium
8. Discovery of Novel Broad Range Neuraminidase inhibitors: A Structured Based Drug Designing Approach, 9th PharmSci@Asia2014 (China) Symposium

Thesis structure

Thesis structure can be divided into four main sections i.e. introduction, methods, neuraminidase and MCL-1. The first section describes the application and importance of CADD in drug discovery process. The components of CADD, especially those applied in our work, are described in chapter 1. The second chapter describes the methods used to achieve our objectives i.e. QSAR, docking and molecular dynamics. The parameters specific to any particular study is described in their respective sections.

The third and fourth sections are divided into three chapters each i.e. introduction, methods, results and discussion. Chapter 3 describes influenza and its life cycle. It also elaborates neuraminidase and its role in the influenza life cycle, thereby making it an appropriate target for influenza inhibition.

The methods used in discovery of neuraminidase inhibitors and parameters specific to it are described in chapter 4. The development of QSAR model and its application to screen ZINC library (J. J. Irwin & Shoichet, 2005; John J. Irwin, Sterling, Mysinger, Bolstad, & Coleman, 2012) along with docking study is explained in this chapter.

Chapter 5 consists of the results and discussion for neuraminidase section. It describes the prediction performance of QSAR, compounds outside the AD of the model and screening of the ZINC library. In addition, the compounds selected as result of docking, their poses in wild and mutant strains are discussed.

The role of apoptosis and its control by BCL-2 protein family is described in chapter 6. This chapter also explains the different role of apoptotic and anti-apoptotic proteins. In addition, the importance of MCL-1 as a drug target is also discussed.

The application of molecular dynamics to predict the poses and understand the dynamics of MCL-1 is described in chapter 7. The use of multiple trajectories to increase the accuracy is also shown. This chapter also highlights the limitation of docking in predicting the accurate pose.

The orientation of compounds predicted by 25ns and 45ns trajectory resulting in MCL-1 inhibition is discussed in chapter 8. This chapter describes the importance of P2 pocket in interaction with ligand. Moreover, the role of electrostatics and scaffold of compounds in determining the activity is discussed.

The last chapter i.e. chapter 9 describes the contributions of the two projects involved in this work and also the limitations and future work.

Chapter 1

Introduction

Computer aided drug design (CADD) is emerging as an important component of drug discovery process as it helps to reduce time to market and cost of the drugs. Traditionally CADD includes ligand-based drug design i.e. quantitative structure activity relationship (QSAR) and structure based drug design i.e. docking. Recently, molecular dynamics emerged as a vital part of the drug discovery process. The first section of this chapter (1.1) describes overview of drug discovery process and application of CADD. The objective and thesis structure are described in 1.5, 1.6 sections respectively.

1.1 Drug discovery process

Drug discovery and development is time-consuming, costly process and risky endeavor. It takes about 15 years and \$1- \$1.5 billion to turn a promising lead compound into a potential drug. Despite the increase in investment in drug discovery, the output is considerably low, mainly due to high rate of drug failure in clinical trials (Allison, 2012). Consequently, in order to reduce the cost and time of a drug to reach market, new technologies were ventured.

With the advancement in areas of genomics and proteomics and development of high-throughput screening (HTS) (Broach & Thorner, 1996; Hertzberg & Pope, 2000), the requirement of new lead compounds was felt. Combinatorial chemistry which can create large population of structurally different compounds became an attractive choice (W. A. Warr, 1997). As combinatorial chemistry grew and was adapted in many research studies, the need for a faster method to screen compounds arise. To cope with these challenges, both experimental and theoretical methods were developed. HTS, for instance, involves screening large libraries of chemicals against a biological target while virtual screening screens large libraries of chemicals computationally and then verifying the predicted compounds in vitro/in vivo (Shoichet, 2004). The purpose of HTS is to speed up the drug discovery process by screening large compound libraries. HTS involves target identification, reagent preparation, compound management, assay development and high-throughput screening which requires great care (Martis E A, 2011) Due to individual biochemical assays with over millions of compounds huge cost and time consumed with HTS (Subramaniam, Mehrotra, & Gupta, 2008). This has led to more faster and effective computational approach i.e. computational virtual screening or virtual screening. In comparison to HTS, virtual screening requires structural information either of ligands (ligand-based virtual screening) or of the target itself (target-based virtual screening) (Ekins, Mestres, & Testa, 2007). Though both virtual screening and HTS are complementary process (Bajorath, 2002), virtual screening gives much higher hit rate (Yun Tang, Weiliang Zhu, Kaixian Chen, & Hualiang Jiang, 2006).

The rapid growth of low-cost computational power in last decades has increased the application of computational technology in the drug discovery pipeline and is known as CADD. CADD is a broad term including different computational tools involved in database, screening potential lead molecules, analyzing the cause of effectiveness or ineffectiveness of a particular drug, modeling and simulation of the compound or biomolecules (Dalkas, Vlachakis, Tsagkrasoulis, Kastania, & Kossida, 2013; Ooms, 2000).

1.2 Computer Aided Drug Design

The general steps of drug discovery can be defined (**Figure1.1**) as disease related genomic, target identification, target validation, lead discovery, lead optimization, preclinical trials and clinical trials (Y. Tang, W. Zhu, K. Chen, & H. Jiang, 2006). Application of computational tools is rapidly gaining implementation in drug discovery and is generally known as CADD (Kapetanovic, 2008). Initially, CADD tools were developed for lead optimization but now they find application in almost all phases of drug discovery (Y. Tang et al., 2006). CADD mainly involves in 1) identification and optimization of new drugs using chemical and biological information of the ligands and structures. 2) filtration compounds with undesirable properties and select most promising compounds (Kapetanovic, 2008; Ou-Yang et al., 2012; Rahman et al., 2012; C. M. Song, Lim, & Tong, 2009).

1.2.1 Target identification

The two main reasons for drug failure are lack of activity against proposed target or its unsafe nature. Hence, target identification and validation is the first and most important stage of any drug discovery process (Hughes, Rees, Kalindjian, & Philpott, 2011). Ideal novel drug targets should be a part of a crucial biological pathway, different from previously known targets, functionally and structurally characterized; and druggable i.e. can bind to small molecules (Bakheet & Doig, 2009). Structure based computational methods have shown promise in predicting targets such as in case of protein kinase inhibitors (Rockey & Elcock, 2006). Potential drug targets have also been identified using inverse docking i.e. docking a compound with a known biological activity against different receptors (Y. Z. Chen & Zhi, 2001) and screening target libraries (Rognan, 2006).

1.2.1.1 Homology Modeling

In absence of experimental structures such as in case of most membrane proteins, homology modeling is used to predict target structure (Cavasotto & Phatak, 2009; Kopp & Schwede, 2004; Elmar Krieger, Nabuurs, & Vriend, 2005). Homology modeling takes advantage of the fact that protein structure is more conserved than sequence and similar sequence have similar structure. Homology modeling is a multistep process (**Figure1.2**) and can be summarized into following steps (Elmar Krieger et al., 2005):

1. Template recognition and initial alignment

2. Alignment correction
3. Backbone generation
4. Loop modeling, Side-chain modeling
5. Model optimization
6. Model validation

Homology works best with sequence identity of more than 40% between the initial sequence and the homologous templates. The accuracy drops considerably as sequence identity drops below 30% (Marti-Renom et al., 2000). Besides dedicated software like Modeller (<https://salilab.org/modeller/>) (Eswar et al., 2007) and multi-utility software Yasara (<http://www.yasara.org/>) (E. Krieger et al., 2009) webserver like SWISS-MODEL (<http://swissmodel.expasy.org/>) (Biasini et al., 2014), are frequently used to predict protein structure. Homology modeling has been utilized in several studies such as in deduction of bovine μ -calpain inhibitor-binding domains (Chai, Lim, Lee, Chai, & Jung, 2014), human muscarinic acetylcholine receptors (T. Thomas et al., 2014), G-protein-coupled receptors (Yarnitzky, Levit, & Niv, 2010) and Human Kynurenine Aminotransferase III (Nematollahi, Church, Nadvi, Gorrell, & Sun, 2014) being some of the recent examples.

1.2.2 Lead Discovery

Once we have a defined target, next step is to find a lead molecule. A lead molecule has at least weak affinity and minimum toxic effects and forms the starting point of the drug like compound (Verlinde & Hol, 1994). Lead structures

should possess following properties: (1) simple chemical features so that they can be easily optimized; (2) have an established structure activity relationship; (3) novel in order to get patent; and (4) good absorption, distribution, metabolism and excretion (ADME) properties (Oprea, Davis, Teague, & Leeson, 2001). Based on the presence or absence of the target structure, lead discovery can be divided into two major class i.e. ligand-based lead discovery and structure-based lead discovery.

1.3 Ligand and Structure based drug design

Depending on availability of structural information of the target, CADD can be divided into two categories (**Figure1.3**) i.e. ligand-based and structure-based CADD. Structure-based CADD relies on the knowledge of the target protein structure to predict potential inhibitors and their binding poses. On the other hand, ligand-based approach utilizes the knowledge of active and inactive compounds to construct quantitative structure-activity relation (QSAR) models for predicting possible ligands (Kalyaanamoorthy & Chen, 2011). Both structure and ligand-based approaches find application in lead discovery as well as lead optimization.

1.3.1 Ligand-based drug design

Ligand-based CADD uses a set of structurally diverse compounds with known activity for a particular target and is based on the hypothesis that compou-



Figure 1.1 Drug Discovery Pipeline

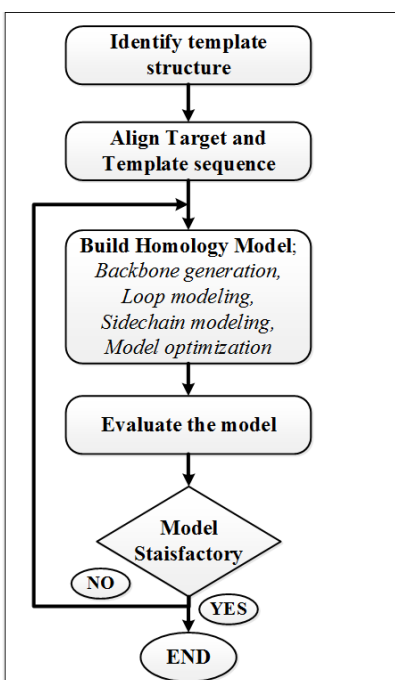


Figure 1.2 Workflow of homology modeling

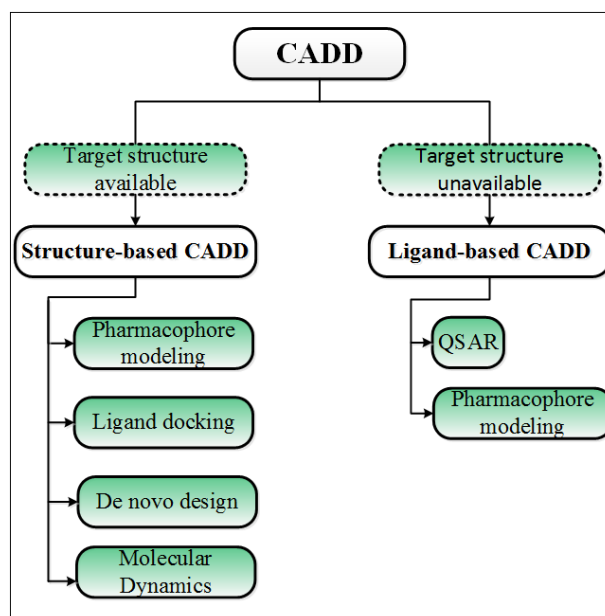


Figure 1.3 Computer aided drug design

-nds with similar structure have similar properties. Based on the presence of either active and/or inactive compounds, ligand-based lead discovery can be divided into two groups (1) selection of compounds based on chemical similarity or (2) the construction of a QSAR model to predict probable lead like compounds.

1.3.1.1 Quantitative structure–activity relationship (QSAR)

QSAR is used often in drug discovery projects to find new lead compounds and works by establishing mathematical relation between structure and function using chemometric method (Kubinyi, 1997; S. Zhang, 2011).

In drug discovery, structure implies physicochemical properties of the compounds; function refers to biological activity and chemometric method includes multiple linear regression (MLR), support vector machine (SVM), artificial neural network (ANN) etc. Since the pioneer work of Hansch and Free-Wilson, a lot of progress has been made in QSAR with the rise of 3D (Verma, Khedkar, & Coutinho, 2010) and even 4D QSAR (Andrade, Pasqualoto, Ferreira, & Hopfinger, 2010).

A QSAR model has following objectives:

1. To identify chemical properties responsible for biological activity.
2. To optimize the existing leads in order to improve their biological activities.
3. To predict the biological activities of novel compounds.

A QSAR workflow consists of creating a combined dataset of active and inactive compounds, calculating descriptors of the compounds present in dataset, splitting dataset into modeling and validation set, creating QSAR model using modeling set and evaluating it by the validation set. The model is used to screen a desired chemical library and accuracy of the screening is judged from the validation results performed earlier. Success of QSAR model not only depends on the dataset but also on the descriptors and methods used for modeling. As effective screening depends on the dataset used for training the model, a diverse dataset can increase the chemical space of the model (Kubinyi, 1997; S. Zhang, 2011).

In addition to the extensive use in predicting the bioactivity, QSAR has also been applied to distinguish drug-like from non-drug-like molecules, explain possible molecular mechanism of the receptor-ligand interactions (G. F. Yang & Huang, 2006). prediction of physicochemical, pharmacokinetic (Xu et al., 2007), and ADMET properties (Klopman, Stefan, & Saiakhov, 2002; Winkler, 2002). Some of the recent application of QSAR in drug discovery are discovery novel GPCR ligands (A. Tropsha & Wang, 2006), inhibitors of acetylcholinesterase in Alzheimer's disease (K. Y. Wong, Duchowicz, Mercader, & Castro, 2012), HIV inhibitors (Debnath, 2005), neuraminidase inhibitors (N. Sharma & Yap, 2012; Zheng et al., 2006) etc.

1.3.2 Structure-based drug design

Unlike ligand-based lead discovery, structure based approach requires protein structure and exploits the protein-ligand interactions to select compounds that bind strongly to the biologically relevant target (Ghosh, Nie, An, & Huang, 2006). In the scenario with no information on the compounds active against the target, *de novo* design approach can be used to identify possible leads (Arakawa, Hasegawa, & Funatsu, 2007). However, *de novo* design is not restricted to a certain condition and can be used whenever a novel lead molecule is required such as in the identification of D816V mutant-selective c-KIT inhibitors (H. Park, Lee, Lee, & Hong, 2014), Aurora A kinase inhibitors (Rodrigues et al., 2013), novel HCV helicase inhibitor (Kandil et al., 2009), and inhibitors of cyclophilin A (Ni et al., 2009), among many others. Structure-based drug design has been successfully applied in many drug discovery projects such as design of GPCR inhibitors (Congreve, Dias, & Marshall, 2014), catechol-O-methyltransferase inhibitors (Ma, Liu, & Wu, 2014), carbonic anhydrase inhibitors (Guzel, Innocenti, Vullo, Scozzafava, & Supuran, 2010), angiotensin-I converting enzyme inhibitors (Anthony, Masuyer, Sturrock, & Acharya, 2012) etc.

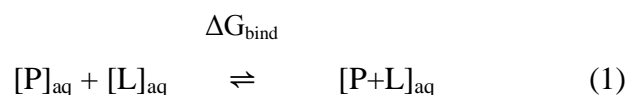
1.3.2.1 Docking

Docking program aims to predict the orientation and conformation of the ligand within the binding site of a receptor. This is achieved by sampling the conformational space of ligand and binding site which are later used to find the binding interactions between ligand and protein. This gives us a score also known

as binding score. The compounds and their respective poses are ranked on the basis of binding score. By this way, docking can predict strength of the interaction of possible lead like molecule prior to its synthesis and *in vitro* or *in vivo* evaluation.

Docking can be divided into of two main steps. In the first step, the algorithm tries to predict the possible binding modes for protein-ligand pair. The aim of the scoring function selected for this step is to roughly distinguish the true binding poses without compromising on speed. The second step involves selection of several poses from the first stage and reevaluating them. The scoring function used in this step is generally more complex and attempts to estimate binding energies as accurately as possible.

Scoring functions are mathematical equations to calculate binding affinity of a ligand towards a receptor. Any protein-ligand interaction can be defined by the equation 1.



The free energy of binding i.e. ΔG is obtained with the Gibbs-Helmholtz equation:

$$\Delta G = \Delta H - T\Delta S \quad (2)$$

Where ΔH is the enthalpy, T is the temperature in Kelvin and ΔS the entropy. The binding constant i.e. K_i can be related with ΔG by following equation:

$$\Delta G = -RT \ln K_i \quad (3)$$

Evaluation and ranking of predicted poses is an important aspect of any docking process. An ideal scoring function should be able to identify true binding poses. An ideal scoring function will be computationally too expensive making them unsuitable for large number of protein-ligand interactions. Therefore every docking program makes its own set of approximations and do not fully account for a number of physical phenomena for example, entropic effects, leading to difference in the results between them (Kitchen, Decornez, Furr, & Bajorath, 2004; Mohan, Gibbs, Cummings, Jaeger, & DesJarlais, 2005). Moreover exhaustiveness of the scoring function can depend on the stage of docking as a less exhaustive scoring scheme is used in pose selection process but more complex scoring scheme is used while estimating binding energies of the selected poses.

Broadly scoring functions can be classified in three different types i.e. force-field based, knowledge based and empirical based scoring functions. The parameters of force field scoring functions such as DOCK (Ewing, Makino, Skillman, & Kuntz, 2001), GOLD (Jones, Willett, Glen, Leach, & Taylor, 1997) are derived from both experimental data and ab initio quantum mechanical calculations. However, a major hurdle lies in the treatment of solvent in ligand binding (Huang, Grinter, & Zou, 2010) . Empirical scoring functions like FlexX (Matthias Rarey, Bernd Kramer, Thomas Lengauer, & Gerhard Klebe, 1996)

estimate the binding affinity of a complex on the basis of a set of weighted energy term. The coefficients are obtained from experimentally determined binding energies and X-ray structural information. Compared to the force field scoring functions, the empirical scoring functions are much faster. On the other hand, dependence on the molecular data sets often yields different weighting factors for the various terms. Hence, terms from differently fitted scoring functions cannot easily be recombined into a new scoring function (Kitchen et al., 2004). Knowledge based scoring functions rely on the information derived from the experimental structures. In comparison to the other two scoring functions, knowledge-based scoring functions have a good balance between accuracy and speed (S. Y. Huang et al., 2010). However, lack of experimental structures can lead to problems (Kitchen et al., 2004). Consensus scoring such as CScore has been utilized to overcome the weakness of individual scoring functions. SYBYL's CScore uses DOCK-like D-Score and GOLD-like G-score which are force field based scoring functions, ChemScore (Eldridge, Murray, Auton, Paolini, & Mee, 1997) an empirical based function and Potential of Mean Force (PMF) (Muegge, 2002) which is knowledge based scoring function.

Despite many attempts to make prediction as accurate as possible, docking methods are still far from being perfect. There are many factors leading to inaccuracy of the docking predictions. The lack of a fast and accurate scoring function is perhaps the most limiting factor (Sousa, Fernandes, & Ramos, 2006). Lack of protein flexibility is another reason for inaccurate predictions by docking. During protein-ligand interaction, protein changes its conformation to achieve

best possible pose of the ligand in a phenomenon known as induced fit. To include the effect of induced fit, flexibility of the protein must be considered. However, including the degrees of freedom (DOF) of a receptor will make docking an even more challenging task. Hence, most docking programs consider ligand as flexible while keeping receptor rigid (Zoete, Grosdidier, & Michielin, 2009). Implicit protein flexibility is achieved in SYBYL by application of soft docking algorithms which work by using a relaxed representation of the molecular surface.

Every year, a number of successful applications of docking are published in literature. Docking has played pivotal role in many drug discovery projects such as development of dipeptidyl peptidase IV (Tanwar, Tanwar, Shaquiquzzaman, Alam, & Akhter, 2014), Sortase A (Uddin & Saeed, 2014), Poly (ADP-ribose) polymerase-1 (Hannigan et al., 2013), HIV protease (Wlodawer & Vondrasek, 1998), neuraminidase (Shan, Ma, Wang, & Dong, 2012), monoamine oxidase (Ferino, Vilar, Matos, Uriarte, & Cadoni, 2012) inhibitors as well as drug molecules against protein kinases and phosphatases (C. F. Wong & Bairy, 2013), potassium ion channels (Dave & Lahiry, 2012), solute carrier transporters (Schlessinger, Khuri, Giacomini, & Sali, 2013) etc.

There are many docking programs that differ in sampling algorithms, the handling of ligand and protein flexibility, and scoring functions (Lyne, 2002). Some of them are mentioned in **Table1.1**.

Table1.1 Brief overview of some of the common docking software			
Software	Algorithm	Scoring function	Success Stories
DOCK (Ewing et al., 2001)	Shape matching (sphere images)	Force field or contact score	(Mohan Sahoo, Chandra Dinda, Ravi Kumar, Panda, & S Brahmkshatriya, 2014)
AutoDock (Morris et al., 1998)	Lamarckian GA	Force field	(Khalid & Paul, 2014; Tan, Khairuddean, Wong, Khaw, & Vikneswaran, 2014)
FlexX (M. Rarey, B. Kramer, T. Lengauer, & G. Klebe, 1996)	Incremental construction	Empirical score	(Saeed, Khan, Rafique, Shahid, & Iqbal, 2014)
GOLD (Jones, Willett, & Glen, 1995)	GA	Empirical score	(Grover et al., 2014)
Glide (Halgren et al., 2004)	Descriptor matching/Monte Carlo (MC)	Empirical score	(H. Sharma et al., 2014)
Fred (McGann, 2011)	Shape matching (gaussian functions)	Gaussian score or empirical scores	(Korošec et al., 2014)
ICM (Abagyan, Totrov, & Kuznetsov, 1994)	Monte Carlo	Mixed force field Empirical score	(Hu et al., 2014)
MOE (MOE)	Monte Carlo simulated annealing	Empirical score	(Abdellatif, Belal, & Omar, 2013; Allen et al., 2013)
SYBYL (SYBYL-X)	Incremental construction	Consensus score	(Dutta Gupta et al., 2014; Jayanthi et al., 2014)

1.3.2.2 Molecular dynamics

Though studies based on crystal structure played a major role in drug discovery projects, the static nature of the proteins have led drug designers to look for computational techniques, such as molecular dynamics, to study systems more dynamically (Durrant & McCammon, 2011). Molecular dynamics simulations,

developed in the late 1970s, is based on Newtonian physics and the general workflow can be described in the following steps (Durrant & McCammon, 2011; H.-J. Huang et al., 2010) 1) the determination of initial positions and velocities of every atom; 2) the calculation of forces applied on the investigated atom using inter-atomic potentials; 3) move the atoms to the position defined by forces calculated in step 2; 4) simulate for a short time period and move to step 2.

Classical molecular dynamics is based on Newton's laws of motion (eq 1) which are integrated in time dependent manner.

$$\mathbf{a}_n = \mathbf{f}_n/m \quad (1)$$

from time n to $n+1$ the velocity and coordinates changes according to following :

$$\mathbf{v}_{n+1} = \mathbf{v}_n + \mathbf{a}_n \Delta t$$

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \mathbf{v}_n \Delta t + \frac{1}{2} (\mathbf{a}_n \Delta t^2)$$

where v is velocity, x is atomic coordinate and Δt is magnitude of the integration time step.

These steps are repeated for long duration resulting in the trajectory of the atomic motions during that period (Beck & Daggett, 2004).

Force field refers to a mathematical formula and associated parameters that describes the potential energy of the protein as a function of its atomic coordinates. AMBER, CHARMM, GROMOS and OPLS-AA are the most commonly used force fields. The force fields are associated with molecular

dynamics suites and thus the choice of force field is a secondary one. There is no consensus which force field is better and often simulations performed with same parameters on different force fields generate consistent results (Hug, 2013; Price & Brooks, 2002).

Molecular dynamics is generally considered as the simulation of all atoms present in the system. This is called as full atomistic simulation. Despite the increase in computational power and use of Graphics Processor Unit (GPU) in molecular dynamics, a full atomistic simulation for longer durations is a daunting task. Hence, different variants of classical molecular dynamics has been designed such as temperature accelerated molecular dynamics, replica exchange molecular dynamics, steered molecular dynamics, coarse grained molecular dynamics etc. (Hug, 2013; Kerrigan, 2013).

MD simulation has two broad applications. First is to analyze the actual dynamics of the system thereby observing the motion of biomolecules at the atomic scale for example, folding/unfolding of peptides or small proteins. The second application is to derive equilibrium and kinetic properties of the system and compare them with experiments to interpret the molecular mechanisms behind a particular biological activity (X. Cheng & Ivanov, 2012).

Some of the recent studies involving molecular dynamics are defining binding mode of phosphoinositide 3-kinase α -selective inhibitor (Bian et al., 2014), analysis of the active site of enzyme mannosyltransferase in *Leishmania major* (Shinde, Mol, Jamdar, & Singh, 2014), analysis of interaction of the

inhibitors against adipocyte fatty-acid binding protein (J. Chen, Wang, & Zhu, 2014), discovery of Hsp90 inhibitors (Li et al., 2014), analysis of CDK2 inhibitors (Tripathi & Singh, 2014) etc.

1.4 Lead optimization

Advances in combinatorial chemistry and HTS have resulted in tremendous increase in number of lead molecules (Chaturvedi, Decker, & Odinecs, 2001). This has made lead optimization a much required step which aims to identify compounds with increased likelihood of success in clinical trials (Korfmacher, 2003). Lead optimization involves chemical modification of promising lead molecules in order to improve potency, selectivity, metabolism and pharmacokinetic parameters (K. C. Cheng, Korfmacher, White, & Njoroge, 2008; Hughes et al., 2011).

Different docking methodologies reproduce the crystallographic binding pose to near perfection but struggle while docking novel ligand to the pocket. Hence, lead optimization can be achieved by accurate prediction of receptor-ligand binding affinities and poses. One of the most commonly used methodologies is the application of molecular dynamics to the selected lead like molecules in order to predict poses and binding free energy. It has been found that rescoring poses generated from docking increases the correlation with the experimental results (Guimarães & Cardozo, 2008; Lindstrom et al., 2011).

1.5 Objective

Pharmaceutical Data Exploration Laboratory (PaDEL) excels in development and application of methods and tools in the biomedical and pharmaceutical fields. The research in PaDEL can be divided into drug discovery, clinical informatics, public health informatics and metabonomics. This study intends to contribute in the drug discovery project leading to discovery of novel drugs against neuraminidase and MCL-1.

The first objective is to discover pan neuraminidase inhibitors. Majority of the drug discovery projects on neuraminidase have focused on a single mutation belonging to either group I or group II neuraminidase. However, mutations causing resistance against oseltamivir or both oseltamivir and zanamivir are not restricted to any specific group. Thus, there is need to develop inhibitors which can be effective against neuraminidase irrespective of mutation or group. The first step is to build QSAR model to screen probable neuraminidase inhibitor. A number of QSAR models have been developed but to the best of my knowledge none of them has considered mutations belonging to both groups of neuraminidase. Moreover, most of the QSAR models built till date lack applicability domain thereby having low reliability. Hence, a consensus QSAR model with applicability domain will be built to screen probable neuraminidase inhibitors. Structure based drug design is an important part of most of the drug discovery projects. Hence, the second aim to achieve the first objective is to do structure based screening using various docking protocols. This will help us to

find a possible solution to the resistance in neuraminidase and help to develop even more potent drugs in future.

Apoptosis plays an important role in cancer and most of the tumors have increased levels of anti-apoptotic proteins. This has led to the development of drugs targeting anti-apoptotic proteins. However, challenging binding site of the anti-apoptotic proteins have led to only few moderately active inhibitors. Moreover, the resistance to these inhibitors due to overexpression of MCL-1 has been a cause of concern. Many studies have been performed to develop selective MCL-1 inhibitors. In recent study compounds with rhodanine scaffold have shown promise as MCL-1 inhibitor. However, the function of the rhodanine scaffold is not understood. Hence, the second objective is to analyze the role of rhodanine scaffold in selective inhibition of MCL-1 in order to guide the development of more potent and selective MCL-1 inhibitors. In order to achieve the second objective the compounds belonging to four different classes i.e. rhodanine, thiazolidinedione, thiohydantoin and hydantoin were prepared by scaffold hopping. The substituents of the compounds were selected for direct comparison with the compounds having rhodanine group. This comparison will help us to understand the role of rhodanine in MCL-1 inhibition as well as develop potential drugs in future.

The above mentioned objectives of the study will provide exposure to the different aspects of the computer aided drug design such as QSAR, docking, molecular dynamics. Moreover, two different targets, with different type of

pockets, i.e. deep and hydrophobic, shallow and surface groove, will help me to get experience in handling different problems.

Chapter 2

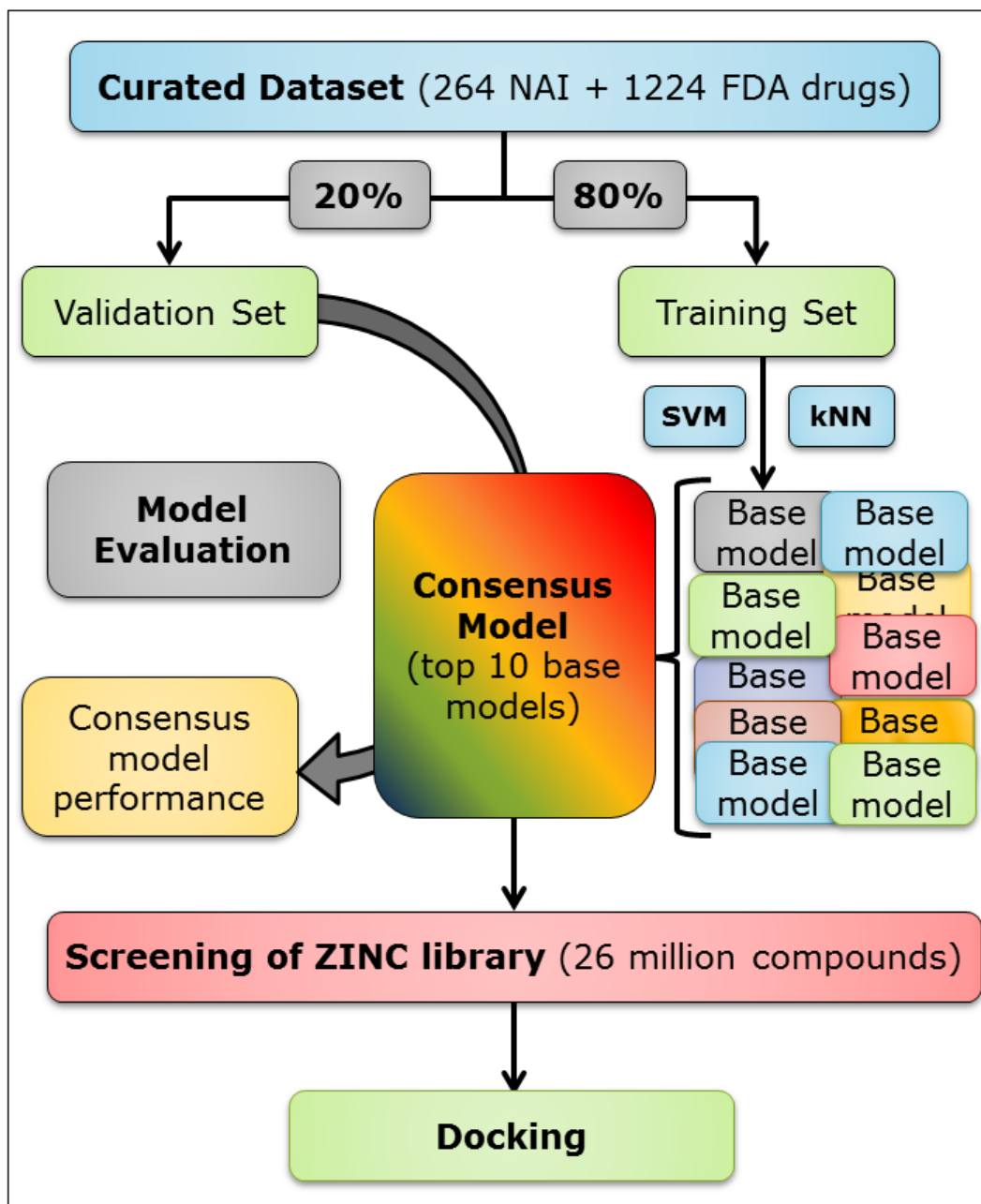
Methods

This chapter describes three possible components of CADD: QSAR (2.1), docking (2.2) and molecular dynamics (2.3). These sections also describe the methodology applied in our study. The specific parameters, if any, used during a particular study are mentioned in respective sections.

2.1 QSAR

In our study, the QSAR modeling consisted of following steps (**Figure2.1**)

1. Dataset selection and curation
2. Descriptor calculations
3. Descriptor selection
4. Base model development
5. Internal validation
6. Consensus model development
7. External validation
8. Prediction i.e. Screening of chemical library

**Figure2.1** General workflow of QSAR

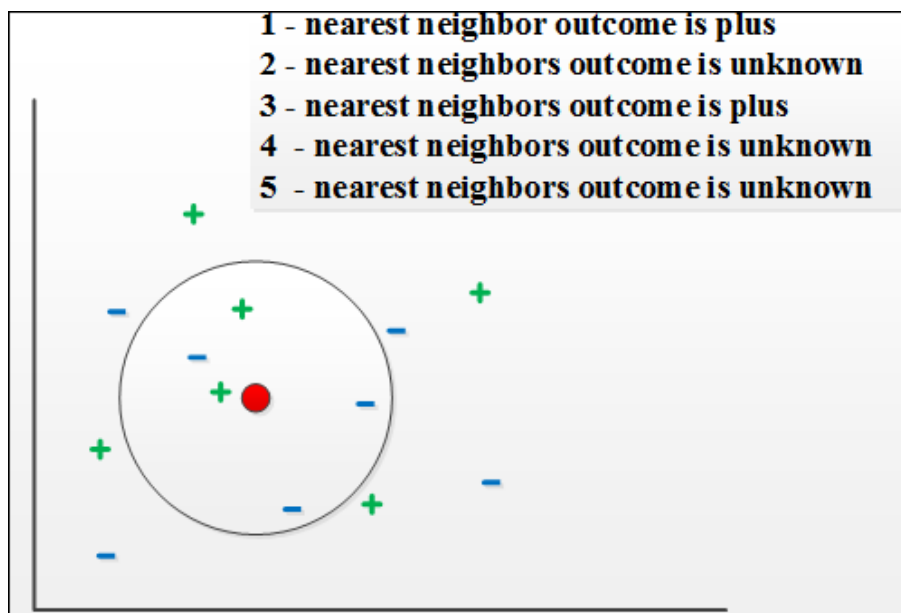


Figure2.2 k-nearest neighbor

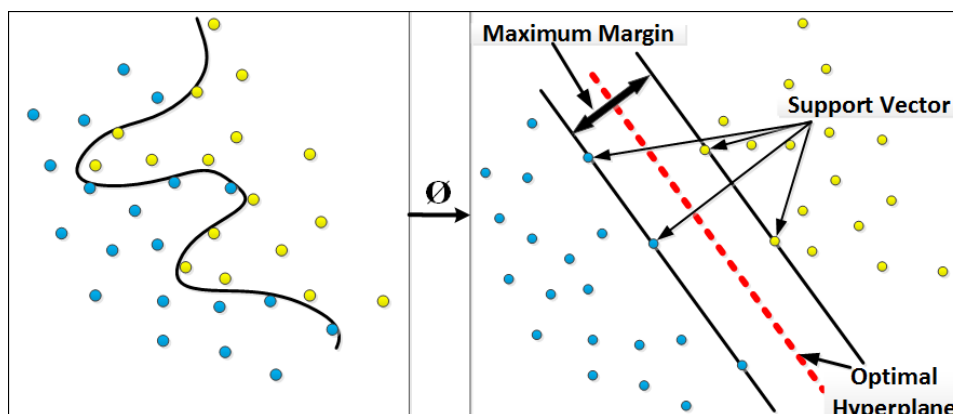


Figure2.3 Support Vector Machine

Kernel function (Φ) SVM converts non-linear classification into linear classification

	Stage 1	Stage 1	Stage 1	Stage 1	Stage 1
Compounds 1-10	Training set	Training set	Training set	Training set	Test set
Compounds 11-20	Training set	Training set	Training set	Test set	Training set
Compounds 21-30	Training set	Training set	Test set	Training set	Training set
Compounds 31-40	Training set	Test set	Training set	Training set	Training set
Compounds 41-50	Test set	Training set	Training set	Training set	Training set

Figure2.4 Five-fold cross validation

2.1.1 Data selection and curation

QSAR model depends on the descriptors, which in turn depends on the structures of compounds in the dataset. Since accuracy of any model depends on the quality of data, any error in the structure will translate into error in the descriptors, leading to inaccurate models.

Some of the steps followed in our study for data curation are removal of the data that cannot be handled by cheminformatics techniques, e.g., inorganic and organometallic compounds, counterions, salts and mixtures; structure validation; ring aromatization; analysis of tautomeric forms; and removal of duplicates (Alexander Tropsha, 2010).

2.1.2 Descriptor calculation

Molecular descriptors are key features of compounds in mathematical values that describe the structure or shape of molecules in order to predict the activity and properties of molecules (Todeschini & Consonni, 2000). The choice of descriptors plays an important role in ligand-based lead discovery. Descriptors can be classified according to 1) their dimension and 2) the properties of the chemical that they represent. According to dimension, descriptors can be classified as 1D, 2D and 3D. 1D descriptors are calculated from the formula itself e.g. number of atoms, molecular weight and 2D descriptors are calculated from 2D structure e.g. number of hydrogen bonds and 3D descriptors are calculated from 3D structure e.g. molecular volume, surface area etc. Descriptors derived on

the basis of property used such as physicochemical e.g. logP and pKa, electronic e.g. polarity and polarizability, topological e.g. connectivity and Weiner index, geometrical e.g. length, width and molecular volume etc and complex e.g. BCUT, WHIM and 3D-MoRSE (Nikolova & Jaworska, 2003; Todeschini & Consonni, 2009; Xue & Bajorath, 2000)

2.1.3 Descriptor selection

2.1.3.1 Pre-processing

A good QSAR model requires a subset of relevant descriptors. However, before selecting a subset of descriptors, a pre-processing step is performed. The descriptors were normalized to ensure that all attributes have equal influence on the model.

A large number of descriptors are not recommended (Dudek, Arodz, & Gálvez, 2006) (Shahlaei, 2013) as:

1. Only some of the descriptors are significantly correlated with the activity
2. Many descriptors are intercorrelated.
3. Prediction accuracy of model might be improved through exclusion of redundant and irrelevant descriptors.
4. Including too many descriptors may lead to overfitting of the model

5. With less number of descriptors the interpretability of relationship between the descriptors and observed activity might be increased.

Hence, an appropriate subset of the descriptors was identified by feature selection to develop the model.

2.1.3.2 Selection

Descriptor/feature selection is a step to reduce the dimensionality by removing irrelevant data. This can lead to increased learning accuracy, and improved result comprehensibility (Lei Yu & Liu, 2003). Feature selection can be supervised or unsupervised. Supervised feature selection can be divided into three general categories i.e. filter methods, wrapper methods and embedded methods (Guyon & Elisseeff, 2003; Saeys, Inza, & Larranaga, 2007). Wrapper method works by selecting subset of the features by search algorithm and then testing the performance of the selected subset by the modeling algorithm (Soto, Cecchini, Vazquez, & Ponzoni, 2008). The performance is tested by cross validation or validation set. The feature subset with best performance is selected to build the model. Forward selection, backward elimination, simulated annealing, tabu search, genetic algorithm (GA) are some of the examples using this methodology (Gammerman, 2014; Goodarzi, Dejaegher, & Vander Heyden, 2012). Wrapper methods use modeling algorithm as a blackbox thus making it simple and universal in approach (Guyon & Elisseeff, 2003). Hence, in our study, GA, which is one of the wrapper methods, was used for descriptor selection.

2.1.3.2.1 Genetic Algorithm

GA belongs to the class of evolutionary algorithms and is based on laws of genetics and biological evolution and can be divided into following steps (Leardi, 2007; Thede, 2004):

1. Initialization: Create large population of random descriptor subsets, represented by binary chromosomes that indicate whether a descriptor is selected.
2. Evaluation: Each chromosome is evaluated and assigned a fitness score. Fitness score is a measure of how well a set of descriptor can predict the properties.
3. Selection: In order to improve the fitness of the population two members are selected from the current population. The chance of being selected is proportional to the chromosomal fitness.
4. Crossover: Crossover is performed to exchange descriptors between two members hoping to achieve descriptor set with better prediction power
5. Mutation: In order to add a little bit randomness into the population, mutation is performed. It is performed by changing the binary value of few descriptors selected randomly.
6. Repeat: Repeat step 2 to 5 until a new population of N members has been created.

2.1.4 Model development

The QSAR model that we developed in our study used the following procedure: (1) Construct many base models using different descriptor subsets and different modeling methods; (2) Select suitable base models to construct a consensus model; (3) Evaluate the performance of the consensus model.

Model development began by splitting the dataset into 4:1 ratio. The four parts were combined to form training set while the 5th part comprised external validation set. Training set was used to develop base models. Base models are individual predictive models that were combined to form consensus model. In this study, base models were constructed from the training set using different descriptor subsets and different modeling methods. The different descriptor subsets were obtained using a GA process. During the GA process, models were developed and evaluated by calculating their Matthew's correlation coefficient (*explained in section 2.1.5.3*) (Vihinen, 2012) using a 5-fold cross validation (*explained in section 2.1.5.1*) process in order to identify relevant descriptors. The GA process was repeated several times while the performance, i.e. Matthew's correlation coefficient (MCC), of each model being evaluated using 5-fold cross validation. All the models having 70% MCC were added to the model pool.

In general, methods for building a QSAR model i.e. modeling methods can be divided into classification and regression methods. Classification method is used to predict the class of the members like inhibitor/non-inhibitor of the dataset screened by QSAR. On the other hand, regression analysis estimates

relationships between dependent (response or outcome) variable and one or more independent (predictor or input) variables and is used for predicting outcomes such as IC_{50} (Gramatica, 2010; Yee & Wei, 2012). For regression analysis, absolute value and minimum error are required. However, this is not available in most of the biological activity assay results, for example inactive compounds are mentioned as having IC_{50} greater than 100 μm . Hence, in our study, classification methods such as kNN and SVM were used for the development of base models.

2.1.4.1 k nearest neighbor

k nearest neighbor (kNN) is one of the easiest modeling methods and is the first choice when there is little or no prior knowledge about the distribution of the data. kNN works by measuring the distance between the test compound to every compound in the training set. The two most common measures to calculate distance are Euclidean distance and Manhattan distance. The class represented by majority of the k number of closest members is returned as the class of test compound (**Figure2.2**) (Yee & Wei, 2012).

As kNN decision is based on a small neighborhood of similar objects it performs well with multi-modal classes. However, treatment of all the features equally can lead to classification errors, especially when only a small subset of useful features is present (J. Kim, Kim, & Savarese, 2012).

2.1.4.2 Support Vector Machine

SVM is one of the most well-known kernel based method for model development. It is based on the concept of decision planes that define decision boundaries. In a binary classification problem the data can be linearly separable or nonlinearly separable. For a linearly separable data, SVM builds a maximal margin hyperplane to separate the two classes. While for nonlinearly separable data, SVM uses a kernel function to map the vectors into a higher dimensional feature space (**Figure2.3**) in order to make them linearly separable (Yee & Wei, 2012).

A main advantage of SVM is that it performs well on datasets having many attributes, even with few samples for the training process. However, main drawbacks of SVM are limitation in speed and size and the selection of the kernel function parameters (J. Kim et al., 2012).

2.1.4.3 Applicability domain (AD)

All base models had inherent AD which is among the five benchmarks set by OECD for any standard QSAR model (Gramatica, 2007). AD has been considered important part of model development by many studies and can be defined as the chemical space where a model can predict properties of compounds with certain accuracy (Jaworska, Aldenberg, & Nikolova, 2004; Jaworska, Nikolova-Jeliazkova, & Aldenberg, 2005; Netzeva et al., 2005; Alexander Tropsha & Golbraikh, 2007; Weaver & Gleeson, 2008).

AD is the chemical space where a model can be used to predict the properties of compounds (Jaworska et al., 2005; Alexander Tropsha & Golbraikh, 2007; Weaver & Gleeson, 2008). In this study, the AD of each base model was defined using the method proposed by Fumera et al (Giorgio Fumera, Fabio Roli, & Giacinto, 2000). This multiple thresholds method uses a value that is computed by modeling methods to specify the confidence for each prediction. Different modeling methods have different algorithms to compute this confidence value. For example, in kNN, the confidence value for predicting a compound as an inhibitor is computed as the proportion of k nearest neighbors of the compound that are inhibitors. So if k is 5 and the number of nearest neighbors of the compound that are inhibitors is 4, the confidence value that the compound is an inhibitor will be 0.8. Usually in a binary classification modeling method, a threshold of 0.5 for the confidence value is used such that if the confidence value is more than 0.5, the compound will be predicted as an inhibitor. Otherwise, it will be predicted as a non-inhibitor. In the multiple thresholds methods, two thresholds are used such that if the confidence value is greater than the higher threshold value, the compound will be predicted as an inhibitor. Conversely, if the confidence value is smaller than the lower threshold value, the compound will be predicted as a non-inhibitor. When the confidence value lies between the two threshold values, the compound will be considered as out of the AD of the model and its activity will not be predicted.

In this study, the two thresholds were determined using the confidence values of the compounds in the testing sets of a 5-fold cross validation. The

confidence values were sorted and those that were found in both inhibitors and non-inhibitors, or those that indicate a transition between inhibitors and non-inhibitors were identified as potential thresholds. All combinations of threshold pairs from the pool of potential thresholds were then tested. The optimum threshold pair was identified using three criteria. The first is the accuracy of the model for those compounds identified as out of the AD should be the lowest when using the optimum threshold pair. The second criterion is the precision of the model for those compounds identified as within the AD should be the highest when using the optimum threshold pair. The last is the number of compounds identified as out of the AD should be the largest when using the optimum threshold pair. The three criteria were applied consecutively. If only one threshold pair satisfied the first criterion, the process was stopped and that pair was identified as the optimum pair. If more than one threshold pairs satisfied the first criterion, the second criterion was applied. The third criterion was used only when more than one threshold pairs satisfied the second criterion. Ties for threshold pairs satisfying the third criterion, if any, were broken randomly.

2.1.5 Validation

Validation is important in order to make a QSAR prediction reliable. Depending upon the dataset used for validation, it can be defined as internal and external validation. Though both internal and external validation are important, it is shown that external validation can improve the reliability of QSAR prediction (Golbraikh & Tropsha, 2002; Alexander Tropsha, Gramatica, & Gombar, 2003).

2.1.5.1 Internal validation

Internal validation is performed by a subset selected from the training set. Some of the methods used for internal validation are random subsampling, bootstrap and cross-validation. Cross validation or n-fold cross validation is most popular among them where n is the number of portions in which training set is distributed.

In our study, 5-fold cross validation was performed (**Figure2.4**). The steps involved in 5-fold cross-validation are as follows:

1. Divide the dataset into 5 equal portions
2. Combine 4 portions to form the training set to build the model
3. Use the remaining portion as a testing set to assess the predictive performance of the model
4. Repeat steps 2 and 3 until all the portions has been used as a testing set once
5. Get the average performance of the five testing sets

2.1.5.2 External validation

Unlike internal validation, the external validation dataset is never used in model development. Moreover, it was shown that higher internal validation is not necessarily sufficient for accurate prediction of model. In addition, emphasis was given on external validation as only way to trust the reliability of QSAR (Golbraikh & Tropsha, 2002; Alexander Tropsha, 2010).

2.1.5.3 Predictive performance

Prediction by QSAR model can be defined as follows:

Table2.1 Confusion matrix showing the predictions made by QSAR model		
	Predicted positive class	Predicted negative class
Positive class	True positives (TP)	False negatives (FN)
Negative class	False positives (FP)	True negatives (TN)

All the base models were then ranked based on their cross-validated MCC value and the top ten base models were selected to construct a consensus model (*explained in section 2.1.6*). The performance of the consensus model was then evaluated using the validation set. The performance was assessed in terms of TP, TN, FP and FN, SE, SP, false positive rate (FPR), false discovery rate (FDR). The overall performance was measured by Q and MCC.

False positive rate: FPR defines expected proportion of FP among compounds belonging to negative class.

$$FPR = \frac{FP}{FP + TN} * 100\%$$

False discovery rate: FDR defines expected proportion of FP among compounds belonging to positive class.

$$FDR = \frac{FP}{FP + TP} * 100\%$$

Sensitivity: SE relates to the model's ability to identify a compound belonging to positive class correctly.

$$\text{Sensitivity} = \frac{TP}{TP + FN} * 100$$

Specificity: SP relates to the model's ability to identify compound belonging to negative class correctly.

$$\text{Specificity} = \frac{TN}{TN + FP} * 100$$

Overall accuracy: is also a statistical measure of how well a QSAR correctly identifies compounds belonging to a particular class. Accuracy is the proportion of true results of both TP and TN in the population.

$$Q = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$

Matthew's correlation coefficient: The MCC is used to measure the quality of binary (two-class) classifications and is regarded as a balanced measure which can be used even with classes of different sizes. MCC is a correlation coefficient between the observed and predicted binary classifications.

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TN + FP) * (TN + FN) * (TP + FP) * (TP + FN)}}$$

In our study, MCC values were used to rank the base models and the top ten base models were selected to construct a consensus model.

2.1.6 Consensus model

Every modeling method has its own strength and weakness. Combining the prediction of the different models reduces the risk of wrong selection by a single poor model, enables the use of different models trained with different portions of the data, reduces risk of overfitting by individual models, supports individual model to be trained on different type of data. In order to have a good consensus model, the individual models should be as unique as possible. Diverse models can be created by different modeling methods, modeling methods with different parameters, different feature selection method etc. In our study, all base models were ranked based on their cross-validated MCC value and the top ten base models were selected to construct a consensus model.

The AD of the consensus model was defined based on the prediction of the base models. Compounds were defined to be out of the AD of the consensus model when all the base models identified the compound to be out of their AD, or if there was a tie in the predictions (i.e. an equal number of base models predicted the compounds to be inhibitors and non-inhibitors). Otherwise, the compounds were defined to be within the AD of the consensus model and were predicted based on majority voting of the base models. In addition, confidence values for the predictions were also computed using a similar algorithm as described earlier using kNN.

2.2 Docking

Docking is the most important component of the structure based drug designing which has been supported by the increase in high-resolution 3D structures and techniques like homology modeling techniques along with the improvements in docking and scoring technologies (Tuccinardi, 2009). Molecular docking has two major roles i.e. to predict the binding mode and to predict the binding affinity of a complex (Huang & Zou, 2010). In our study docking was used as a means to screen probable inhibitors and deduce their binding poses. Docking method in our study involved the following steps i.e. receptor structure preparation, identification of active site, ligand structure preparation, docking and analysis of the results.

2.2.1 Receptor Preparation

The 3D structure was downloaded from the protein data bank (PDB) (Berman et al., 2000) and prepared for the docking process. The preparation involved deletion of natural ligand and water, addition of missing hydrogen atoms, termini treatment, correction of protonation states according to a particular pH, energy minimization of the hydrogen atoms keeping heavy atoms fixed.

2.2.2 Identification of active site

The active site was selected according to the data present in literature. If the active site is not explicitly mentioned in literature, residues within 6Å of natural ligand were selected as active site.

2.2.3 Ligand preparation

Ligands were prepared within MOE or SYBYL which included prediction of ionization states, generation of tautomers, conformational isomers and energy minimization.

2.2.4 Docking

Optimized ligands were docked against the receptor using MOE or SYBYL. Parameters for the docking were selected according to the problem. Ligands were ranked according to the scores generated by the docking. The selected poses predicted for the ligands were analyzed to deduce reason behind the biological activity of the ligand.

2.3 Molecular Dynamics

All biological systems are dynamic in nature and cannot be understood merely looking at their static structure which is just a snapshot from real dynamics (X. Cheng & Ivanov, 2012; Durrant & McCammon, 2011). Molecular dynamics was performed in our study as it enables to understand the interactions in real time which can give more insights leading to better understanding of the biological system. In this study, molecular dynamics was performed by following steps:

1. System Preparation
2. Minimization

3. Heat up system
4. Equilibration
5. Production run

2.3.1 System Preparation

System preparation is the most important step for any process. A system with defects like atomics clashes, missing atoms and improper charge can result in dubious results. The aim of system preparation is to:

1. Add missing atoms via homology modeling
2. Modify ionization according to the desired pH
3. Preparation of parameter files for novel residues and molecules
4. Fix bond
5. Add hydrogen
6. Add counterions to neutralize the system

2.3.2 Minimization

Energy minimization is the first real step in molecular dynamics protocol. The purpose of energy minimization is to adjust the system according to a particular force field, uniform distribution of solvent (in case of explicit solvent) and counter-ions and remove any steric clashes between atoms. It was performed in two stages:

1. Energy minimization of water molecules and counter-ions with restraints on the protein and ligand was first performed

2. In second step, the entire system was minimized.

Mostly, energy minimization end up finding local minimum structure. To avoid local minimum the system is heated in subsequent step.

2.3.3 Heating up the system and equilibration

The initial positions of the atoms taken from crystal structure have velocity corresponding to absolute zero temperature. The velocity reassigned randomly is not an accurate method. Hence, the system is heated from 0K to 310K over a period of time thus reassigning velocities to particular temperature. The gradual increase of the temperature ensures that system has uniform temperature.

2.3.4 Production run

Molecular dynamics can be performed by using three different canonical ensembles i.e. microcanonical ensemble (constant N,V,E), canonical ensemble (constant N,V,T) and isothermic-isobaric ensemble (constant N,P,T). Any ensemble can be selected according to the problem in hand. All biological process occurs at nearly constant pressure (e.g., atmospheric pressure). Thus, isothermic-isobaric ensemble seems like a reasonable choice. In our work production run was performed at NPT conditions. As molecular dynamics is a random process, multiple trajectories were generated for a conclusive result.

Chapter 3

Neuraminidase

This chapter provides an introduction to influenza virus and neuraminidase. The role of neuraminidase in virus life cycle and its importance as drug target is also discussed. Influenza, commonly called as flu, is a communicable respiratory viral disease that affects mainly nose, throat, bronchi and, occasionally lungs. Although flu is often confused with other influenza-like illnesses such as common cold during the early phase of infection, it is a more severe disease and is caused by a completely different virus (Apisarnthanarak et al., 2004; Nicholson, 1992; Yuen et al., 1998).

3.1 Influenza virus

Influenza viruses are enveloped viruses belonging to the family Orthomyxoviridae. The influenza virus can be distinguished into three categories, influenza A, B and C on the basis of antigenic differences between their matrix and nucleoprotein (NP) (Lamb & King, 2001).

Influenza C is less common and Influenza B is confined to human and rarely infects other species (Taubenberger & Morens, 2008). Although both influenza A and B co-circulate and have been responsible for epidemics, the impact of influenza A has been much higher than influenza B. This is because the

lower rate of genetic variability, combined with its limited host range reduces the chances of pandemics caused by influenza B (Hay, Gregory, Douglas, & Lin, 2001).

3.1.1 Influenza A

Influenza A virus is the most virulent of all three due to greater genetic variability and host range such as, such as humans, birds, horses, dogs and pigs. Influenza A can be sub-divided into different subtypes according to antigenic properties of Hemagglutinin (HA) and Neuraminidase glycoproteins. Numerous combinations can be found by the combination of 16 HA (H1–H16) and 9 neuraminidase (N1–N9) glycoproteins. The neuraminidase is further classified into two phylogenetic groups based on sequence analysis: N1, N4, N5 and N8 are in group I, while N2, N3, N6, N7 and N9 in group II (Russell et al., 2006).

3.1.2 Structure of Influenza A virus

The influenza virus consists of a lipid membrane derived from the host cell comprising of three surface proteins, the HA, neuraminidase and membrane ion channel (M2) protein (**Figure3.2**). Enclosed inside the lipid membrane are internal proteins which includes NP, the matrix protein (M1) and the RNA dependent RNA polymerase complex (RPC) composed of polymerase basic protein 1 (PB1), polymerase basic protein 2 (PB2) and polymerase acidic protein (PA) and two nonstructural proteins, nonstructural protein 1 (NS1) and nonstructural protein 2 (NS2) (Swayne, 2008).

HA is a trimer composed of a globular domain and a stem domain. HA has receptor binding site responsible for the attachment of the virus to the cell sialic acid (SA) receptors and fusion peptide responsible for fusion of the viral and cell membrane (Wiley & Skehel, 1987).

Neuraminidase is an enzyme made up of four co-planar and roughly spherical and identical subunits. Each subunit consists of head, which possesses enzymatic activity, along with a centrally attached stalk. The stalk is embedded in the viral membrane by a hydrophobic region. The enzyme is an exo-glycohydrolase and cleaves α -ketosidic linkage between terminal sialic acid and an adjacent sugar residue (Air & Laver, 1989; Colman, 1994).

Membrane ion channel protein (M2) is a single-pass membrane protein. The function of M2 is to reduce the pH across the viral membrane to allow the fusion of viral and endosome membrane. (Pielak & Chou, 2011).

3.1.3 Virus life cycle

Like all viruses influenza virus need the biological machinery of the host cell for its replication (**Figure3.3**). To achieve this, the virus needs to enter the cell. The attachment of influenza virus to the cell is initiated by the interaction between the HA and SA receptors present on cell surface. The attachment is followed by the internalization of the virus by the process of endocytosis. During

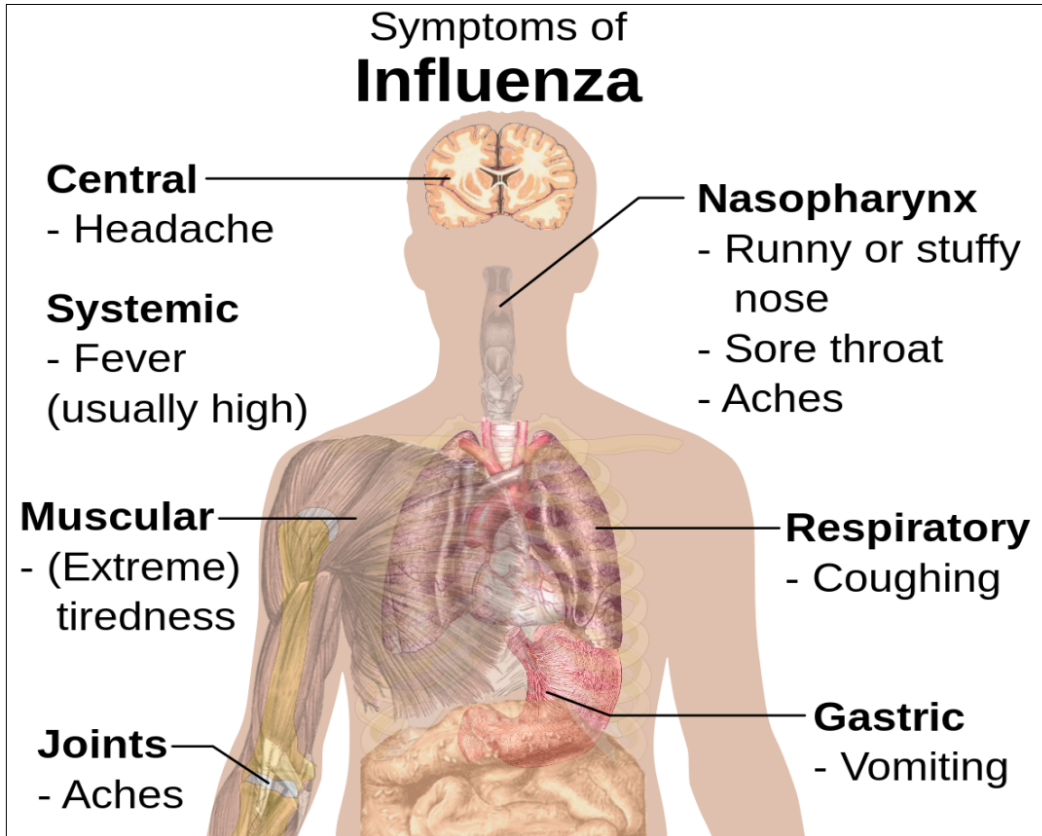


Figure3.1 General symptoms of Influenza (Hägström, 2014)

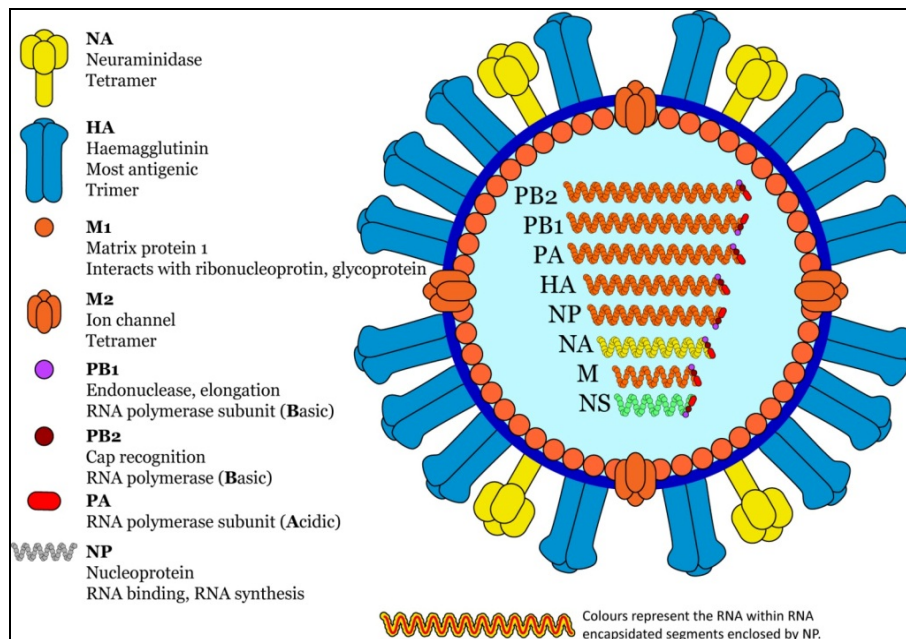


Figure3.2 Structure of Influenza Virus (Mackay)

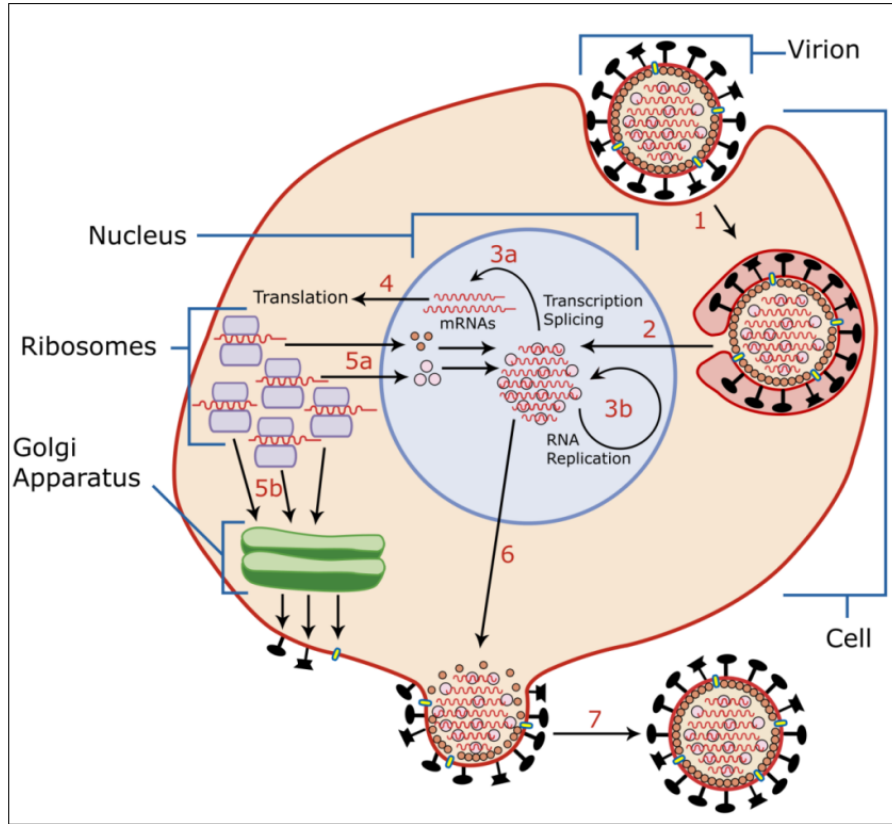


Figure 3.3 Overview of influenza virus life cycle (Times, 2007)

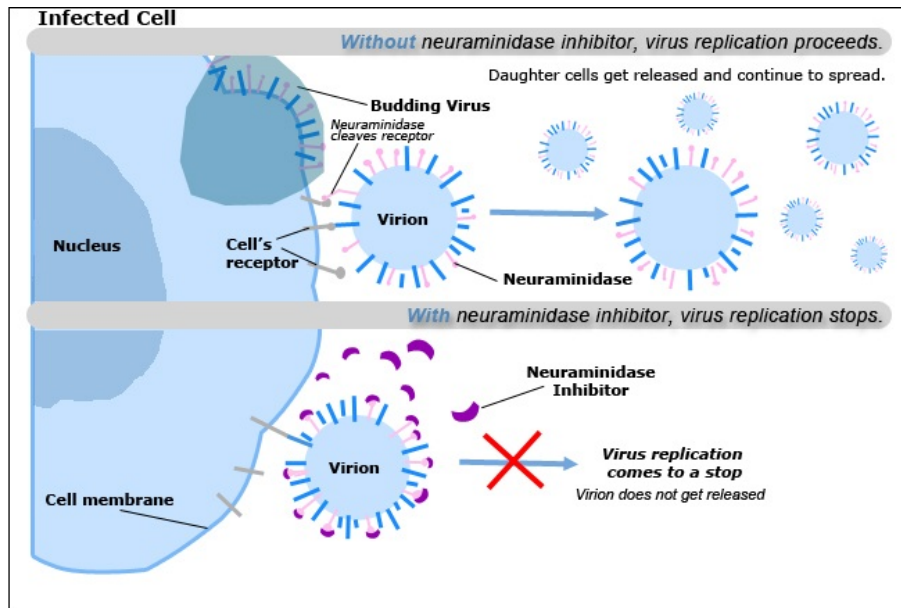


Figure 3.4 Role of neuraminidase in influenza life cycle (Can005, 2011).

Neuraminidase plays an important role in virus life cycle and is required for release of budding virion from the host cell. Neuraminidase achieves that by cleavage of bond formed between hemagglutinin and sialic acid receptors

the process of endocytosis drop in pH lead to the fusion of the membrane and unpacking of the genome (Ari, 1992; Pinto & Lamb, 2006). The fusion domain causes the fusion of viral and endosome membrane followed by the release of RNP into the cellular cytoplasm leading to generation of virus progeny (Fodor & Brownlee, 2002; Pielak & Chou, 2011). The HA that facilitates the viral entry inhibits the release of virion by attaching again to the SA receptors. Neuraminidase cleaves of $\alpha(2-6)$ ketosidic linkage between a terminal SA and an adjacent sugar residue thus releasing the budding virions and blocking them to get clumped with each other (**Figure3.4**).

3.1.4 Antigenic variation

Antigenic variation in influenza viruses occurs by mainly two mechanisms (**Figure3.5**) (Hay et al., 2001; Nelson & Holmes, 2007) i.e. point mutations (antigenic drift) and gene reassortment (antigenic shift).

3.1.4.1 Antigenic Drift

Antigenic drift is caused due to the accumulation of point mutations in viral proteins. The high rate of mutation is the result of low accuracy of RNA-dependent RNA polymerase. The viruses having certain advantages over other are selected. Eventually this leads to changes sufficiently large enough to avoid existing antibodies. Antigenic drift happens while a single type of virus passes through the host.

3.1.5 Antigenic Shift

The genetic shift occurs when two viruses, having property to infect two different hosts, co-infect a single host e.g. swine. The selectivity of a virus depends on the type of SA acid linkage present in a host. The SA has $\alpha(2,6)$ linkage in humans and have $\alpha(2,3)$ linkages in case of avians and equines while swine has SA with both type of linkages (Kimble, Nieto, & Perez, 2010). As genome of the influenza virus is segmented, reassortment of the genome between two strains of viruses results in a new strain. This new strain can have the capability to infect human but due to genetic reassortment is significantly different to avoid immune response.

3.1.6 Characteristic function of Neuraminidase

The role of neuraminidase is to catalyze the cleavage of $\alpha(2-6)$ - or $\alpha(2-3)$ -ketosidic linkage that exists between a terminal SA and an adjacent sugar residue. This removal of SA has two major effects. Firstly, it assists in the mobility of virus in the respiratory tract as well as facilitates the release of virion progeny from infected cells. Secondly, the removal of SA from the carbohydrate moiety of newly synthesized HA and neuraminidase is necessary to prevent self-aggregation of the virus after release from host cells (Air & Laver, 1989; Gong, Xu, & Zhang, 2007). In addition, neuraminidase plays an important part in secondary infection such as bacterial pneumonia possibly by the destruction of respiratory epithelium; virus-induced immunosuppression; and inflammatory response to viral infection (Gong et al., 2007).

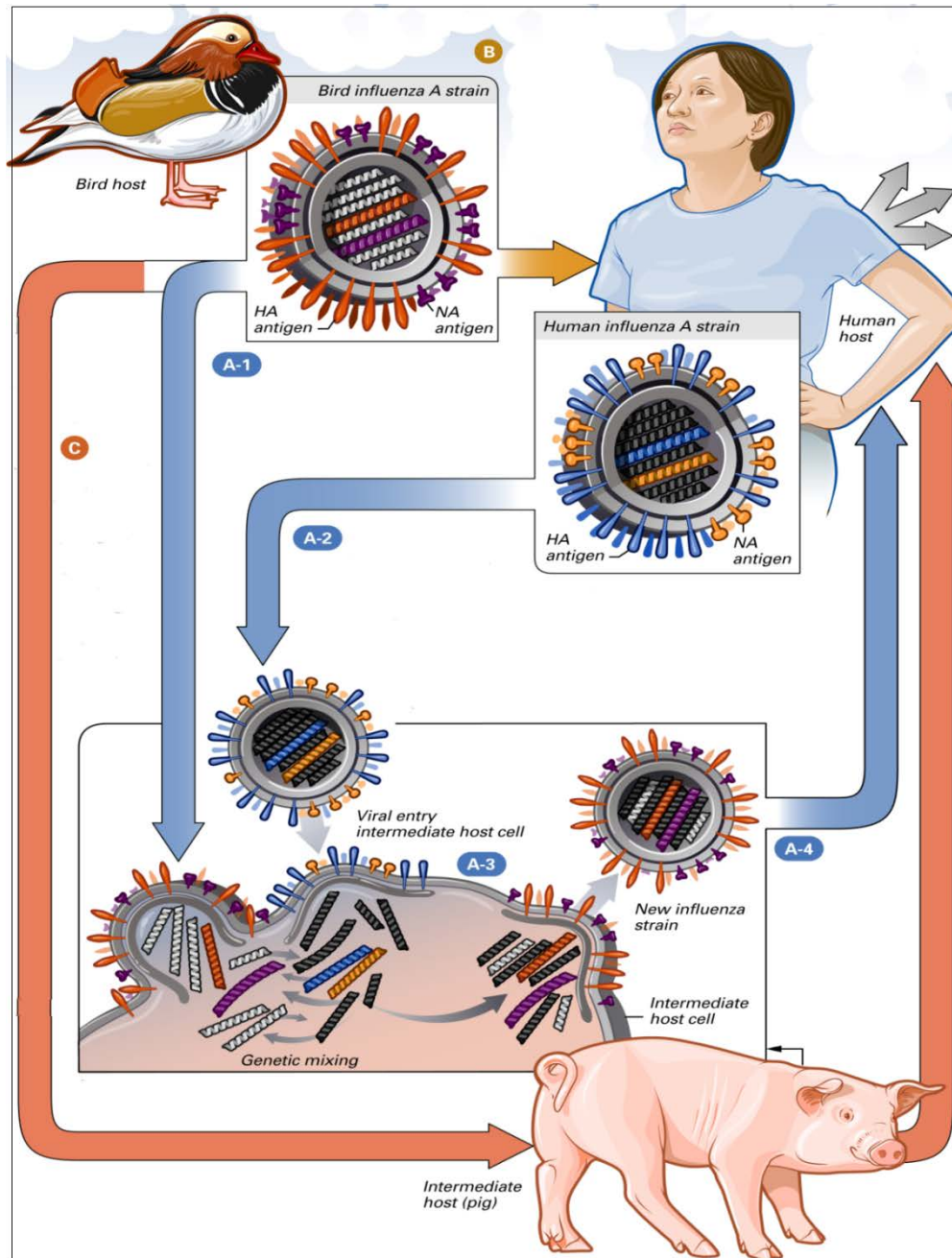


Figure 3.5 Schematic representation of different ways causing virus mutation (NIAID, 2011). In step B the virus is transferred from the bird to the human directly. The virus leaving the human have gained some mutation by the process of genetic drift. Genetic drift also leads to mutation on path C where virus is first transferred to swine and then later to humans. However, the virus specific to bird (A-1) and human (A-3) infect common host swine in step A-3. This leads to new strain of virus having capability to infect human but with different antigen to avoid immune system by the process of genetic reassortment.

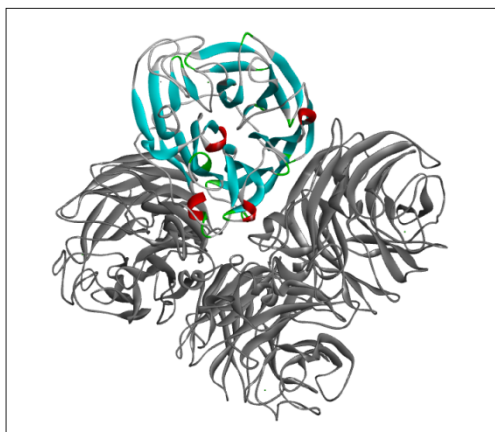
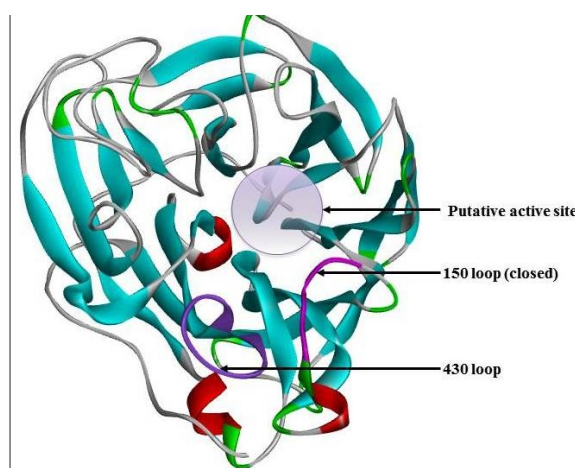
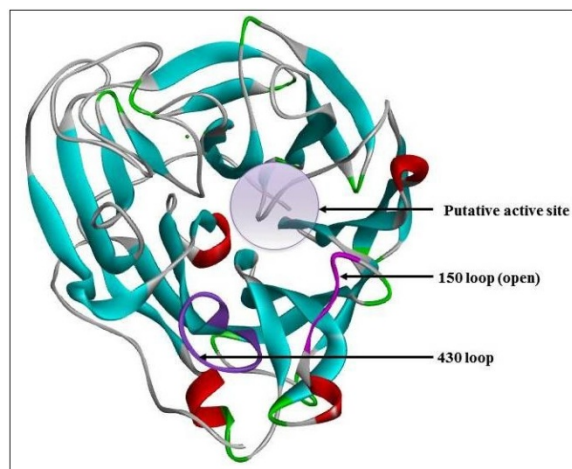


Figure3.6 Neuraminidase tetramer [2HTY]



a)



b)

Figure3.7 Neuraminidase group 1 monomer depicting putative active site, 430 loop and (a) closed 150 loop [2HU4] and open 150 loop [2HTY]

3.1.7 Neuraminidase as a drug target

M2 ion channel allows the uncoating of virus particles in endosome and thus plays a crucial part in virus life cycle. Considering its important role, M2 ion channel was the first target for influenza treatment. However, the M2 inhibitors were effective against influenza A infection as only the A strains of the virus have M2 ion channel proteins. Moreover, cases of CNS side effects and drug-resistant viral strains were reported against M2 inhibitors (Q. Liu, Liu, & Yang, 2013; Mark von Itzstein, 2007). This resulted in the hunt for other possible targets for influenza treatment. Neuraminidase is another protein that plays crucial role in virus life cycle and has been considered as a prime target for influenza inhibitors.

As can be seen from **Figure3.4** neuraminidase inhibitors halts the release of budding influenza viruses from infected host cells thereby preventing infection of new host cells and interrupting the infection (Moscona, 2008). Studies have revealed that despite low sequence identity, the structure is well conserved among group I and II as well as neuraminidase of influenza B virus. Moreover, the active site is more or less conserved among all neuraminidase (Russell et al., 2006).

3.1.8 Structure of neuraminidase

The structure of neuraminidase was deduced by Varghese, Laver and Colman for N2 subtype (J. N. Varghese, Laver, & Colman, 1983). According to them, neuraminidase is a mushroom-shaped tetrameric protein (**Figure3.6**), attached to the viral membrane by slender stalk composed of hydrophobic sequence at N-terminus. The enzyme active site and calcium binding domain are

situated in the head of neuraminidase. The neuraminidase's head region is propeller-like structure formed by six blades. Each blade is formed of four anti-parallel β sheets stabilized by disulfide bonds. Blades are connected by loops of variable length and structure. (Air, 2012; Gong et al., 2007; Shtyrya, 2009)

3.1.9 Active site of neuraminidase

There is up to 75% sequence variation but the active site is highly conserved in most influenza A and B viruses and there are large numbers of charged residues in the pocket enclosing SA binding site also known as SA cavity (Gong et al., 2007). However, group-1 and group-2 neuraminidase differ from each other centered on the 150-loop (residues 147–152). A major consequence of these differences in structure is that there is a large cavity (150-cavity) adjacent to the active site in group-1 but not in group-2 neuraminidase (**Figure3.7**). This leads to an increase in width of the active site cavity by about 5Å in group-1 viruses (Russell et al., 2006).

It has been found in the apo simulations of group-1 neuraminidase that motion is coupled to an outward movement of the adjacent 430-loop (residues 430-439). This had led to two additional cavities in group-1 neuraminidase besides SA cavity i.e. 150-cavity and 430 cavity which has been explored to identify potential residues for novel drug interactions (Amaro et al., 2007; L. S. Cheng et al., 2008; Landon et al., 2008).

Out of the residues comprising the active site, eight amino acids i.e. Arg118, Asp151, Arg152, Arg224, Glu276, Arg292, Arg371, and Tyr406 are

referred as catalytic residues due to their critical role for functioning and their direct contact to the substrate, while eleven others i.e. Glu119, Arg156, Trp178, Ser179, Asp198, Ill222, Asp227, Asp277, Asn294, and Asp425 are known as framework residues due to their primary role in stabilization of the active site (Ferraris & Lina, 2008). Catalytic residues, except Arg224, are in direct contact with sialic acid and form polar contacts. Arg224 forms nonpolar contact with the glycerol moiety of the N-acetylneuraminic acid (Neu5Ac2en) (Shtyrya, 2009).

Table3.1 Binding cavity residues	
SA cavity	Arg118 , Glu119, Leu134 , Val149*, Lys150*, Asp151 , Arg152, Ser153, Pro154, Arg156 , Trp178, Ser179, Ser195, Gly196, Ile222, Arg224, Glu227, Ser246, Glu276, Glu277, Arg292, Asn294, Tyr347 , Arg371 , Tyr406
150 cavity	Val116, Ile117, Arg118 , Leu134 , Thr135, Gln136, Ser145, Gly147, Thr148, Val149*, Lys150*, Asp151 , Arg156 , Arg430 , Pro431 , Ile437 , Trp438 , Thr439
430 cavity	Asn325, Pro326, Tyr347 , Asn369, Ser370, Arg371 , Trp403, Ser404, Tyr406 , Ile427, Arg428, Gly429, Arg430 , Pro431 , Lys432, Glu433, Ile437 , Trp438 , Thr439

Binding cavity residues are sorted according to the location in different cavities. Residues in bold face participate in interactions in more than one cavity across cavity boundaries

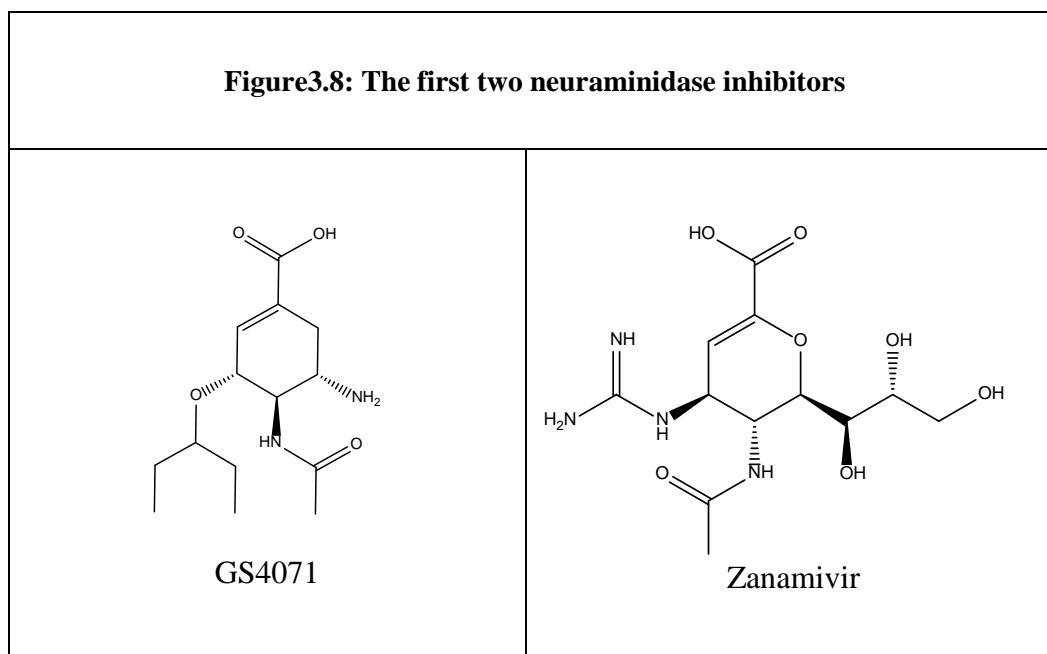
On the basis of the occurrence of residues at SA cavity, 150 loop or 430 loop Cheng et al. classified the residues as shown in **Table3.1**. While some residues are restricted to a particular loop or pocket, other are shared between two cavities. Moreover residues such as Val149 and Lys150 are present in 150 cavity but form a part of SA cavity as 150 loop attains closed conformation.

3.1.10 Neuraminidase inhibitors

The mechanism of neuraminidase action proceeds via formation of a sialosyl cation intermediate. This catalytic intermediate is subsequently released as α -Neu5Ac (Mark von Itzstein, 2007). C-1 carboxylate group α -Neu5Ac and Neu5Ac2en forms charge–charge interactions with positively charged arginine triad i.e. Arg118, Arg292 and Arg371 (M. von Itzstein & Thomson, 2009). As a result, a number of Neu5Ac2en mimetic was discovered as potential neuraminidase inhibitors. Among these, 2-deoxy-2,3-didehydro-N-acetylneuraminic acid (DANA) proved to be the most potent inhibitor. The deduction of apo and holo crystal structures with ligands α -Neu5Ac and Neu5Ac2en turned out to be critical in the discovery of neuraminidase inhibitors (Colman, Varghese, & Laver, 1983; J. N. Varghese et al., 1983; J. N. Varghese, McKimm-Breschkin, Caldwell, Kortt, & Colman, 1992). By applying structure based drug design, the most potent neuraminidase inhibitor 4-deoxy-4-guanidino-Neu5Ac2en i.e. zanamivir was discovered. However, due to its poor bioavailability, intraoral inhalation was chosen for drug delivery (Mark von Itzstein, 2007; M. von Itzstein & Thomson, 2009).

The discovery of zanamivir provided valuable details and laid the foundation of other neuraminidase inhibitors such as GS4071 (**Figure3.8**). The crystal structure of GS4071 revealed that the orientation of the cyclohexene ring and key interactions involving the carboxylate, amino and acetamido substituents with the neuraminidase active site are similar to those observed for Neu5Ac2en

and its derivatives (C. U. Kim et al., 1997; C. U. Kim et al., 1998). The pentyloxy side chain occupied the region utilized by glycerol group in zanamivir. This was possible due to the reorientation of Glu276 thus creating a hydrophobic cavity. However, GS4071, similar to zanamivir, had poor bioavailability and thus its pro-drug was used. This ethyl ester pro-drug is known as oseltamivir (Mark von Itzstein, 2007; M. von Itzstein & Thomson, 2009).



3.1.11 Drug resistance

Both zanamivir and oseltamivir resemble the natural substrate. It was believed that due to this close resemblance, they are less likely to face resistance. However, several mutations have been observed to cause resistance against one or both of them (Abed, Baz, & Boivin, 2006; Ferraris & Lina, 2008; Richard et al., 2008). Currently three major neuraminidase mutations i.e. H274Y, N294S in N1 and R292K in N9 have been reported (Collins et al., 2008; Yan Wu et al., 2013).

The R292K mutation have been shown to compromise viral growth and reduced neuraminidase catalytic efficiency (Yan Wu et al., 2013) but the H274Y and N294S mutations are stably maintained. The high resistance against oseltamivir due to H274Y and R292K in N1 and N9 respectively has caused major concerns (Collins et al., 2008; Woods, Malaisree, Long, McIntosh-Smith, & Mulholland, 2013; Yan Wu et al., 2013). A number of studies has been conducted to deduce the mechanism behind the oseltamivir and zanamivir resistance due to various mutations in order to develop better neuraminidase inhibitors (Chachra & Rizzo, 2008; Collins et al., 2009; Aeron C. Hurt, Lowther, Middleton, & Barr, 2010; Karthick & Ramanathan, 2014; Malaisree et al., 2009; Mihajlovic & Mitrasinovic, 2008; J. W. Park & Jo, 2009; Shu et al., 2011; Vergara-Jaque et al., 2012; Woods et al., 2012). However, not much success has been achieved and pan neuraminidase inhibitor remains a dream.

Chapter 4

Neuraminidase Methods

This chapter describes methods used to achieve our goal to discover pan neuraminidase inhibitors. Consensus QSAR model (4.1) with AD was built to screen ZINC library for probable neuraminidase inhibitors (PNI). The probable inhibitors were docked (4.2) against group I and group II neuraminidase with open, closed conformations. The three most important mutations, i.e. H274Y, N294S and R292K, resulting in resistance against all three influenza drugs i.e. oseltamivir, zanamivir and laninamivir were also used.

4.1 QSAR

4.1.1 Dataset curation

The literature was extensively surveyed to obtain as many structurally diverse inhibitors of influenza A neuraminidase group I (NA1) with their respective IC₅₀ values. A total of 279 NA1 inhibitors were obtained (Abed, Nehme, Baz, & Boivin, 2008; Dao et al., 2011; Dao et al., 2010; Honda, Masuda, Yoshida, Arai, Kaneko, et al., 2002; Honda, Masuda, Yoshida, Arai, Kobayashi, et al., 2002; A. C. Hurt et al., 2007; C. U. Kim et al., 1998; Lew et al., 2000; Lew et al., 1998; A. L. Liu, Wang, Lee, Wang, & Du, 2008; Y. Liu et al., 2011;

Masuda, 2003; Rao et al., 2010; Shie et al., 2007; Smith et al., 1998; Sun, Zhang, Huang, & Zhou, 2006; Wen et al., 2010; Williams et al., 1997; Yamashita et al., 2009). The 2D structures of the inhibitors were drawn using ChemDraw Pro 12.0 (ChemDraw, 2011) and then processed using Pipeline Pilot student edition (Pipeline, 2011), following the general guidelines suggested by recent reviews on the importance of data curation in QSAR modeling work (Dearden, Cronin, & Kaiser, 2009; Fourches, Muratov, & Tropsha, 2010). The process started with the removal of compounds containing one or more metal atoms as the majority of chemical descriptors cannot be calculated reliably for such compounds. The protonation states of common functional groups were set according to pH 6.5, which is the standard assay conditions for neuraminidase assays (L. S. Cheng et al., 2008; Garozzo, Timpanaro, Stivala, Bisignano, & Castro, 2011). Compounds with several tautomeric forms were standardized to a single tautomer. The compounds were then energy minimized using the default settings in Pipeline Pilot energy minimization component. Duplicate compounds were identified by generating canonical SMILES string for each compound and removing those compounds with duplicate canonical SMILES string. For enantiomeric pairs, the enantiomer with the higher IC_{50} value was removed because only 1D and 2D chemical descriptors will be calculated and such descriptors cannot discriminate between the enantiomers. This reduced the number of NA1 inhibitors to 264.

The inhibitors were then divided into two groups based on their IC_{50} values. A total of 173 compounds with IC_{50} values $\leq 10 \mu\text{M}$ were placed in the potent inhibitors group, and 91 compounds with IC_{50} values $> 10 \mu\text{M}$ were placed

in the weak inhibitors group. In order to increase the diversity and size of the dataset, 1326 drugs from FDA Orange Book were obtained and processed in the same manner as the NA1 inhibitors. Drugs that were known neuraminidase inhibitors were removed. A similarity check protocol based on Tanimoto coefficient in Pipeline Pilot was also performed to remove those drugs that were very similar to the NA1 inhibitors. The remaining 1224 drugs were considered as non-inhibitors and were combined with the weak inhibitors group to form the weak/non-inhibitors group (henceforth referred to as non-inhibitors for brevity).

A training set and a validation set were formed from the two groups by splitting the compounds in the groups in a ratio of 8:2. The training set was used to develop the QSAR models and contained 138 potent inhibitors and 1052 non-inhibitors. The validation set was used to determine the prediction performance of the final QSAR model and contained 35 potent inhibitors and 263 non-inhibitors.

4.1.2 Descriptor calculation

A total of 672 1D and 2D different molecular descriptors were calculated using PaDEL-Descriptor v2.7 (Yap, 2011). These include descriptors from different classes such as autocorrelation descriptors (e.g. autocorrelation (charge)); chi indices descriptors (e.g. chi chain); electrotopological state indices descriptors (e.g. atom type electrotopological state); BCUT descriptors; constitutional descriptors (e.g. weight, ring counts); topological descriptors (e.g. Zagreb index, Wiener numbers). Descriptors with zero standard deviation in their values were then removed.

4.1.3 Development of QSAR model and screening

The models were built according to the method defined in chapter 2 with following parameters. For kNN, different k values were used and these include 3, 5, 7, 9, 11 and 13. For SVM, γ used were 0, 0.02, 0.0002, 0.000002, 0.00000002 with $C = 100000$.

In order to identify novel neuraminidase inhibitors, the ZINC library containing 26 million compounds were screened with the consensus model.

4.2 Docking

4.2.1 Structure preparation

Both group I and II neuraminidase crystal structures were downloaded from PDB. These include N1 crystal structures with open and closed loop, H274Y and N294S mutation (henceforth referred as N1_open, N1_closed, N1_H274Y, and N1_N294S respectively) and N9 crystal structures (N9_closed, N9_R292K). The neuraminidase strains used in this study and their corresponding PDB files can be found in **Table4.1**.

Calcium has an important role in the enzymatic activity of neuraminidase and its use in docking studies has been suggested (Lawrenz et al., 2010). Among the structures used in the study only N1_closed and N1_N294S lacked calcium ion. Hence, a calcium ion was added to N1_closed and N1_N294S by superimposing it to N1_open and placing a calcium ion in the same location as in

Table4.1 Neuraminidase strains used for docking study				
Name	PDB ID	Susceptibility towards oseltamivir	Susceptibility towards zanamivir	Susceptibility towards laninamivir
N1_open	2HTY (Russell et al., 2006)	Susceptible	Susceptible	Susceptible
N1_closed	2HU4 (Russell et al., 2006)	Susceptible	Susceptible	Susceptible
N1_H274Y	3CKZ (Collins et al., 2008)	Resistant	Susceptible	Susceptible
N1_N294S	3CL2 (Collins et al., 2008)	Low resistant	Susceptible	Susceptible
N9_closed	7NN9 (Joseph N. Varghese, Chandana Epa, & Colman, 1995)	Susceptible	Susceptible	Susceptible
N9_R292K	4MWL (Yan Wu et al., 2013)	Resistant	Low resistant	Low resistant

N1_open. Ligands and water molecules were deleted from PDB structures and the resulting structures were prepared using MOE and SYBYL (SYBYL-X). Structure preparation includes structure correction, protonation and energy minimization. The protonation states were determined at pH 6.5, which is the standard assay condition for neuraminidase assays (L. S. Cheng et al., 2008; Garozzo et al., 2011). The calculation of partial charges and energy minimization

of only hydrogen atoms was performed using AMBER99 within MOE and SYBYL.

4.2.2 Active site

In our study, residues described by Lily et al., (L. S. Cheng et al., 2008) were used defined active site. For receptors with closed 150 loop, residues of SA cavity i.e. R118, E119, L134, V149, K150, D151, R152, S153, P154, R156, W178, S179, S195, G196, I222, R224, E227, S246, E276, E277, R292, N294, Y347, R371, Y406 constituted the active site. In addition, residues from 150 and 430 cavity i.e V116, I117, T135, Q136, S145, G147, T148, N325, P326, N369, S370, W403, S404, I427, R428, G429, R430, P431, K432, E433, I437, W438 and T439 were used to define the active site for the structures with open 150 loop.

4.2.3 Dataset for virtual screening

Compounds predicted as PNI as result of screening by QSAR model was used for virtual screening (N. Sharma & Yap, 2012). In addition, oseltamivir and zanamivir, the two well-known neuraminidase inhibitor, as well as laninamivir were added to the library.

4.2.4 Molecular docking

Docking was performed in two stages i.e. filtering compounds that are by a less exhaustive approach and exhaustive search against various mutant strains (**Figure4.1**). It has been mentioned in many studies that 150 loop closes after binding to ligand which enables it to make tighter interaction with the ligand

(Russell et al., 2006; M. Wang et al., 2011; Yan; Wu et al., 2013). Hence, the preliminary filtering was performed using N1_open. In the first stage docking was performed using MOE and SYBYL. The resulting poses were sorted according to the docking score and top 50% of the ligands from each were combined to form a non-redundant library of compounds. High penalty was opted for ligands leaving the active site while docking using SYBYL. The side chains were kept free during refinement of the poses while docking with MOE.

Before moving on to second stages of docking, drug-like filters were applied to remove the compound with more likeness to be, for instance, bioavailable. The remaining compounds went through second stage of docking. Docking was performed against N1_closed, N1_H274Y, N1_N294S, N9_closed and N9_R292K using the more rigorous Surflex-GeomX protocol in SYBYL. Compounds with score lower than oseltamivir were filtered at every stage of docking.

4.2.5 Energy minimization and rescoring

Considering the amount of time required only top 10 PNI were selected for energy minimization. The three best poses for each PNI, selected by visual inspection of docking results, were used for energy minimization. Partial charges and force field parameters for the compounds were generated automatically by antechamber suite in AMBER12 (D.A. Case, 2012). The general AMBER force field (gaff) (J. Wang, Wolf, Caldwell, Kollman, & Case, 2004) was used for com-

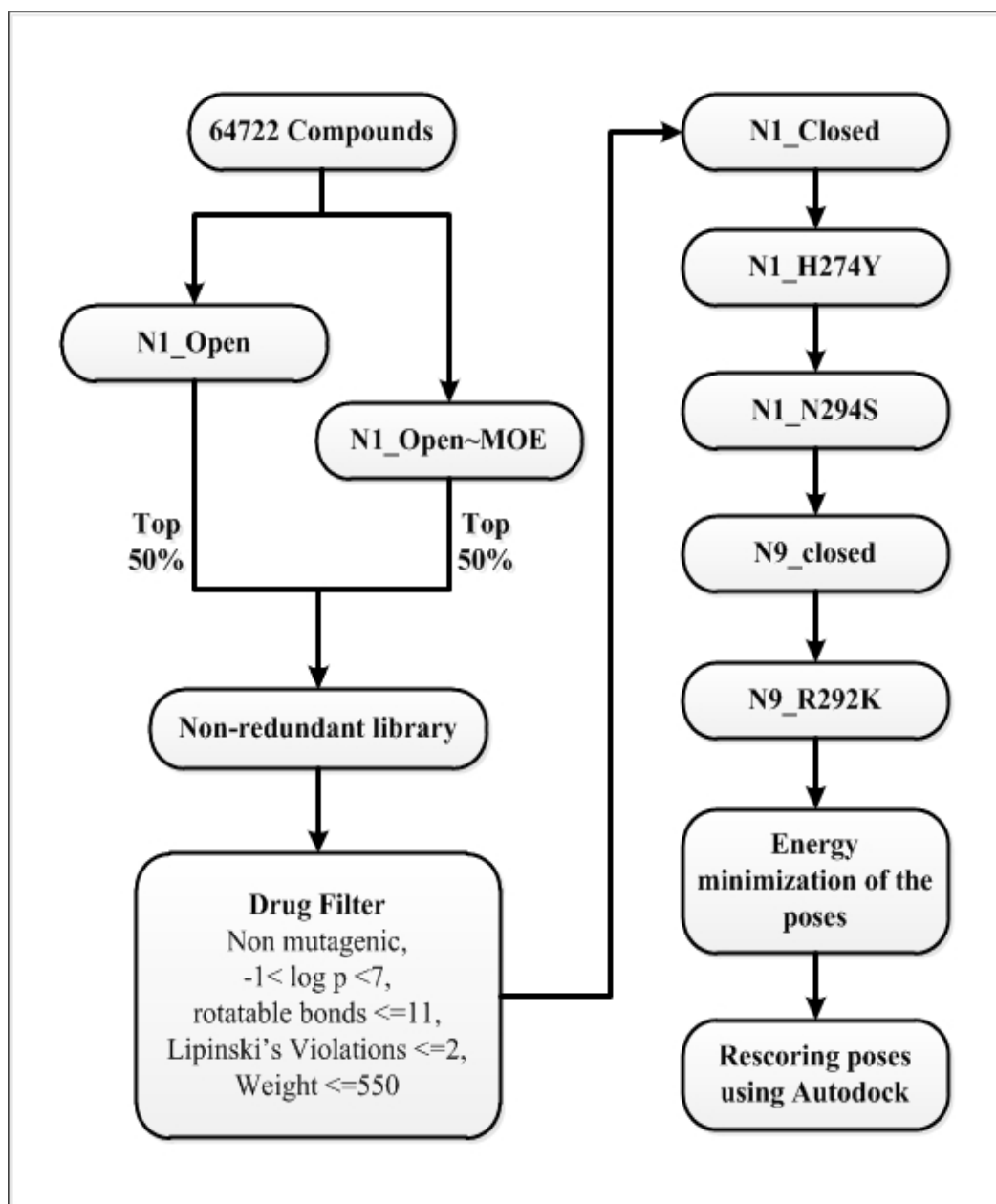


Figure4.1 Overview of docking process

-pounds, and AMBER ff12SB force field for the proteins. Hydrogen atoms for the neuraminidase were added using the LEaP module in AMBER12. Systems were solvated using TIP3P (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983)

water in a octahedron box extending 12 Å beyond any solute atom. The system was made neutral by adding appropriate number of counter ions.

Energy minimization was performed using pmemd.cuda in AMBER12. Water molecules and counter ions were minimized with 500 cycles of steepest descent followed by same number of conjugate gradient cycles. This was followed by energy minimization of the entire system during which only ligand, protein residues within 6Å of ligand and hydrogen atoms were allowed to move, maintaining other atoms fixed. The energy minimization of atoms only in proximity of the ligand was performed as fully minimized poses could lead to different local minima thereby creating noise in analysis (Guimarães & Cardozo, 2008). The minimized poses were rescored using Autodock (Morris et al., 2009).

Chapter 5

Neuraminidase Results and Discussion

This chapter describes the results for neuraminidase including the performance of base models, overall performance of consensus model and the compounds predicted outside AD. In addition, results of docking and comparison of the poses of approved neuraminidase inhibitors as well as probable inhibitors against wild and mutant type are discussed.

5.1 QSAR

In this work, the first phase of drug design was to develop a QSAR model with a defined AD and low FPR for screening large chemical libraries for novel and potentially potent NA1 inhibitors. We have developed such a model using a multiple thresholds AD method, potent NA1 inhibitors, use of marketed drugs as additional non-inhibitors and consensus modeling.

Only compounds having inhibitory activity against NA1 were collected while those having activity against other neuraminidase were not added into the dataset to reduce the FPR. It is common to use $IC_{50} \leq 10 \mu\text{M}$ to identify potent inhibitors (Birchall, Gillet, Harper, & Pickett, 2008; Lange et al., 2010; Naik, Santoshi, & Joshi, 2012). Although $IC_{50} \leq 1 \mu\text{M}$ has been suggested to identify

very potent inhibitors, it would have reduced the number of inhibitors in this study, which may reduce the diversity of inhibitors and cause the model to miss the identification of substantial numbers of potential inhibitors. Hence, in this study, we have decided to use the common threshold of $IC_{50} \leq 10 \mu\text{M}$. In the future, when there are more very potent inhibitors available, another QSAR model could be developed for the prediction of very potent neuraminidase inhibitors. A similar consideration is the use of two cutoff values to ensure better separation of inhibitors and non-inhibitors. However, this would have removed a large portion of the weak inhibitors from the negative set, which may reduce the AD of the model. The use of single cutoff value ensured that the molecules with weak activity were categorized as non-inhibitors. Hence, a single cutoff value of $IC_{50} \leq 10 \mu\text{M}$ was chosen to divide the dataset into potent and weak inhibitors.

There are more published potent NA1 inhibitors than non-inhibitors. Models trained using such datasets often had high FPR. Hence, in this study, we added marketed drugs to increase the size and diversity of non-inhibitors so as to reduce the FPR of the models. The results suggest that such approach is practical and useful. A considerably low FPR and FDR of the consensus model increase the efficiency to screen large chemical libraries.

Table 5.1 Performance of the base models selected to form consensus model

Base model	Value of k	No. of descriptors	Training						Cross Validation			
			TP	FN	TN	FP	No. of inhibitors outside AD of model	No. of non-inhibitors outside AD of model	AUC	SE (%)	SP (%)	MCC
1	3	227	110	0	1023	0	28	29	0.992	100	99.6	0.980
2	3	224	110	0	1023	0	28	29	0.991	100	99.6	0.980
3	3	221	112	0	1023	0	26	29	0.991	100	99.6	0.980
4	3	228	110	0	1023	0	28	29	0.992	100	99.6	0.980
5	3	225	110	0	1023	0	28	29	0.991	100	99.6	0.980
6	5	221	124	3	1039	5	11	8	0.993	97.4	99.5	0.962
7	5	222	124	3	1039	5	11	8	0.993	97.4	99.4	0.959
8	5	224	124	3	1039	5	11	8	0.993	97.4	99.4	0.959
9	5	223	124	3	1039	5	11	8	0.993	97.4	99.4	0.959
10	5	224	124	3	1039	5	11	8	0.993	97.4	99.4	0.959

5.1.1 Base Models

From a total of 38655 base models, top models were selected on the basis of their performance (**Table 5.1**) to form consensus model. Analysis of the base models used to construct the consensus model showed that all the base kNN models had k as 3 or 5.

Base kNN models with $k > 5$ had larger FP and FN. This could be due to the fact that though large k yields smoother decision regions (Y. Song, Huang, Zhou, Zha, & Giles, 2007), k that were too large are detrimental as it destroys the locality of the estimation since farther unrelated compounds are also used for the prediction of the unknown compounds (V. C. P. Chen, 2010). Base SVM models usually had larger FN, which may have led to the rejection of many novel inhibitors. Hence these models were not used to form the consensus model. From the initial 672 descriptors, a maximum of 228 descriptors were found to be useful for developing the base models.

5.1.2 Performance of consensus model

The performance of the consensus model on the training set and validation set were similar, with an overall accuracy of 99.3% and 98.0% respectively (**Table5.2**). This suggests that the consensus model was unlikely to be overfitted.

It can be seen that the consensus model has a low FPR, the proportion of negative predicted as positive out of total negative compounds available, of 0.8%

and low FDR, which is a measure of how many of total positives predicted are not TP, of 6.3%.

	TP	FN	TN	FP	No. of inhibitor outside AD of model	No. of non-inhibitor outside AD of model	AUC	SE (%)	SP (%)	Q (%)	MCC
Training set	127	3	1042	5	8	5	0.916	97.7	99.5	99.3	0.966
Validation set	30	4	260	2	1	1	0.851	88.2	99.2	98.0	0.898

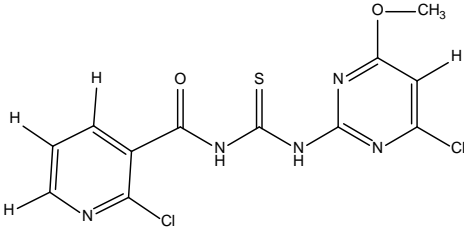
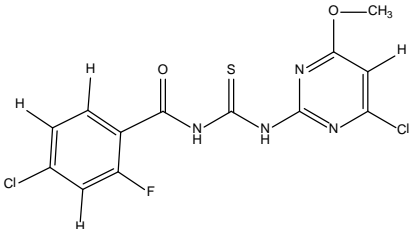
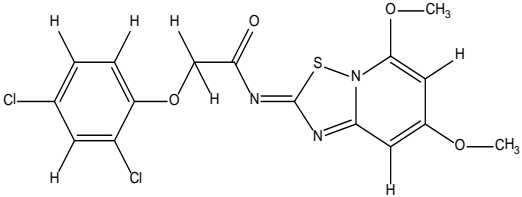
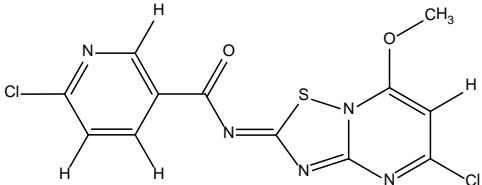
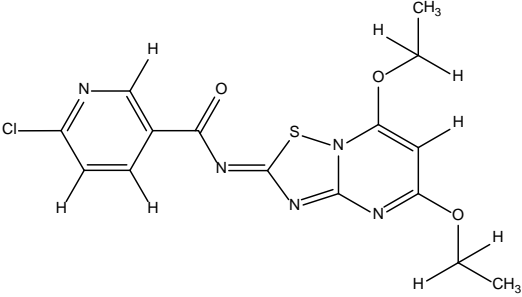
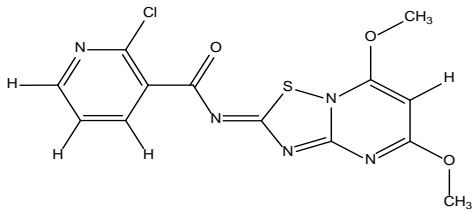
The numbers of descriptors used in the base models varied between 200 and 250. MCC value was calculated as it is less influenced by imbalanced test sets due to consideration of accuracy and error rate of both classes (Bekkar, Djemaa, & Alitouche, 2013). In comparison to overall accuracy MCC dropped for validation set which was more significant for validation set i.e. 0.898 in comparison to 0.966.

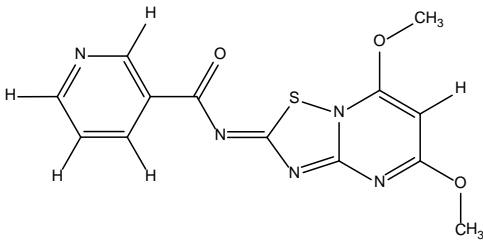
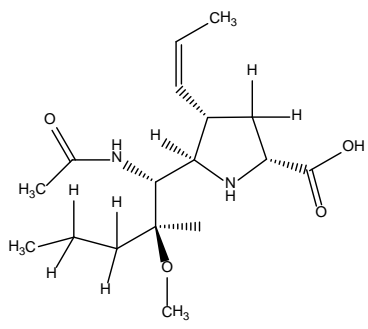
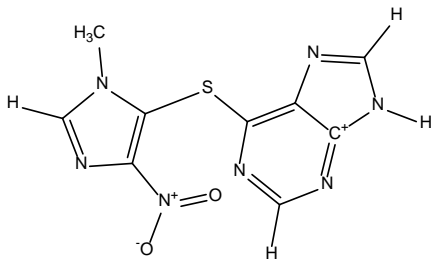
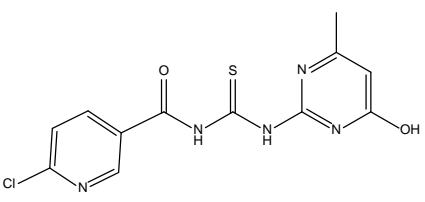
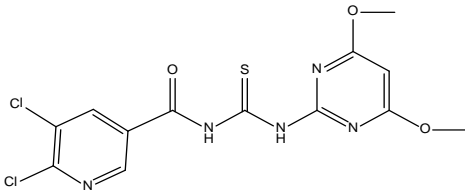
5.1.3 Compounds outside AD

AD for a model can be defined in many ways. Some models define AD using chemical classes that are found in the training set. In other cases, AD is defined using the range of descriptors values or determining the probability density. The disadvantage of these AD methods is that the AD is dependent on the training set but independent of the modeling method. An improved method is to define the AD on the basis of prediction confidence (Tong et al., 2004). This allows the AD to be dependent on both the training set and modeling method.

Hence, AD of the base models was defined using prediction confidence in this study.

Table 5.3 Compounds outside the AD of the consensus model		
Compound	Name as in reference	Chemical class
a) Training set		
<p>The structure shows a thiourea group (-NH-C(=S)-NH-) connecting two pyrimidine rings. The left pyrimidine ring has a chlorine atom at the 5-position. The right pyrimidine ring has a methoxy group (-OCH₃) at the 2-position.</p>	37 ⁺	Acyl(thio)ureas bearing a substituted pyrimidine ring
<p>The structure is similar to 37+, but the chlorine atom is at the 4-position of the left pyrimidine ring, and the methoxy group is at the 4-position of the right pyrimidine ring.</p>	38 ⁺	Acyl(thio)ureas bearing a substituted pyrimidine ring
<p>The structure shows a thiourea group (-NH-C(=S)-NH-) connecting a substituted pyrimidine ring to a substituted benzene ring. The pyrimidine ring has a chlorine atom at the 5-position. The benzene ring has a methyl group (-CH₃) at the 2-position and a chlorine atom at the 4-position.</p>	32 ⁺	Acyl(thio)ureas bearing a substituted pyrimidine ring
<p>The structure shows a thiourea group (-NH-C(=S)-NH-) connecting a substituted pyrimidine ring to a substituted furan ring. The pyrimidine ring has a methoxy group (-OCH₃) at the 2-position. The furan ring has a nitro group (-NO₂) at the 5-position.</p>	35 ⁺	Acyl(thio)ureas bearing a substituted pyrimidine ring

	44 ⁺	Acyl(thio)ureas bearing a substituted pyrimidine ring
	53 ⁺	Acyl(thio)ureas bearing a substituted pyrimidine ring
	55 ⁺	<i>2H</i> -1,2,4-thiadiazolo [2,3-a] pyrimidine
	59 ⁺	<i>2H</i> -1,2,4-thiadiazolo [2,3-a] pyrimidine
	60 ⁺	<i>2H</i> -1,2,4-thiadiazolo [2,3-a] pyrimidine
	61 ⁺	<i>2H</i> -1,2,4-thiadiazolo [2,3-a] pyrimidine

	62 ⁺	2 <i>H</i> -1,2,4-thiadiazolo [2,3- <i>a</i>] pyrimidine
	A-315675*	5-[(1 <i>R</i> ,2 <i>S</i>)-1 (acetylamino)-2-methoxy-2-methylpentyl]-4-[(1 <i>Z</i>)-1-propenyl]- (4 <i>S</i> ,5 <i>R</i>)- <i>D</i> -proline
	Azathioprine**	
b) Validation set		
	40 ⁺	Acyl(thio)ureas bearing a substituted pyrimidine ring
	46 ⁺	Acyl(thio)ureas bearing a substituted pyrimidine ring

⁺ From Reference (Sun et al., 2006), *From reference (Abed et al., 2008), ** Drug from FDA orange book

Table 5.4 Functional group present in the compounds outside of AD	
Group	Number of appearance in different compounds
OMe	14
6-Cl-3-Py	5
OEt	2
Cl	4
Me	3
2-Cl-3-Py	2
2-Me-1-(4-Cl-Ph)-Pr	1
(2,4-Cl ₂ -Ph)-OCH ₂	1
5,6-Cl ₂ -3-Py	1
5-(4-NO ₂ -Ph)-2-Furyl	1
2-F-4-Cl-Ph	1
(2,4-Cl ₂ -Ph)-OCH ₂	1
3-Py	1
OH	1

Me : Methyl, Et : Ethyl, Ph: Phenyl, Py: Pyrimidine

A total of 13 compounds in the training set and 2 compounds in the validation set were outside the AD of the consensus model **Table 5.3**. Analysis of the compounds outside AD of the consensus model reveals that the majority of these compounds are of the acyl(thio) urea class with substituted pyrimidine rings. Other than substituted pyrimidine rings, other groups were also present in these compounds (**Table 5.4**).

Screening of the ZINC library using the consensus model identified 64772 compounds as potential neuraminidase inhibitors. A total of 173674 compounds were outside the AD of the consensus model. Most of the identified potential NA1 inhibitors from the ZINC library have one or more aromatic rings. The scaffold connecting the rings had C=S, N-H, and C=O, which may form hydrogen bonds in the active site of NA1.

5.2 Docking

The extensive docking process reduced the number of compounds to 1148. The structures of the top 10 PNI selected for energy minimization can be observed in (**Table5.5**). The Tanimoto coefficient was calculated between the top 10 PNI and established drugs i.e. oseltamivir, zanamivir and laninamivir. The patent records were checked for any possible patents of the inhibitors suggested by our work. In addition the inhibitors were inspected for role in any other disease. Lastly, commercial availability of the compounds was looked into.

As can be observed from the (**Table5.6**) none of the PNI is closely related to established drugs. However, PNI were found to be more similar to zanamivir

and laninamivir which could be attributed to the presence of guanidino group. ZN37 was found to be most similar to all three established inhibitors which can be observed by highest tanimoto coefficient of 0.549, 0.609 and 0.307 for zanamivir, laninamivir and oseltamivir respectively. ZN17 and ZN88 were found to be most dissimilar compounds in comparison due to lack of central scaffold present in other inhibitors. ZN33 and ZN99 are other compounds similar to established inhibitors. Despite having guanidino group and central ring similar to zanamivir and laninamivir, ZN88 has low tanimoto coefficient of 0.216, 0.194 and 0.189 against zanamivir, laninamivir and oseltamivir respectively. The tanimoto coefficient of ZN46 was also found on lower side.

As be seen from (**Table5.7**) that none of the probable inhibitors have been reported in literature for any other diseases. However, two compounds ZN21 and ZN37 was found to be covered in patent by IBM and Thomson pharma respectively. Despite being screened from ZINC library, none of the PNI is commercially available.

Table 5.5 The final 10 PNI and their ZINC codes

	Zinc Code	Structure		Zinc Code	Structure
1	ZINC13443833 (ZN33)		6	ZINC13443835 (ZN35)	
2	ZINC13443799 (ZN99)		7	ZINC13778721 (ZN21)	
3	ZINC13778743 (ZN43)		8	ZINC36355417 (ZN17)	
4	ZINC29556337 (ZN37)		9	ZINC33818946 (ZN46)	
5	ZINC34341188 (ZN88)		10	ZINC38494578 (ZN78)	

Table5.6 Tanimoto coefficient of the PNI against established inhibitors				
	Zinc Code	Zanamivir	Laninamivir	Oseltamivir
1	ZN37	0.549	0.609	0.307
2	ZN33	0.398	0.498	0.291
3	ZN99	0.537	0.495	0.265
4	ZN35	0.37	0.46	0.28
5	ZN21	0.353	0.343	0.217
6	ZN43	0.332	0.333	0.301
7	ZN46	0.245	0.294	0.259
8	ZN88	0.216	0.194	0.189
9	ZN17	0.157	0.142	0.098
10	ZN78	0.115	0.14	0.132

Table5.7 Information related to PNI						
	Zinc Code	Literature Reference	Organism	Patent	Other Diseases	Vendor
1	ZN33	Bioorg Med Chem Lett. 2002 Aug 5;12(15):1925-8	Influenza A	None	None	None
2	ZN35	Bioorg Med Chem Lett. 2002 Aug 5;12(15):1925-8.	Influenza A	None	None	None
3	ZN99	Bioorg Med Chem Lett. 2002 Aug 5;12(15):1921-4.	Influenza A	None	None	None
4	ZN21	Bioorg. Med. Chem. Lett., (1996) 6:15:1805	Influenza A/B	IBM Patent Data EP0833825; US5990156; WO1996036628	None	None
5	ZN43	J. Med. Chem., (1998) 41:6:787	Influenza A/B	None	None	None
6	ZN17	None	None	None	None	None
7	ZN37	Bioorg Med Chem Lett. 2002 Aug 5;12(15):1925-8.	Influenza A	Thomson Pharma	None	None
8	ZN46	None	None	None	None	None
9	ZN88	None	None	None	None	None
10	ZN78	None	None	None	None	None

5.2.1 Energy Minimization and Rescoring

Scores generated as a result of energy minimization can be observed in **Table 5.8**. The energy minimization and pose rescoring resulted in decrease of the energy for ZN17 and ZN78. The binding free energy of zanamivir and laninamivir was found to be close to each other. For all three i.e. oseltamivir, zanamivir and laninamivir, lowest binding free energy was found in N9_R292K. Besides N9_R292K the binding free energy of zanamivir and laninamivir showed no effect of mutation. On the other hand, the binding free energy of oseltamivir dropped slightly in N1_N294S and even more in N1_H274Y. Binding free energy of the top five PNI were not affected by N1_H274Y and N1_N294S mutation but some slight drop was observed in N9_R292K.

5.2.1.1 Standard Deviation of the docking scores

The standard deviation (SD) of the docking scores was calculated (**Table 5.8**) in order to indicate how susceptible the different inhibitors are to different mutation environments. The lowest SD was observed for ZN78 but it also had lowest average docking score. The SD of ZN43, ZN88, ZN35 and ZN46 was found close to each other. Among these four ZN88 had highest average binding free energy followed by ZN46, ZN35 and ZN43. Though ZN33 and ZN21 ranked 2nd and 3rd according to the average binding free energy but they showed higher SD in comparison to the above mentioned inhibitors. The highest SD was observed for laninamivir, zanamivir and oseltamivir indicating maximum

susceptibility towards mutations. Hence, ZN43, ZN88, ZN35 and ZN46 are least influenced by mutations and also has good binding free energy.

Table 5.8 Binding free energy (kcal/mol) of 10 PNI along with oseltamivir, zanamivir and laninamivir.							
Compound	N1_closed	N1_N294S	N1_H274Y	N9_closed	N9_R292K	ABFE	SD
ZN78	-5.51	-5.17	-5.02	-5.15	-6.01	-5.372	0.4
ZN43	-9.24	-9.28	-9.7	-8.71	-8.19	-9.024	0.584
ZN88	-11.17	-11.6	-11.63	-10.64	-10.25	-11.058	0.604
ZN35	-9.79	-9.61	-10.16	-9.31	-8.5	-9.474	0.625
ZN46	-9.21	-9.43	-10.7	-9.12	-9.39	-9.57	0.644
ZN37	-7.49	-8.15	-9.58	-8.04	-8.21	-8.294	0.733
ZN99	-7.92	-10.13	-9.31	-8.51	-8.93	-8.96	0.833
ZN33	-10.94	-9.71	-10.61	-10.08	-8.75	-10.018	0.853
ZN21	-9.02	-10.18	-11.16	-9.21	-8.95	-9.704	0.952
ZN17	-7.75	-7.7	-5.28	-6.46	-5.99	-6.636	1.079
Laninamivir	-7.65	-8.32	-8.85	-7.57	-5.59	-7.596	1.237
Zanamivir	-8.1	-7.83	-8.73	-7.97	-5.28	-7.582	1.332
Oseltamivir	-8.58	-8.38	-7.72	-8.87	-5.16	-7.742	1.504

5.2.1.2 Correlation between IC50 and average binding free energy (ABFE)

Table 5.9 Average binding free energy (kcal/mol) and IC50 (nM) oseltamivir, zanamivir and laninamivir						
	Zanamivir		Laninamivir		Oseltamivir	
	IC50*	ABFE	IC50*	ABFE	IC50*	ABFE
N1_closed	0.15	-8.1	0.28	-7.65	0.31	-8.58
N1_N294S	0.48	-7.83	1.4	-8.32	28	-8.38
N1_H274Y	1.5	-8.73	7.5	-8.85	1100	-7.72
N9_closed	0.65	-7.97	0.74	-7.57	0.28	-8.87
N9_R292K	35.68	-5.28	16.33	-5.59	2870	-5.16
Correlation	0.959		0.659		0.986	

The IC50 was of zanamivir, laninamivir and oseltamivir for N1 and N9 strains was collected from the work of Marjuki H et.al (Marjuki et al., 2014) and Yamashita (Yamashita, 2010) respectively. Correlation was calculated between IC50 and average binding free energy obtained for zanamivir, laninamivir and oseltamivir. High correlation was obtained for oseltamivir and zanamivir i.e. 0.986, 0.959 which dropped slightly for laninamivir i.e. 0.659 (**Table 5.9**).

Among the 10 PNI, IC₅₀ for six inhibitors are reported for neuraminidase A i.e. N1_closed in our study. The correlation of 0.486 was observed between the documented IC₅₀ and calculated average binding free energy (**Table 5.10**).

Table 5.10 Correlation between IC₅₀ and calculated binding free energy		
Compound	IC₅₀ (μg/ml)	BFE
ZINC13443833	0.016 (Honda, Masuda, Yoshida, Arai, Kaneko, et al., 2002)	-10.94
ZINC13443835	0.0141 (Honda, Masuda, Yoshida, Arai, Kaneko, et al., 2002)	-9.79
ZINC13443799	0.0181 (Honda, Masuda, Yoshida, Arai, Kaneko, et al., 2002)	-7.92
ZINC13778721	0.28 (Sollis, Smith, Howes, Cherry, & Bethell, 1996)	-9.02
ZINC13778743	0.0001 (Smith et al., 1998)	-9.24
ZINC29556337	0.264 (Honda, Masuda, Yoshida, Arai, Kaneko, et al., 2002)	-7.49

5.2.2 Conformations of Glutamic276 in non-mutant strains

Comparison of the poses of N1_closed with oseltamivir, zanamivir and laninamivir (together all three will be referred as OZL) reveals that Glu276 can occupy two conformations suitable for both polar and non-polar interactions (**Figure5.2**). In the first conformation, i.e. with zanamivir and laninamivir as ligands, the carboxylic group faces the binding pocket thereby forming hydrogen bond with 8- and 9- hydroxyl of the glycerol group of zanamivir and laninamivir. On the other hand while oseltamivir is the ligand, carboxylic group of Glu276 faces away from binding pocket thus making the pocket more hydrophobic in nature. This facilitates oseltamivir pentyloxy side to form non-polar interactions with Glu276.

5.2.3 Conformation of Glutamic276 leading to resistance

All three mutations, i.e. N294S, H274Y and R292K, results in movement of carboxylic group inside hydrophobic cavity with varying strength. This results in the disruption of hydrophobic cavity leading to resistance against oseltamivir.

5.2.3.1 N294S and H274Y mutations

Mutation of Asn294 to Ser294 (N294S) results in slight change in the position of Glu276 which has no effect on hydrogen bond formation with polar side chain of zanamivir or laninamivir (**Figure5.3**). The non-polar side chain of oseltamivir was pushed outside by about 1Å. The effect of Tyr347 being slightly flipped did not result in loss of hydrogen bond. However, the bond length of

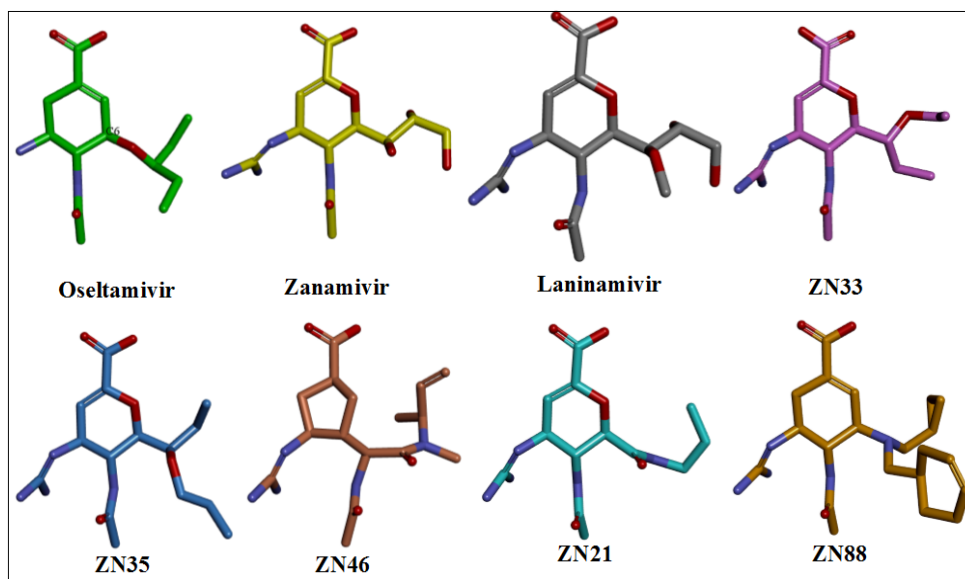


Figure5.1 Structures of oseltamivir, zanamivir, laninamivir and top 5 PNI according to ABFE

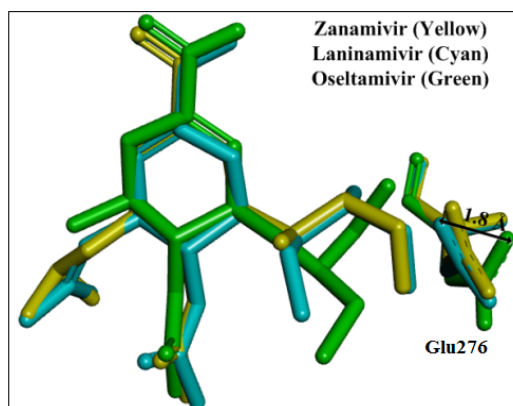


Figure5.2 Conformation of Glu276 with oseltamivir, zanamivir and laninamivir as inhibitors

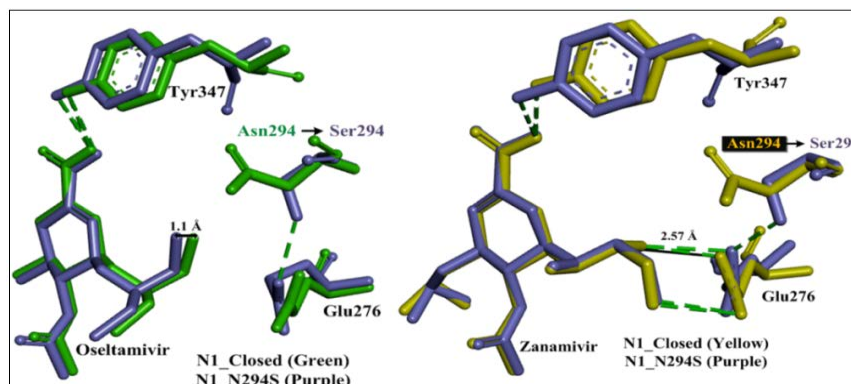


Figure5.3 Comparison of oseltamivir and zanamivir poses in N1_closed and N1_N294S.

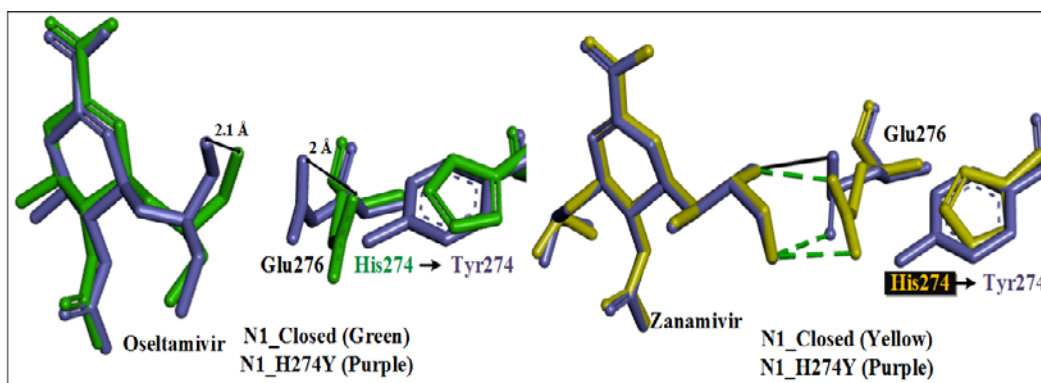


Figure 5.4 Comparison of pose of oseltamivir, zanamivir in N1_closed and N1_H274Y.

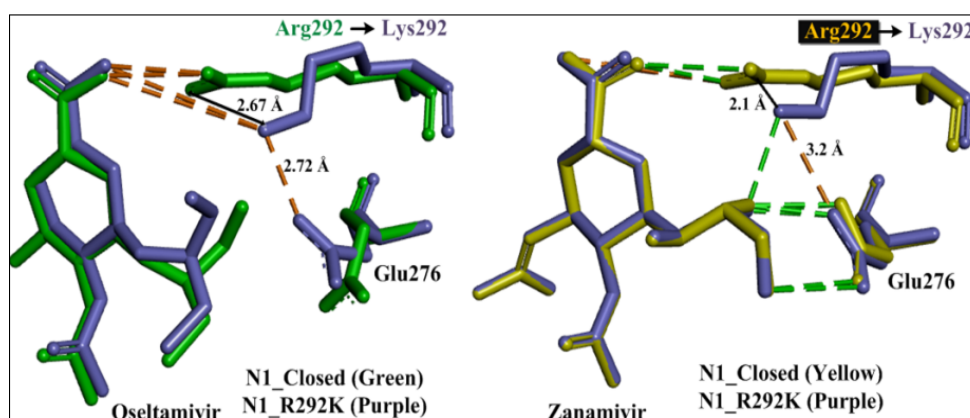


Figure 5.5 Comparison of poses of oseltamivir, zanamivir in N9_closed and N9_R29K.

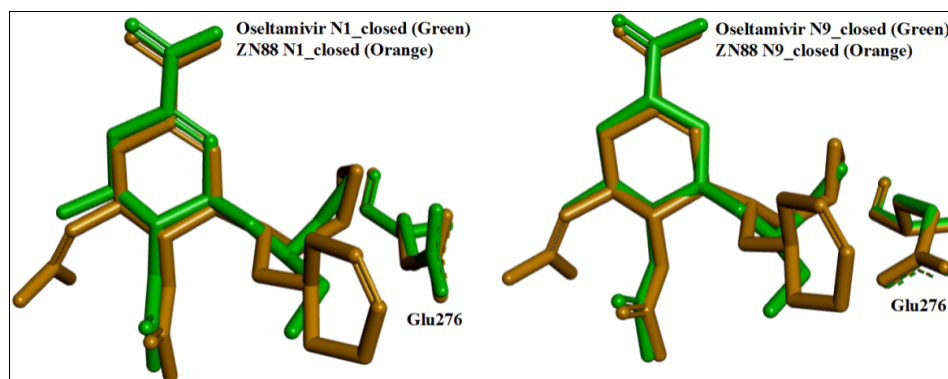


Figure 5.6 Comparison of ZN88 and oseltamivir pose in N1_closed and N9_closed

hydrogen bond formed was higher by 0.2Å for mutated system. These results are in concordance with study performed by Collins et.al (Collins et al., 2009; Collins et al., 2008).

As a result of H274Y, i.e. His274 to Tyr274, mutation, Glu276 moved inside the cavity by about 2Å thereby pushing the oseltamivir outside by about 2.1Å (**Figure5.4**). However, the mutation resulted in orientation of Glu276 close to one required by zanamivir and laninamivir for formation of hydrogen bond. Hence, binding mode of zanamivir and laninamivir was not affected and they formed hydrogen bond with Glu276 in both closed and mutant strains.

Effect of large inward movement of Glu276 in N1_H274Y as compared to N1_N294S was also evident by high binding free energy difference of oseltamivir between N1_H274Y and N1_closed in comparison to N1_N294S and N1_closed.

5.2.3.2 R292K mutation

The comparison of poses of oseltamivir in N9_closed and N9_R292K shows that mutation of Arg292 to Lys292 (R292K) causes the Glu276 to adopt a conformation enabling it to form salt bridge with Lys292 (**Figure5.5**). This inward movement disrupts the hydrophobic cavity required by oseltamivir's pentyloxy group thereby shifting oseltamivir by 2.96Å which is in good agreement with a recent study (Yan Wu et al., 2013). The change in orientation of Glu276 has limited effect on zanamivir and laninamivir due to the formation of hydrogen bond between polar group of the side chain and Glu276. However, presence of Lys292 instead of Arg292 resulted in loss of hydrogen bond with the

carboxylic group of the ligand. This causes resistance against OZL which is indicated by lowest binding free energy in N9_R292K.

5.2.4 Comparison of the poses of potential inhibitors with wild strains

Comparison of the poses of ZN88 with oseltamivir in N1_closed and N9_closed (**Figure5.6**) indicates that side chain of ZN88 and oseltamivir are close to each other. In both N1_closed and N9_closed (**Figure5.7**) amino group of side chain of ZN88 forms salt-bridge with Glu276 and Glu277. Salt-bridge between amino group of side chain and Asp151 was observed in N1_closed. In addition, non-polar part of the side chain is involved in hydrophobic interactions.

On the other hand, side chain of ZN33 shows more flexibility as observed from the poses with N1_closed and N9_closed (**Figure5.7**). Similar to ZN33, the flexible side chain of ZN35 was found close to oseltamivir exploring the hydrophobic cavity in N1_closed and N9_closed (**Figure5.9**). In both ZN33 and ZN35, the ether group of the side chain was not able to form any hydrogen bond.

The longer pentyl side chain of ZN21 was found to be close to the terminal of oseltamivir side chain in N1_closed. The side chain was found to be in different conformation in N1_closed thus indicating the flexibility of the side chain (**Figure5.10**). Despite of the presence of ketone group and amine group no polar interaction was observed.

The side chain of ZN46 was found in close proximity to oseltamivir and formed non-polar interaction similar to other inhibitors (**Figure5.11**). In addition

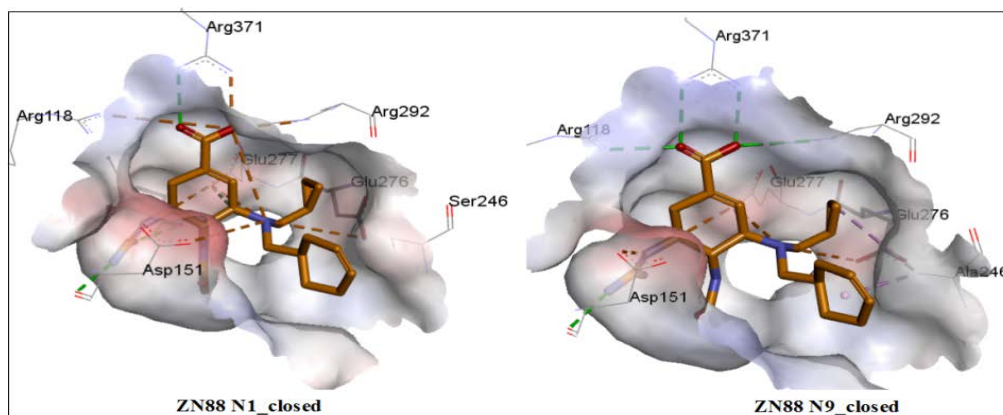


Figure5.7 Comparison of the Interactions of ZN88 in N1_closed and N9_closed.

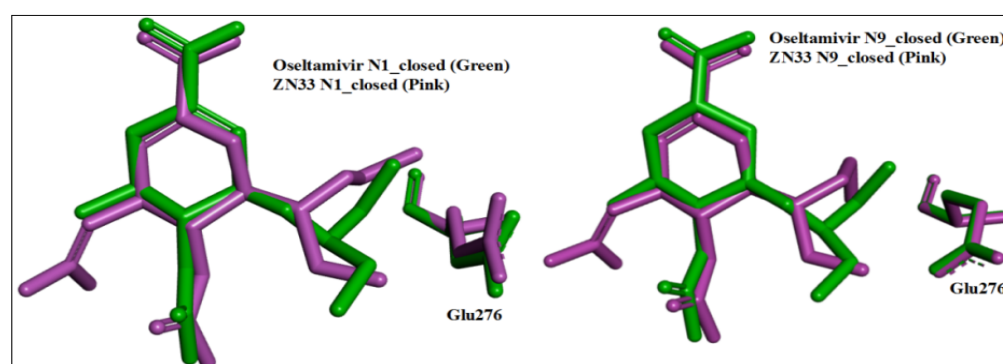


Figure5.8 Comparison of ZN33 and oseltamivir pose in N1_closed and N9_closed

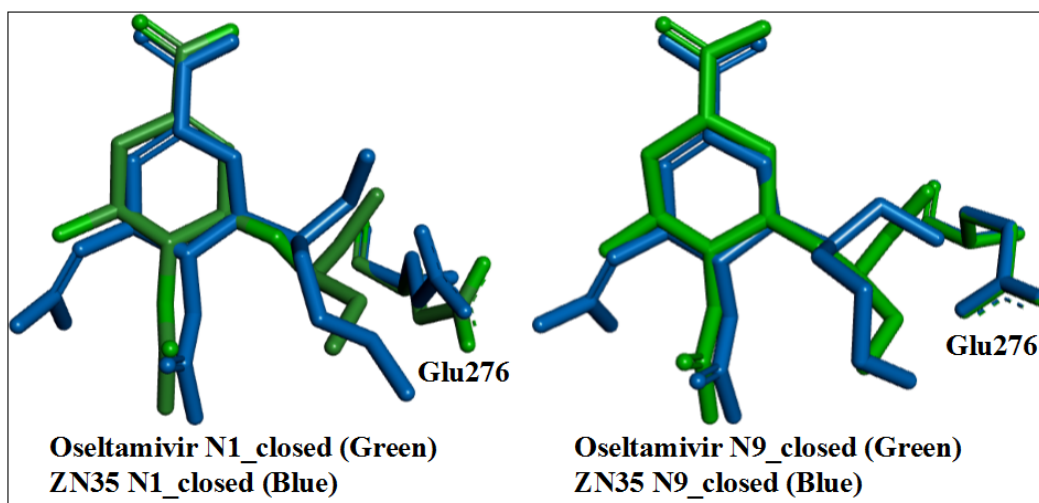


Figure5.9 Comparison of ZN35 and oseltamivir pose in N1_closed and N9_closed.

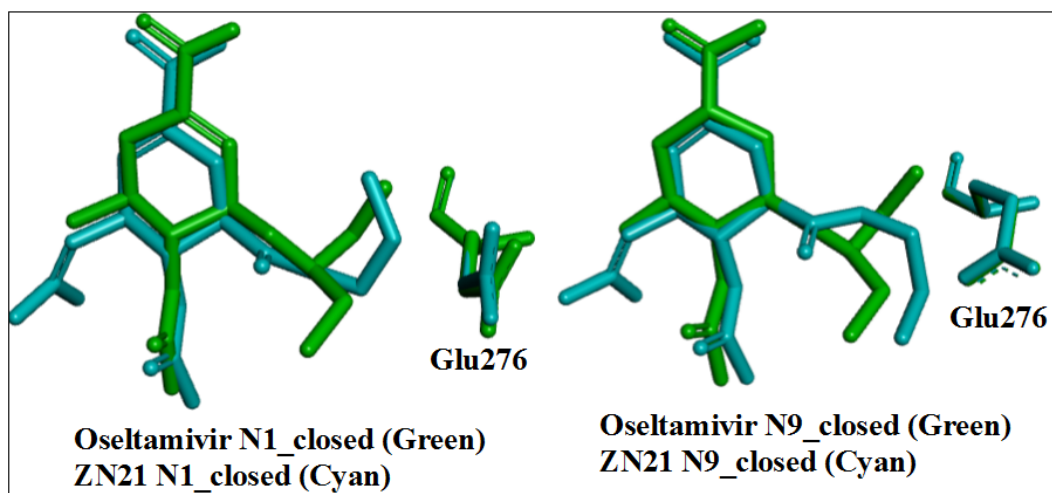


Figure5.10 Comparison of ZN21 and oseltamivir pose in N1_closed and N9_closed

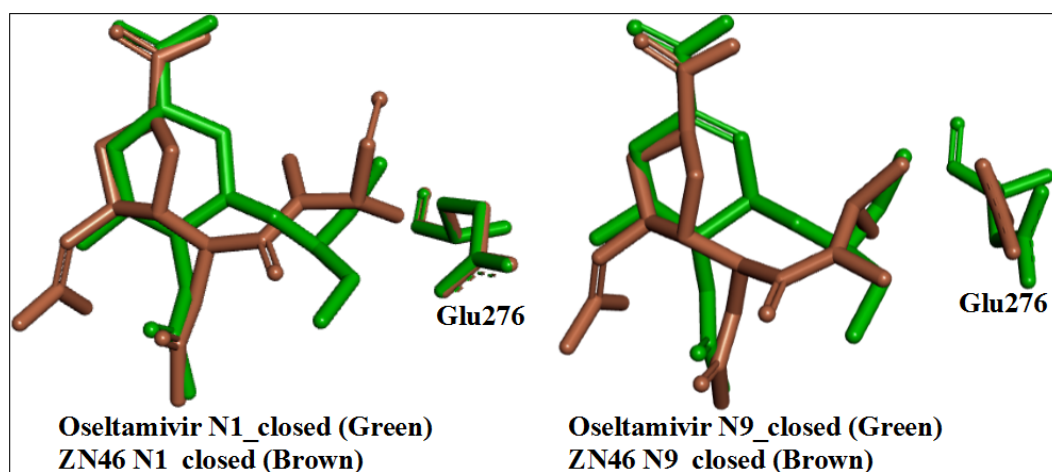


Figure5.11 Comparison of ZN46 and oseltamivir pose in N1_closed and N9_closed

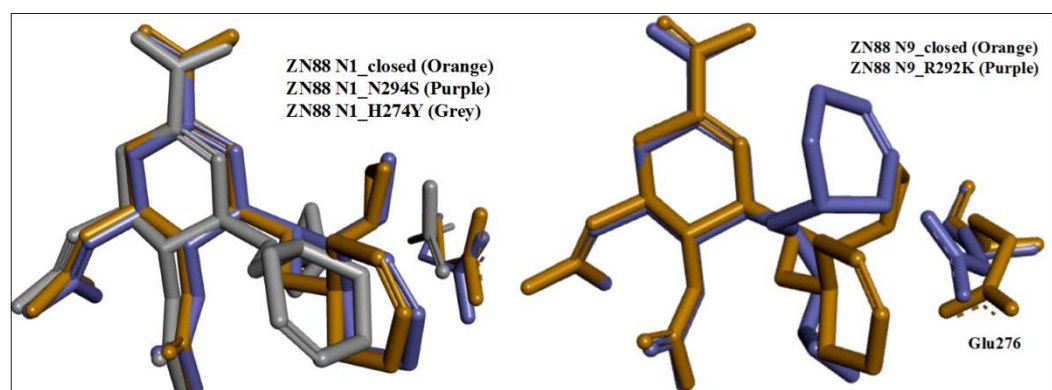


Figure5.12 Comparison of the poses of ZN88 in different strains

the ketone group was also able to form additional hydrogen bond with Arg152. However, the amino group did not participate in hydrogen bond with Glu276, Glu277 and Asp151 like ZN88.

Due to absence of side-chain amino group, ZN33, ZN35 and ZN21 lacked additional salt-bridge formed by ZN88. All inhibitors due to guanidino group formed hydrogen bond with Glu119, Asp151, Arg152, Trp178, Glu227 and between Arg152 and amide group similar to zanamivir and laninamivir.

Increase of the side chain by one carbon in ZN35 in comparison to ZN33 cause the shorter side chain to move inside the pocket. This indicates the side chain, only as long as propyl can occupy the binding cavity. The effect of longer chain was also observed in ZN21 where longer side chain hindered its entrance into the pocket. The unprotonated form of amino group in ZN46 resulted in absence of hydrogen bond with Glu276, Glu277 and Asp151.

5.2.5 Comparison of the poses of potential inhibitors with mutant strains

The comparison of poses of ZN88 across different mutant strains reveals that pose with N1_N294S is almost similar to one with N1_closed (**Figure5.12**). However, for mutant N1_H274Y, where Glu276 is present more towards the pocket, the side chain of ZN88 is pushed backwards. In N9_R292K, the presence of Lys292 resulted in smaller hydrophobic cavity (**Figure5.13a**). It can be observed that side chain of pose in N1_H274Y crashes in the surface created by N9_R292K. This causes the side chain to rotate and still form the non-polar

interactions. In addition, the amino group of ZN88 was able to form salt-bridge with Glu276 in N9_R292K (**Figure5.13b**) and in N1_H274Y (**Figure5.14**). The interaction between Asp151 and ZN88 was not observed in N1_N294S (**Figure5.14**).

Like ZN88, the pose of ZN33 in N1_closed was found similar to N1_N294S. However, for N1_H274Y the shorter ethyl side chain was found to be deep inside the cavity while propyl side chain was pushed outside (**Figure5.15**). For N9_R292K, the ethyl side chain was found inside the pocket while propyl was found facing outwards. The longer side chain was found to be shifted as a result of constricted hydrophobic cavity. The interactions with guanidine and amide group, similar to non-mutant strains, were found in all mutant systems.

The effect of increase in length of the side chain can be observed from the orientations of ZN35. Except for N1_N294S, the side shorter chain was found inside the pocket for N1_closed and N1_H274Y. For N9_R292K the side chain was found to be twisted due to constrain in binding cavity (**Figure5.16**).

Almost similar poses were observed for ZN21 in N1_closed, N1_N294S and N1_H274Y (**Figure5.17**). The side chain was found to be rotated in opposite direction in N9_R292K as compared to N9_closed. However, the effect of the change in hydrophobic cavity was not observed in both N9_closed and N9_R292K as the side chain was not found facing the base of the cavity.

ZN46 was able to deal with the mutation by the change in orientation of side chain (**Figure5.18**). The vinyl group was able to form non-polar interaction with

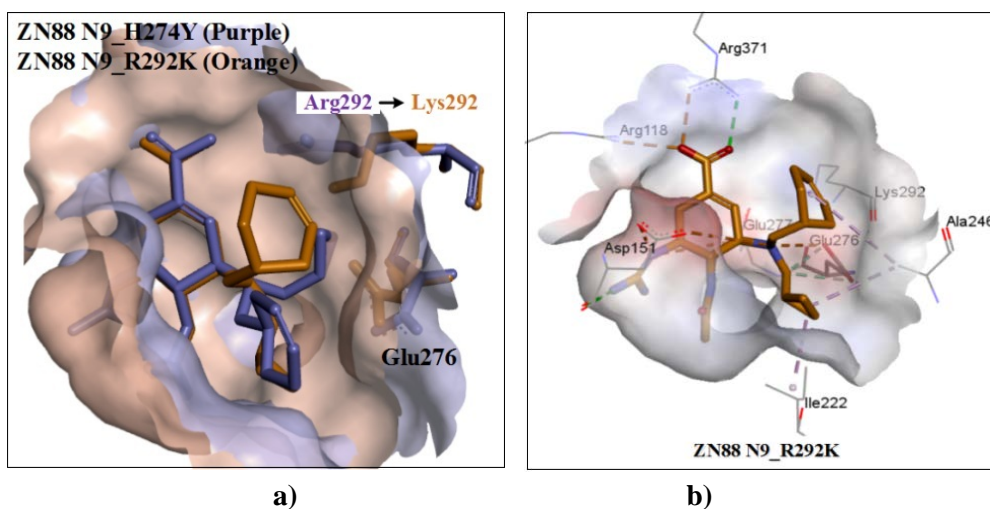


Figure5.13 a) Comparison of the poses of ZN88 in N1_H274Y and N9_R292K b) Interaction of ZN88 in R292K

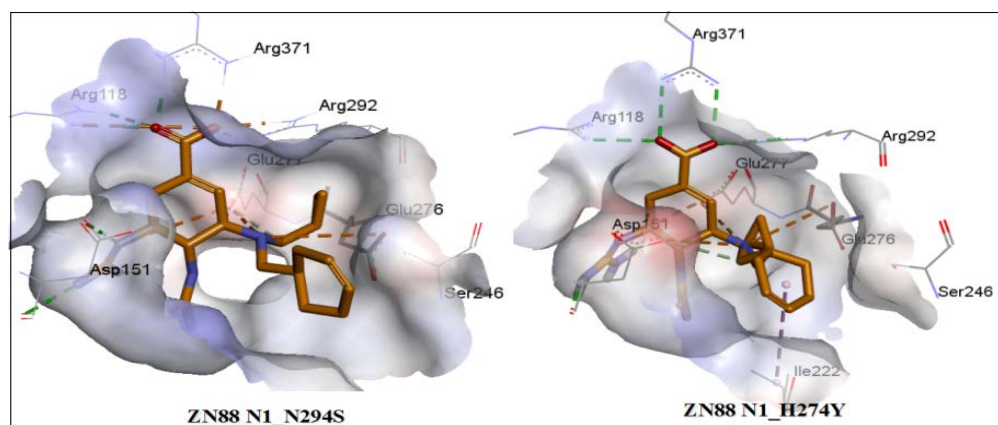


Figure5.14 Comparison of the poses of ZN88 in N1_N294S and N1_H274Y

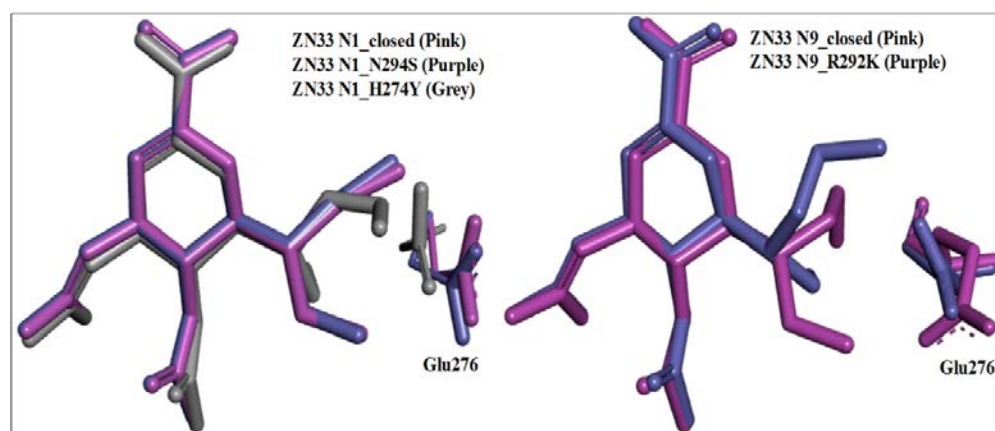


Figure5.15 Comparison of the poses of ZN33 in different strains

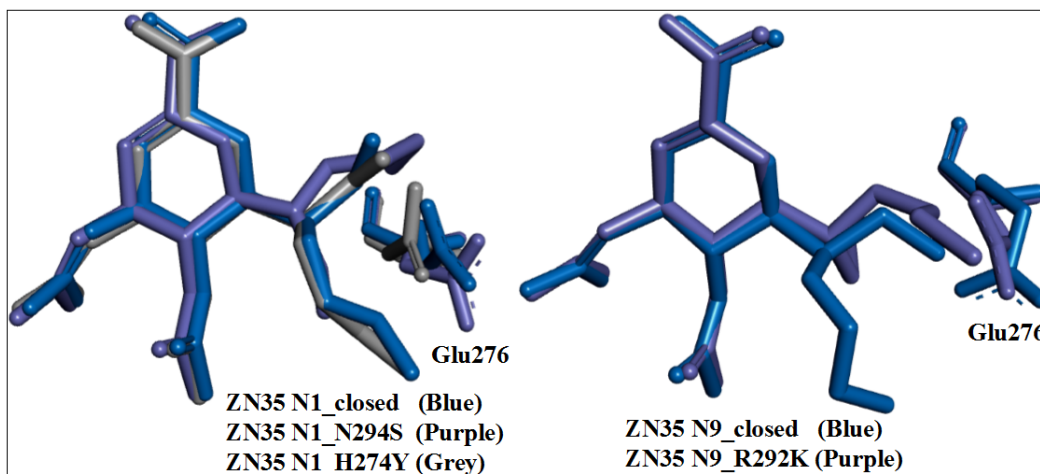


Figure5.16 Comparison of the poses of ZN35 in different strains

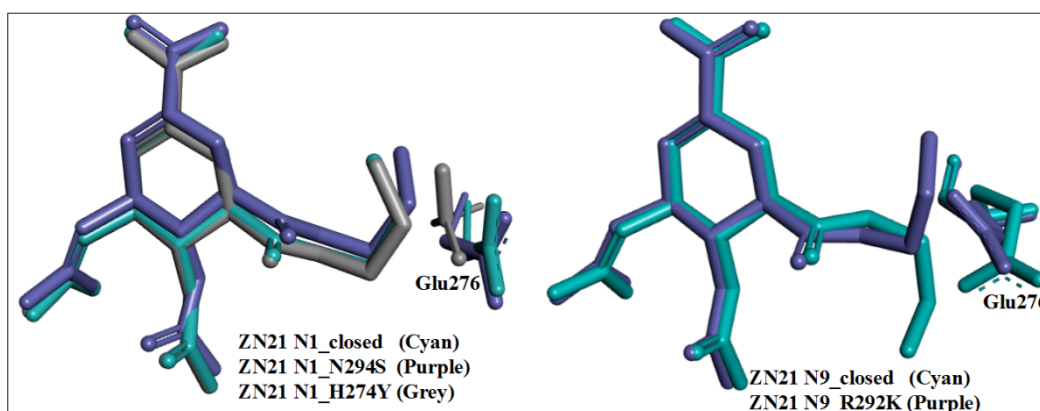


Figure5.17 Comparison of the poses of ZN21 in different strains

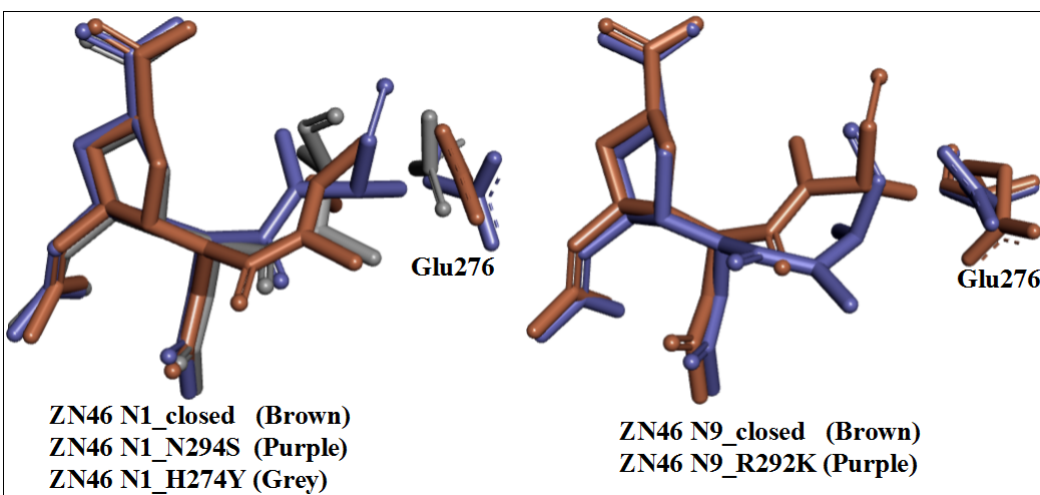


Figure5.18 Comparison of the poses of ZN46 in different strain

Tyr347 in N1_N294S and N1_H274Y; with Ala246 in N9_closed; and with Ile222 in N9_R292K. In addition, the ketone group was able to hydrogen bond with Arg152 in N1_closed and N9_R292K.

For both ZN88 and ZN33 in N1_N294S and N1_closed, side chain was found close to each other due to less inwards movement of Glu276 inside hydrophobic cavity. On the other hand greater inward movement of Glu276 in N1_H274Y resulted in pushing side chains deeper in the pocket. ZN35, ZN21 and ZN46 indicate that the length of the side chain and the flexibility is a key determinant in formation of non-polar interactions. The presence of Lys292, in addition to inward movement of Glu276, caused the shift in side chain of ZN88 and ZN33. Effect of constriction in hydrophobic cavity was also observed in ZN35 and ZN46. Nevertheless, all inhibitors were able to form non-polar interaction in all the systems.

The central ring and guanidino group were unaffected by any mutation. Hence, interaction between Glu119, Asp151, Arg152, Trp178, Glu227, Arg152 and guanidino group was still maintained similar to N1_closed and N9_closed. All inhibitors formed hydrogen bond with arginine triad i.e. Arg118, Arg292 and Arg371 in N1_closed except in N9_closed where Arg292 was replaced with Lys292.

In addition, ZN88 was able to form hydrogen bond with Glu276 and Glu277 in all the cases. Though hydrogen bond was not observed between Asp151 and side chain of ZN88, the distance was close enough to make

interactions. It is possible that a slight change in conformation might lead to formation of hydrogen bond. The concluding remarks of this work have been described in Chapter 9.

Chapter 6

MCL-1

This chapter describes one of the most important biological processes known as apoptosis and its role in maintaining cell balance (6.1). It also describes the control of apoptosis by BCL-2 family of proteins and implications in cancer (6.2). Importance of Myeloid cell leukemia-1 (MCL-1) in apoptosis and as a drug target is discussed in 6.3 section.

6.1 Apoptosis

The study by Kerr et.al. (Kerr, Wyllie, & Currie, 1972) highlighted the importance of a process known as cell death or apoptosis. According to them, apoptosis plays a complementary and opposite role to mitosis, thereby regulating cell population. In multicellular organisms, the delicate balance between the number of cells to be eliminated and number of new cells produced is maintained by apoptosis (Meier, Finch, & Evan, 2000). Apoptosis is activated whenever tissue modeling is required and has many essential roles (Golstein, 1998) such as it eliminates the less fit cells in vertebrate epiblast (Claveria, Giovanazzo, Sierra, & Torres, 2013), shapes the embryo (e.g. by removing interdigital cells during limb formation) (Zuzarte-Luis & Hurle, 2002), eliminates autoreactive and non-functioning immature lymphocytes (Rathmell & Thompson, 2002).

6.1.1 Apoptosis and Cancer

In a seminal study, Weinberg and Hanahan (Hanahan & Weinberg, 2000) described six hallmarks traits that govern the transformation of normal cells to cancer cells. They include sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis. Recently, this list was updated to include four more traits i.e., genome instability and mutation, tumor-promoting inflammation, avoiding immune destruction and deregulating cellular energetics (Hanahan & Weinberg, 2011). Among all these, apoptosis has been the center of attraction as cancer cells try to evade cell death either by inactivation of pro-apoptotic or up-regulation of anti-apoptotic factors (Marsoni & Damia, 2004).

6.1.2 Apoptotic Pathways

Apoptosis involves the following steps: detection of stress signals, suppression of anti-apoptotic proteins, activation of pro-apoptotic proteins, release of apoptosis inducing agents from mitochondria inter-membrane space (IMS) such as cytochrome c and Smac/DIABLO, activation of caspases and finally cell death (Tait & Green, 2010). Caspases are effector molecules in apoptotic pathwa-

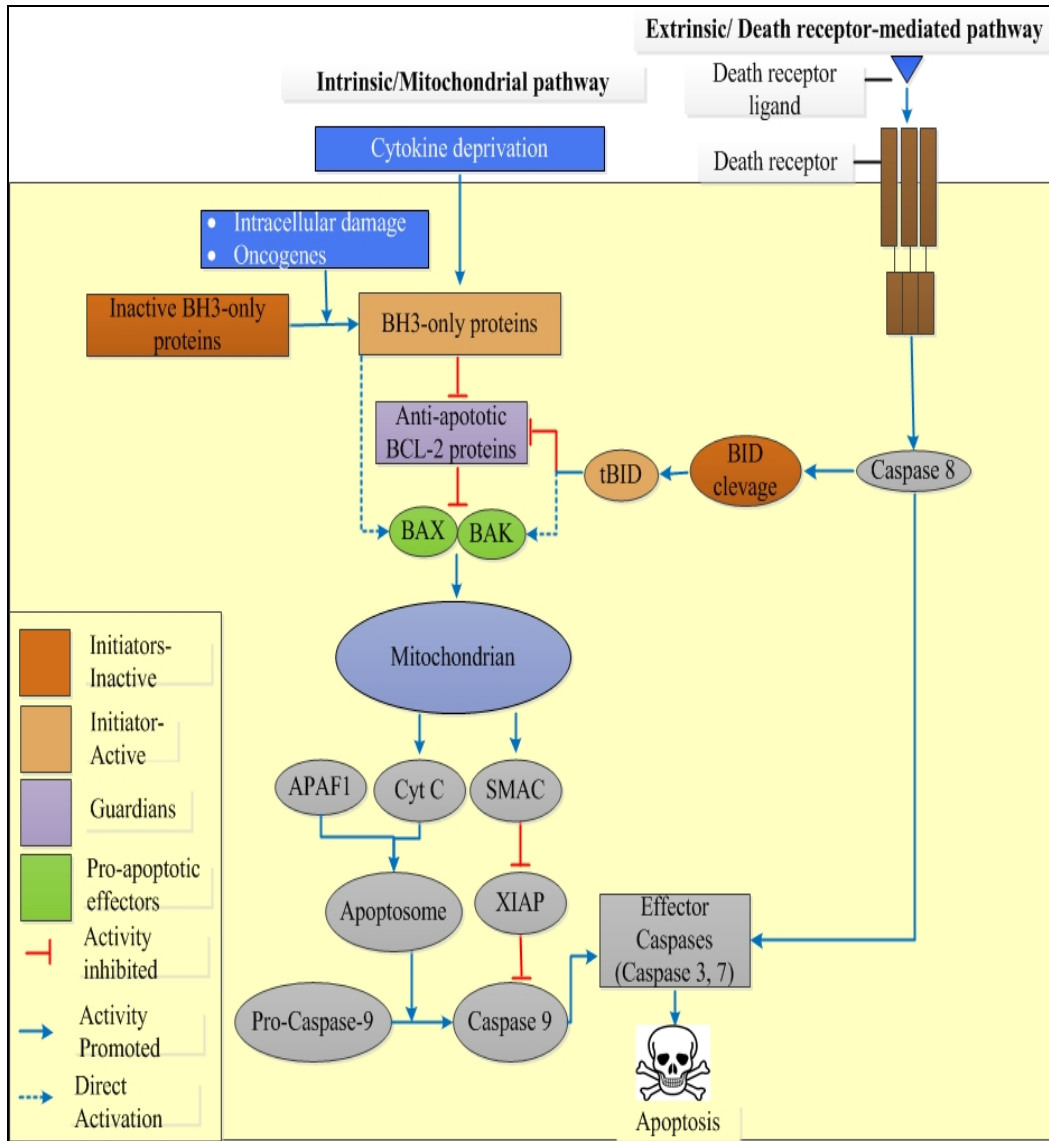


Figure6.1 The intrinsic and extrinsic apoptotic pathways (adapted from (Peter E. Czabotar, Lessene, Strasser, & Adams, 2014; Youle & Strasser, 2008)).

Cytotoxic stimulus like intracellular damage, cytokine deprivation initiates the intrinsic pathway by activation of BH3-only proteins. BH3 proteins inhibit anti-apoptotic BCL-2 proteins thus indirectly activating apoptotic proteins BAX and BAK. However, Some BH3-only proteins, such as BIM and PUMA, may also be able to directly activate BAX or BAK. BAX and BAK after oligomerization cause mitochondrial outer membrane permeabilization (MOMP) thereby releasing cytochrome C (Cyt C) and SMAC. Cytochrome C combines with APAF1 to form apoptosome which in turn activates pro-caspase 9 to caspase 9. SMAC inhibits XIAP (X-linked inhibitor of apoptosis protein) which can inhibit caspase 9. Caspase 9 activates effector caspases e.g. caspase 3, caspase 7 causing cell death. Extrinsic pathway is activated by various factors such as genotoxic agents. Death receptor on activation recruits protein with a death domain (DD) which in turn recruits pro-caspase 8 leading to its activation. Caspase 8 creates active form of BID i.e. tBID by cleavage. tBID links extrinsic pathway to intrinsic pathway.

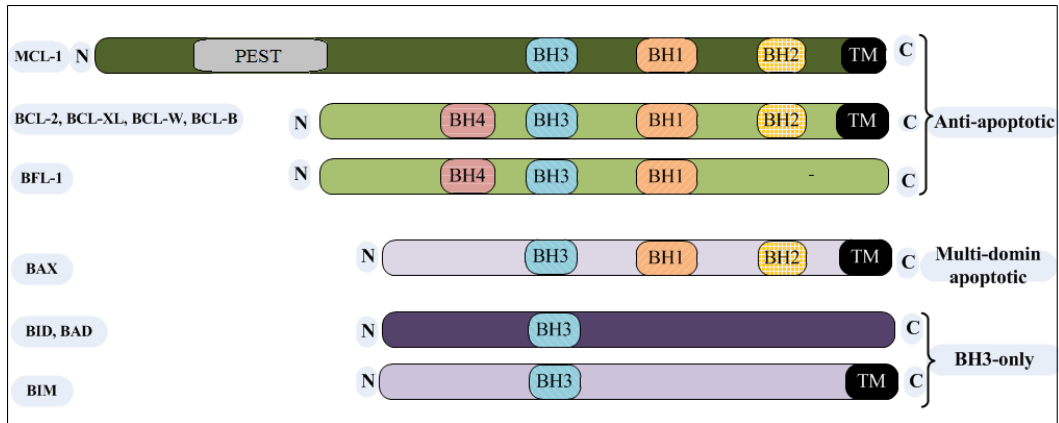


Figure6.2 Classification of core B-cell lymphoma-2 (BCL-2) family proteins on the basis of BCL-2 homology (BH) domains (adapted from (L. W. Thomas, Lam, & Edwards, 2010))

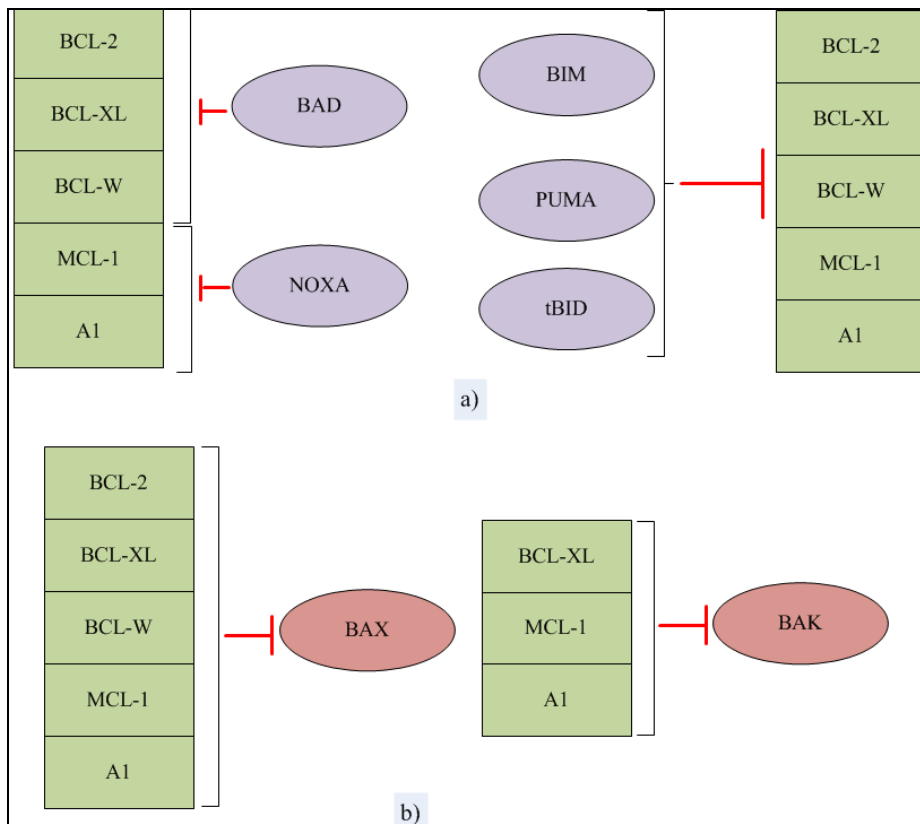


Figure6.3 The selective interactions within BCL-2 family members. (adapted from (Peter E. Czabotar et al., 2014))

a) It can be observed that promiscuous binders BIM, PUMA and tBID can inhibit all the anti-apoptotic proteins. On the other hand, NOXA interact only with MCL-1 and A1 and BAD only with BCL-2, BCL-XL and BCL-W and are known as selective binders b) Similarly BAX can be inhibited by all anti-apoptotic members while BAK is only inhibited by BCL-XL, MCL-1 and A1

-ys and are synthesized as inactive pro-caspases. On initiation of apoptotic pathway, these pro-caspases convert to active form via cleavage of a pro-domain mostly by other caspases (Favaloro, Allocati, Graziano, Di Ilio, & De Laurenzi, 2012). Caspases can be activated by two major pathways (**Figure6.1**) i.e. extrinsic pathway and intrinsic pathway (Fulda & Debatin, 2006).

6.2 BCL-2 Protein Family

BCL-2 was discovered over 20 years ago as a result of its upregulation in follicular B-cell lymphoma. This was a milestone discovery as overexpression of BCL-2 inhibited the cell death instead of promoting cell proliferation like most previously discovered oncogenes (Tsujimoto, Ikegaki, & Croce, 1987). It has been widely accepted that BCL-2 family of proteins, is essential for the development and maintenance of homeostasis.

BCL-2 family proteins are characterized by the presence of four regions of sequence homology i.e. BCL-2 homology (BH) domains. Based on structural and functional features, BCL-2 proteins family can be divided into three sub-families i.e. anti-apoptotic proteins, multi-domain pro-apoptotic proteins (also known as effector proteins) and BH3-only pro-apoptotic proteins (**Figure6.2**). The anti-apoptotic proteins includes BCL-2, BCL-XL, MCL-1, BCL-W and BCL-2A1; effector proteins consists of BAX, BAK and BOK; and BH3-only pro-apoptotic proteins comprise of BIM, BID, PUMA, BAD, HRK, NOXA and BMF.

6.2.1 BCL-2 family protein-protein interactions

The BH3-only proteins act upstream of BAX and BAK which can be deduced from the fact that in the absence of BAX and BAK, BH3-only proteins cannot induce apoptosis on their own. Initially, it was believed that activated BH3-only proteins can bind to all pro-survival counterparts. However, due to subtle differences in their BH3 domains and in the grooves of the anti-apoptotic proteins, it was found that different BH3-only proteins have varying affinity towards anti-apoptotic counterparts (**Figure 6.3**). BIM, PUMA and tBID can bind to all the pro-survival proteins and are known as promiscuous binders. On the other hand, NOXA interact only with Mcl-1 and A1 and BAD only with BCL2, BCL-XL and BCL-W and are known as selective binders (Weyhenmeyer, Murphy, Prehn, & Murphy, 2012).

6.2.2 BCL-2 family proteins as therapeutic targets

BCL-2 family members have essential roles right from early embryogenesis to adult tissue homeostasis. BCL-2 family of proteins regulates apoptosis which is important for embryonic development and prevention of cancer. In addition to apoptosis, Bcl-2-family proteins regulate other types of cell death, including necrosis and autophagy, thus acting as nodal points where multiple pathways converge (Yip & Reed, 2008). BCL-2 proteins also play critical roles in non-cancerous cells by maintaining neuronal activity, autophagy, calcium handling, mitochondrial dynamics and energetics (Hardwick & Soane,

2013). Some of the important effects caused by their deletion can be seen in **Table 6.1** (Peter E. Czabotar et al., 2014; Youle & Strasser, 2008)

Table 6.1 Physiological role of BCL-2 protein families	
Anti-Apoptotic members	
BCL-2	Abnormal death of renal epithelial progenitors, death of mature B and T lymphocytes causing defective immune system, premature greying
BCL-XL	Death of fetal erythroid progenitors and neuronal cells. Loss of a single allele encoding BCL-XL decreases spermatogenesis and reduces platelet numbers, while loss of both alleles kills hepatocytes resulting in liver fibrosis
BCL-W	Death of developing sperm cell leading to male sterility
A1	Death of granulocytes and mast cells
MCL1	Loss of MCL1 causes failure in implantation, premature death of B and T lymphoid cells, cardiomyopathy and heart failure
Effector Proteins	
BAX	Mild lymphoid hyperplasia, male sterility
BAK	Elimination of BAK results in excess platelets as a result of increased platelet lifespan
BH3-only Proteins	
BIM	Lymphoid and myeloid cell hyperplasia, SLE-like autoimmune disease, abnormal resistance to cytokine deprivation, deregulated calcium flux
BID	Resistance to hepatocyte killing, fatal hepatitis
PUMA	Resistant to DNA damage, cytokine deprivation and glucocorticoids
BAD	Resistance to loss of epidermal growth factor or insulin growth factor
HRK	Mild resistance to deprivation of nerve growth factor
BIK	No defects till date
NOXA	Mild resistance of fibroblasts to γ -irradiation , but high resistance of fibroblasts and keratinocytes in the skin to UV radiation

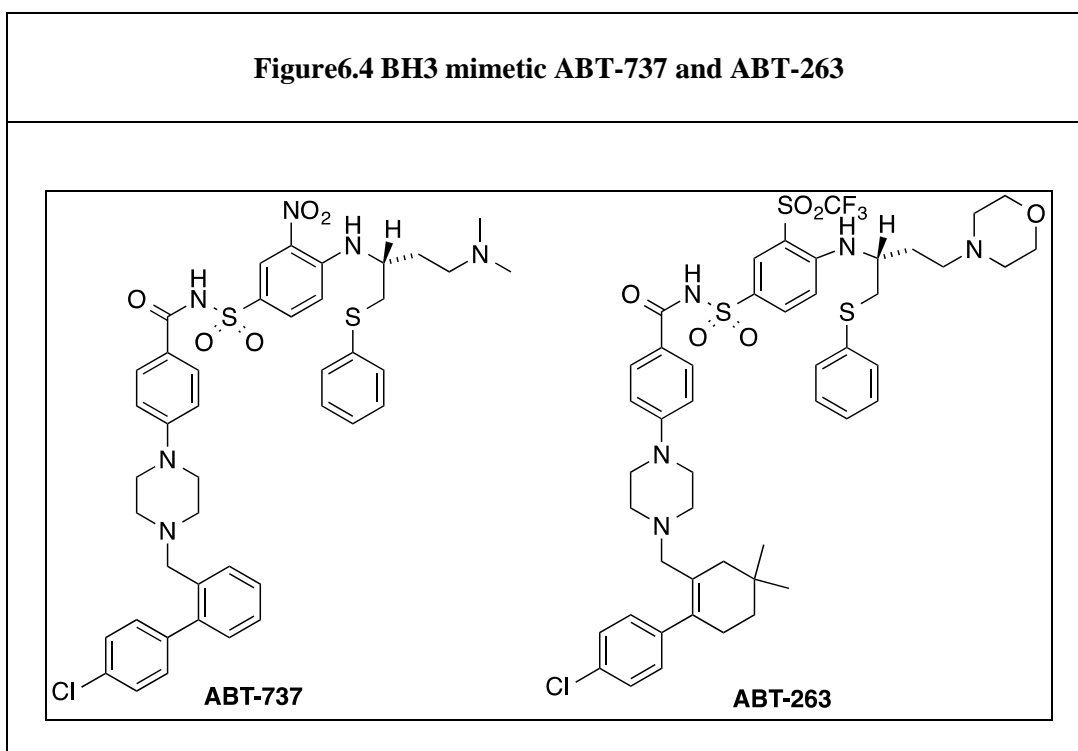
Any flaw in the apoptotic pathway causes several diseases, particularly cancer and autoimmunity. Inhibition of apoptotic pathway not only causes tumor progression but also leads to resistance of diverse tumours against chemotoxic drugs (Peter E. Czabotar et al., 2014). On the other hand too much apoptosis can enhance ischaemic conditions leading to neurodegeneration (Peter E. Czabotar et al., 2014).

6.2.3 BH3 mimetic as potential drugs

Targeting the regulation of anti-apoptotic BCL-2 family members is the most attractive approach. Most of the tumors show increased levels of BH3-only proteins due to high levels of anti-apoptotic proteins. Such findings along with the success of stapled BH peptides have resulted in the search for BH3 mimetics. However, the long, shallow and mainly hydrophobic groove of the anti-apoptotic BCL-2 proteins is more challenging and thus only a few slightly potent MCL1 inhibitors have been discovered. Several small-molecule BH3 mimetics have been identified but most of them bind their targets with moderate affinity (Peter E. Czabotar et al., 2014; Juin, Geneste, Gautier, Depil, & Campone, 2013).

Following the line of BH3 mimetics, two potential drugs ABT-737, a BH3 mimetic modeled after BH3 domain of BAD (Oltersdorf et al., 2005), and ABT-263 (Tse, 2008) have been launched (**Figure 6.4**). Both ABT-737 and ABT-263 bind strongly to BCL-2, BCL-XL and BCL-W but not to MCL-1 (Lee et al., 2007; Oltersdorf et al., 2005). Surprisingly the target of these drugs is BIM-BCL-2 complex instead of unoccupied BCL-2. Furthermore, their affinity towards

BIM-BCL-2 complexes is much more than BIM-BCL-XL or BIM-BCL-W complexes (Merino et al., 2012). Thus, increased level of BCL-XL, MCL-1 or A1 can lead to the resistance against ABT-737 and ABT-263. Moreover, the ability of ABT-263 to act on BCL-XL causes toxicity due to role of BCL-XL in controlling the platelet lifespan (Roberts et al., 2012). These findings indicate that BH3 mimetics targeting single anti-apoptotic proteins might be better approach. The selective target approach was applied in ABT-199, which is a selective BCL-2 inhibitor.



6.2.4 MCL-1 as a drug target

MCL-1 is among the most frequently amplified genes in human cancer (G. Wei et al., 2012) and has been linked to several cancer including lung, breast,

prostate, pancreatic, ovarian, melanoma, leukemia and cervical cancers, making it an appealing target (Akgul, 2009; Brotin et al., 2010; Cavarretta et al., 2006; Ding et al., 2007; Goncharenko-Khaider, Matte, Lane, Rancourt, & Piche, 2012; Gores & Kaufmann, 2012; Mitchell et al., 2010; L. Song, Coppola, Livingston, Cress, & Haura, 2005; H. Zhang et al., 2011; Zhou et al., 2013). Moreover, MCL-1 overexpression has emerged as a resistance mechanism against a number of anticancer drugs such as ABT-263 (Tse, 2008), ABT-737 (Yecies, Carlson, Deng, & Letai, 2010), BT-199 (Souers et al., 2013) and WEHI-539, as well against gemcitabine (S. H. Wei et al., 2008), the widely prescribed drug for pancreatic cancer (Lessene, 2013). Though the crucial physiological role of MCL1 suggests that targeting it might produce severe side effects, many cancers like acute myeloid leukaemia cells are more sensitive to the loss of MCL1. Hence, there is need of MCL-1 inhibitors with a well-defined therapeutic window.

6.3 MCL-1

The human MCL-1 gene is located on chromosome 1q21 and consists of three exons. As a result of alternative splicing, MCL-1 gene translate into two isoforms i.e. MCL-1L and MCL-1S. MCL-1L (or generally called as MCL-1) unlike other anti-apoptotic members possess three BH domains (**Figure6.2**) i.e. BH1-BH3 and lacks BH4 domain (Akgul, 2009). MCL-1 contains two PEST sequences i.e. sequences containing mainly of proline (P), glutamic acid (E), serine (S) and threonine (T) at N-terminal (Le Gouill, Podar, Harousseau, & Anderson, 2004). These PEST sequences are characteristic of MCL-1 and respon-

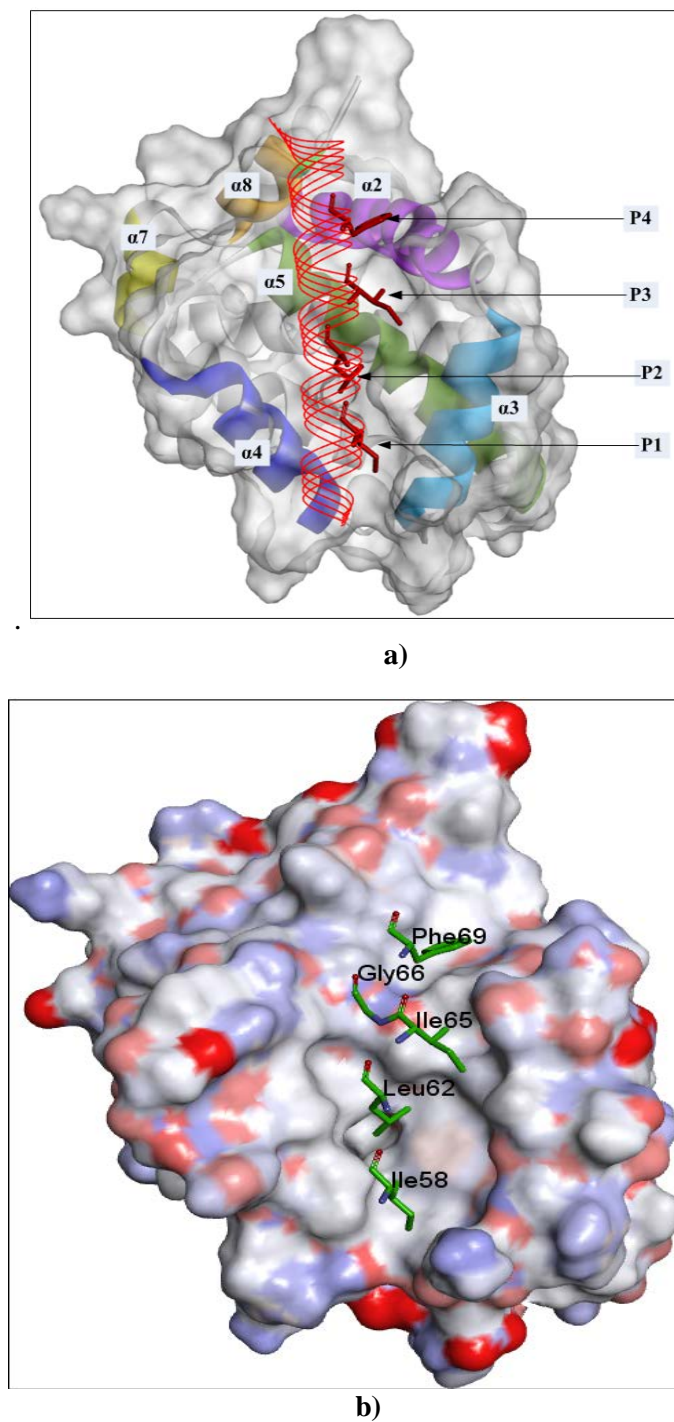


Figure6.5 Structure of MCL-1

a) Structure of MCL-1 revealing the hydrophobic groove formed by $\alpha 2$ -4 and $\alpha 8$ helices with $\alpha 5$ forming core of the hydrophobic groove. The pockets P1- P4 are defined by residues 2d, 3a, 3d and 4a respectively of BIM-BH-3 only protein. **b)** It can be observed that pocket P2 is deepest among all the pockets and pocket P4 is separated from the rest three pockets by a saddle like intervention with which Gly266 is aligned. The hydrophobic groove is flanked by electropositive residues shown by blue colour of the surface.

-sible for its short half-life (Akgul, 2009). Similar to other anti-apoptotic members, MCL-1 contains C-terminal transmembrane domain that helps to insert into different intracellular membranes including mitochondria (Michels, Johnson, & Packham, 2005). MCL-1 is localized at intracellular membranes including mitochondria as well as in cytoplasm. Localization of MCL-1 at outer mitochondrial membrane is important for its anti-apoptotic function. MCL-1 localized at mitochondrial membrane inhibits apoptosis by forming dimer with pro-apoptotic member BAK thereby suppressing cytochrome-c release from mitochondria (Akgul, 2009).

6.3.1 MCL-1 function

In vitro studies indicate that similar to other anti-apoptotic members, MCL-1 sequesters BAK, thereby preventing its oligomerization. This interaction is disrupted by NOXA, BIM, BID and PUMA causing cell death (M. R. Warr & Shore, 2008). Different BH3-only proteins have different effect upon binding with MCL-1. It has been shown that BIM and PUMA stabilizes MCL-1 while NOXA induces proteasomal-dependent MCL-1 degradation (P. E. Czabotar et al., 2007).

6.3.2 MCL-1 versus BCL-2 family member's specificity

The topology of MCL-1 is similar to other BCL-2 family member proteins. MCL-1 is composed of eight helices i.e. $\alpha 1$ (residues 172-191), $\alpha 2$ (residues 204-221), $\alpha 3$ (residues 225-235), $\alpha 4$ (residues 244-256), $\alpha 5$ (residues 261-281), $\alpha 6$ (residues 288-301), $\alpha 7$ (residues 303-308) and $\alpha 8$ (residues 312-

318) where central hydrophobic $\alpha 5$ helix is enveloped by rest seven helices. However, the groove is composed of $\alpha 2-4$ and $\alpha 8$ helices surrounding $\alpha 5$ helix which forms core of groove (**Figure6.4a**). Despite the fact that MCL-1 has similar structure, it has a distinct BH3 binding profile (**Figure6.3**).

Although MCL-1 has a very similar structure with other anti-apoptotic BCL-2 proteins, having less than 2Å backbone root mean square deviation (RMSD) over conserved helices, it only shares ~25% sequence identity (Day et al., 2005; Fire, Gullá, Grant, & Keating, 2010). Comparison of MCL-1 and BCL-XL reveals that $\alpha 3$ helix is longer in MCL-1. MCL-1 surface is electropositive in nature due to fourteen lysine and four histidine residues present at $\alpha 3$, $\alpha 3-\alpha 4$ loop, and $\alpha 4$ while BCL-XL groove is essentially uncharged (**Figure6.4b**). Moreover, the binding groove in BCL-XL is tightly packed due to hydrogen bond between Gln111 of $\alpha 3$ and Glu129 of $\alpha 4$ in comparison to Lys215 and His233 in MCL-1. This results in significantly different BH3 binding profile of MCL-1 in comparison to other anti-apoptotic members (Day et al., 2005; Quinn et al., 2011).

6.3.3 BH3 and interaction with MCL-1

The sequences of BH3 peptides are highly variable. However, for all BH3 the residues follow a heptad repeat, [abcdefg]_n, with ‘‘a’’ and ‘‘d’’ positions buried in the hydrophobic pocket of the respective partner. It has been observed that residues at position 3e (mostly having small residue like glycine), aspartic acid at 3f and four hydrophobic residues at positions 2d, 3a, 3d and 4a (4a means residues at a position in 4th heptad repeat) are more or less conserved. The

pockets corresponding to the four hydrophobic residues 2d, 3a, 3d and 4a in anti-apoptotic members are known as P1, P2, P3 and P4 respectively (**Figure6.4a**).

The difference in residues at positions 2d, 3a, 3d and 4a can lead to difference in specificity (Fire et al., 2010). In BH3-only proteins, 2d is the least conserved of all four hydrophobic residues i.e. 2d, 3a, 3d and 4a while the hydrophobic residues 3a and 3d are highly conserved. Though residues 3a and 3d provide stability but their role in selectivity is yet not clear (Day et al., 2005).

As can be seen from **Figure6.4b** the pockets are of varying depths and sizes. The P2 pocket is the deepest among them and is contiguous with the shallower pockets P1 and P3. Residue 3e aligns with a saddle point which separates the pocket P1, P2 and P3 from P4. Structure compounds targeting BCL-2, BCL-XL or MCL-1 reveals that all compounds are anchored in the P2 hydrophobic pocket. The P2 pocket appears to be critical for binding all of the ligands because of its plasticity and ability to accommodate the BH3 mimetics into deep cavities not present with natural ligands (P. E. Czabotar et al., 2007). The important role of the conserved residues 2d, 3a, 3d, 4a and 3g are as follows:

6.3.3.1 Position 2d

The binding pocket P1 surrounding 2d is formed by residues of $\alpha 3$ and $\alpha 4$ helices i.e. Met231, Lys234, Leu235, Val249 and Phe270. Moreover, it seems that the interaction with pocket P1 forms the selective binding criteria (Day et al., 2005).

MCL-1 can undergo local conformational change at the 2d position to accommodate isoleucine, smaller residue alanine or even larger residues like tyrosine. In case of smaller residue alanine Leu235 shifts to fill the cavity. While for large residue like tyrosine, a more prominent movement of $\alpha 3$ region of MCL-1 by 1.7 Å away from the peptide and 1.2 Å movement of peptide away from MCL-1 is observed. This movement is facilitated by the local conformational changes in peptide and $\alpha 3$ helix of MCL-1. (Fire et al., 2010).

6.3.3.2 Position 3a

Pocket P2, engaged by 3a, is formed by Met231, Val249, Val253, Leu267, Phe270 in case of MCL-1. This pocket cannot accommodate charged or polar residues (Fire et al., 2010).

6.3.3.3 Positions 3d

Residues Val220, His224, Ala227, Phe228, Met231 and Thr266 forms pocket P3 enveloping 3d. Due to less helical $\alpha 3/\alpha 2$ region of the receptor, 3d site is less tightly packed in BCL-XL in comparison to MCL-1. In MCL-1, the site is constrained to isoleucine whereas BCL-XL can accommodate a range of residues. (Dutta et al., 2010; Fire et al., 2010).

6.3.3.4 Position 4a

The pocket P4 formed by Val216, Val220, Val265, Phe319 for MCL-1 is found to be more open and solvent exposed in MCL-1 as compared to BCL-XL. This makes MCL-1 tolerant to many mutations at site 4a while BCL-XL is more

constrained at this site. If 4a site is replaced with charged or polar amino acids, affinity of the BH3 only protein to BCL-XL drops significantly (Fire et al., 2010).

6.3.3.5 Positions 3g

Presence of negatively charged residues at 3g position makes BH3 only proteins BCL-XL specific due to presence of arginine. On the other hand, NOXA, a MCL-1 specific BH3-only protein, has a lysine at this position. Mutation of lysine to glutamic acid increases the affinity of NOXA towards BCL-XL (Fire et al., 2010).

Comparison of apo and holo forms of both BCL-XL and MCL-1 reveals that binding groove needs to open in order to accommodate BIM BH3 peptide. The structural change accompanied with binding groove opening is largest in P2 and P3 (P. E. Czabotar et al., 2007). While, MCL-1 achieves it by reorientation of carboxy-terminal end of $\alpha 4$ helix, BCL-XL does it by shifting the residues that compose $\alpha 3$ helix (Fire et al., 2010). This can be attributed to the higher flexibility of $\alpha 4$ helix in MCL-1 and $\alpha 3$ helix in BCL-XL (C.-Y. Yang & Wang, 2012).

6.3.4 Targeting MCL-1

Numerous approaches have been applied to target MCL-1. However, most of them are not MCL-1 specific and target multiple anti-apoptotic proteins. BH3 mimetic is one of the most promising approaches and has been applied to discover some of the recent drugs.

6.3.4.1 *ABT-737*

It is one of the most successful BH3 mimetic and has strong affinity towards anti-apoptotic members of the BCL-2 family. The chloro-biphenyl moiety and thio-phenyl moiety of ABT-737 occupies P2 and P4 pocket respectively. MCL-1 and BCL-XL differs in P2 pocket. Moreover, in MCL-1, P4 pocket is relatively more open and exposed towards the solvent. Hence, ABT-737 does not bind with MCL-1, thereby fails in cancer with overexpression of MCL-1 (Lee et al., 2007).

Chapter 7

MCL-1 Methods

The methods used to explain the activity of the compounds against MCL-1 are elaborated in this chapter. Docking and molecular dynamics are explained in 7.1 and 7.2 sections respectively. The specific parameters for MCL-1 are also described in these sections.

7.1 Docking

7.1.1 Structure preparation

The structure of MCL-1 (PDB id 2NL9) was downloaded from PDB. Ligand and water molecules were deleted from PDB structure and the resulting structure was prepared using MOE. The structure was prepared by addition of hydrogen atoms, structure correction, protonation states according to pH 7 and energy minimization of hydrogen atoms keeping heavy atoms fixed. The calculation of partial charges and energy minimization was performed using AMBER99 within MOE.

7.1.2 Active site

Active site was determined by selecting the residues within 6Å of the natural peptide. It comprised of Arg215, Val216, Gly217, Gly219, Val220, His224, Ala227, Phe228, Gly230, Met231, Leu232, Lys234, Leu235, Asp236, Lys244, Ser245, Leu246, Arg248, Val249, Met250, His252, Val253, Phe254, Ser255, Asp256, Gly257, Val258, Asn260, Trp261, Gly262, Arg263, Ile264, Val265, Thr266, Leu267, Phe270, Phe318, Phe319, His320 and Val231.

7.1.3 Dataset for docking

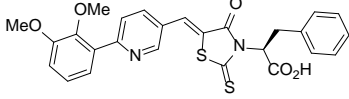
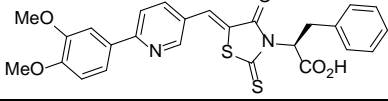
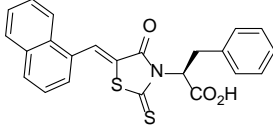
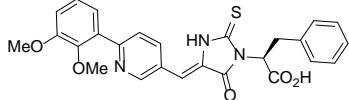
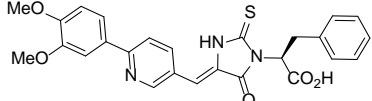
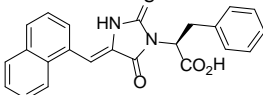
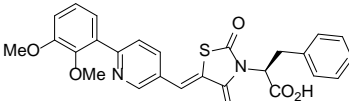
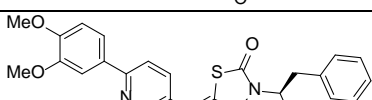
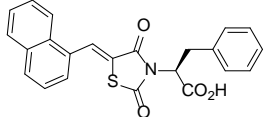
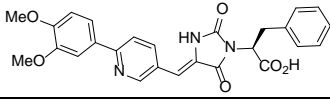
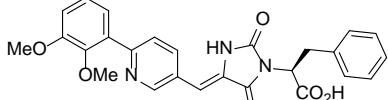
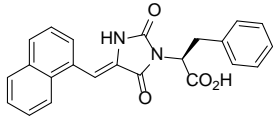
In the study by Bernardo et al. (Bernardo et al., 2010), compounds having rhodanine scaffold, i.e. ST_1_046 and ST_1_109 **Table 7.1**, were reported to be active against MCL-1. ST_1_046 and ST_1_109 exhibited different pro-survival protein selectivity. While ST_1_046 acted as selective Mcl-1 inhibitor, ST_1_109 showed activity against both Mcl-1 and Bcl-XL. The selectivity of ST_1_046 towards Mcl-1 was credited to the para-methoxy group, which fits perfectly in the binding groove of Mcl-1 but faces steric repulsion in the binding groove of Bcl-XL. In the work by Bernardo et al., effect of substituent groups towards selectivity and activity was studied but function of the rhodanine scaffold was not understood. Rhodanine-based compounds are known as pan assay interference compounds (PAINS) due to frequent hits in screening campaigns. Rhodanine-based compounds can non-specifically interact with proteins in multiple ways due to more than one reactive site present on rhodanine. Thus activity of compounds

possessing a rhodanine moiety is considered very skeptically despite convincing results (Tomasic & Peterlin Masic, 2012).

To understand the functional role of the scaffold in interaction with anti-apoptotic proteins, our collaborator Miss Tang Shi Qing graduate student Dr. Christina CHAI, used scaffold-hopping strategy to swap the rhodanine scaffold with structurally related five-membered multiheterocyclic rings, i.e. thiohydantoin, hydantoin and thiazolidinediones **Table 7.1** were used for docking. These rings differ in the presence of the exocyclic and endocyclic sulfur atoms. The compounds synthesized have biaryl substituents similar to ST_1_046 and ST_1_109 for direct comparison. In addition to similar biaryl substituents, bicyclic methylene substituent and naphthalene-1-yl methylene was also synthesized to analyze the effect of multicyclic arylidenes versus biaryl arylidenes in the different heterocyclic systems.

7.1.3.1 Fluorescence polarization assay

The inhibitory activity of the compounds against Mcl-1 and Bcl-XL was evaluated with fluorescence polarization assay (FPA) by our collaborator Miss Tang Shi Qing graduate student Dr. Christina CHAI. FPA was performed by using a fluorescein tagged BH3 domain of Bak peptide as the fluorescent source. The Bak peptide was chosen because it is known to bind to both proteins of interest. The optimal protein concentration to be used for the assay was firstly determined via a protein titration with 60 nM of Flu-Bak ligand. From the titration curves, the optimal protein concentration for Mcl-1 and Bcl-XL were found to be

Table7.1 Dataset for Mcl-1 studies				
	Name	Structure	Class	IC ₅₀ (μM)*
1	ST_1_109		rhodanines	22.31
2	ST_1_046		rhodanines	10.64
3	ST_1_R1N		rhodanines	17.84
4	ST_1_202		thiohydantoin	58.5
5	ST_1_208		thiohydantoin	38.97
6	ST_1_247		thiohydantoin	38.46
7	ST_1_222		hydantoin	57.68
8	ST_1_227		hydantoin	45.90
9	ST_1_261		hydantoin	57.07
10	ST_1_159		thiazolidinedione	>100
11	ST_1_162		thiazolidinedione	>100
12	ST_1_249		thiazolidinedione	>100

* The IC₅₀ values are the average of three independent experiments; Values are the average ± standard error of the mean .

63.75 $\mu\text{g/ml}$ and 37.5 $\mu\text{g/ml}$ respectively. Using the optimal protein concentration, the IC₅₀ values of our compounds were obtained by performing a 10-point 2-fold dilution of the compounds with 60 nM of Flu-Bak and 63.75 $\mu\text{g/ml}$ and 35.60 $\mu\text{g/ml}$ Mcl-1 and Bcl-XL respectively. All the IC₅₀ values determinations were carried out in triplicates and average IC₅₀ was reported Table 7.1

7.1.4 Molecular Docking

Docking was performed in MOE by default parameters. Primary challenge in a docking study is prediction of the correct binding pose. The involvement of rhodanine moiety enhanced this problem thus post-processing of the docking results was performed. Different methods for post-processing of the docking results such as using MPBSA (Lindstrom et al., 2011) clustering (Kozakov, Clodfelter, Vajda, & Camacho, 2005) and short MD simulation have been applied (Yuriev & Ramsland, 2013). In our work poses predicted by MOE were clustered and the pose with best docking score of highest cluster was selected for further analysis.

7.2 Molecular Dynamics

7.2.1 System preparation

For molecular dynamics structure with PDB id 4HW4 was used which has complete human sequence as compared to 2NL9. The best pose predicted by docking was overlapped with 4HW4 and was used as starting structure. The partial charges and force field parameters for the compounds were generated

automatically by the antechamber suite in AMBER12 (D.A. Case, 2012). The general AMBER force field (gaff) (J. Wang et al., 2004) was used for the compounds, and AMBER ff12SB force field for the proteins. All missing hydrogen atoms of the proteins were added using the LEaP module in AMBER12. Systems were solvated using TIP3P (Jorgensen et al., 1983) water in a octahedron box extending 12 Å beyond any solute atom. The system was made neutral by adding appropriate number of counter ions.

7.2.2 Minimization, heating up and equilibration of system

Molecular dynamics simulations were performed using pmemd.cuda in AMBER12. Water molecules and counter ions were minimized with 500 cycles of steepest descent followed by same number of conjugate gradient cycles. This was followed by energy minimization of the entire system. Molecular dynamics simulations were performed using a time step of 2 fs with a cutoff radius of 8Å for the non-bonded interactions. Long range electrostatic interactions were calculated by particle-mesh Ewald (PME) method (Darden, York, & Pedersen, 1993). The bonds involving hydrogen were constrained using SHAKE algorithm (Ryckaert, Ciccotti, & Berendsen, 1977) and Langevin dynamics was used for temperature control. The temperature was gradually increased from 0k to 310K with protein restraints over a period of 20ps at NVT conditions. This was followed by 1000ps of NPT equilibration at 1 atm pressure and later succeeded by production run. The periodic boundary conditions (PBC) were used during molecular dynamics simulations.

7.2.3 Production run

It has been suggested that single longer trajectory approach does not necessarily lead to good results (T. Hou, J. Wang, Y. Li, & W. Wang, 2011). Moreover, in a recent study it was shown that a single run samples the system inadequately and replicate trajectories enhances the search of conformational space (Adler & Beroza, 2013). Coordinates of not more than 1% of the atoms of the receptor was altered by 0.001 Å while keeping the initial velocities unchanged of the equilibrated systems. The production run was performed on these modified coordinates to generate five replicate trajectories of 10ns each. However, considering the flexibility of the $\alpha 4$ helix, comprising key residues forming P1 and P2 pocket, a single longer simulation of 50ns was also generated.

The RMSD of the backbone atoms was calculated for the trajectories. Depending on the analysis the later 5ns of each replicate trajectory was merged into a single trajectory of 25 ns for further analysis. Similarly first 5ns of the trajectory was discarded for longer 50ns trajectory. However, this single longer trajectory 45ns (will be referred to as 45NsT from now on) was not merged with shorter 25ns one (will be referred to as 25NsT from now on) to avoid bias for one particular pose. The clustering analysis was performed for the both 25NsT and 45NsT by cpptraj of AMBER12.

7.2.4 Binding free energy

The binding free energy for each system was calculated by the MM/GBSA approach. MM/GBSA was selected as it has been shown in many recent studies that MM/GBSA produces results with comparable accuracy to MM/PBSA and outperforms its counterpart while calculating relative binding free energies at less computational cost (Tingjun Hou, Junmei Wang, Youyong Li, & Wei Wang, 2011; Srivastava & Sastry, 2012). MMGBSA approach combines molecular mechanical energies with the continuum solvent approaches using sander program from Amber12. The binding free energy (ΔG_{bind}) of the compounds in each complex was calculated according $\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{protein}} - G_{\text{ligand}} = \Delta H + \Delta G_{\text{solvation}} - T\Delta S = \Delta E_{\text{MM}} + \Delta G_{\text{GB}} + \Delta G_{\text{SA}} - T\Delta S$ (1 where ΔE_{MM} is the gas-phase interaction energy between protein and ligand, including the electrostatic and Van der Waals interactions; ΔG_{GB} and ΔG_{SA} are the polar and non-polar components of the desolvation free energy and $T\Delta S$ is the change of the conformational entropy upon ligand binding.

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{protein}} - G_{\text{ligand}} = \Delta H + \Delta G_{\text{solvation}} - T\Delta S = \Delta E_{\text{MM}} + \Delta G_{\text{GB}} + \Delta G_{\text{SA}} - T\Delta S \quad (1)$$

The gas phase free energy contributions are calculated by sander within the Amber program suite. The solvation free energy contributions can be further decomposed into an electrostatic and hydrophobic contribution. The electrostatic portion was calculated using Generalized Born (GB) method with $\text{igb} = 2$ (Onufriev, Bashford, & Case, 2000, 2004), with salt concentration 0.1 Molarity and PM3 semi-empirical Hamiltonian used for the quantum calculation. The

hydrophobic contribution is approximated by the LCPO method (Weiser, Shenkin, & Still, 1999) implemented within sander.

Chapter 8

MCL-1 Results and Discussion

The results of the docking and molecular dynamics study to explain activity against MCL-1 are explained in this chapter. The results of the clustering and analysis of the subsequent poses are discussed.

8.1 MCL-1 versus BCL-XL

As a result of overlapping residues it was observed that residues such as Ala93, Glu96, Phe97, Arg100, Tyr101, Leu112, Ser122, Gln125, Val126, Leu130, Gly138, Arg139, Ala142, Phe143, Phe146 and Tyr196, that are part of binding pocket or close to it are more pronounced in BCL-XL. More than other residues, Phe146 and Phe97 seem to be having more effect by reducing depth of P2 and P3 cavity respectively. The P2 cavity was observed to be distinguishing feature as it seems to be deeper than BCL-XL.

8.2 Docking

The poses of the compounds indicated that the active compounds utilize the P2 pocket. However, the results could not explain the reason why certain compounds were strongly active while others were not.

8.3 Molecular Dynamics

The trajectories for most of the simulations were found to be equilibrated as shown by the RMSD calculation of the backbone atoms (**Figure8.1**, **Figure8.2**).

8.3.1 Clustering

Clustering analysis performed on 25NsT (**Table 8.1**) shows that for ST_1_159, ST_1_261 there was no consensus and ligand occupied different orientations. Moreover, difference in size of first and second cluster ST_1_227 and ST_1_249 was not too large. The representative poses from both first and second cluster was considered for analysis of these compounds. For the remaining compounds, their representative pose from top cluster was selected for analysis.

Table 8.1 The cluster size of top three clusters is shown.			
Compounds*	Cluster_1	Cluster_2	Cluster_3
ST_1_109	1715	500	285
ST_1_159	1098	1000	397
ST_1_162	2000	440	60
ST_1_208	1500	500	471
ST_1_202	2500		
ST_1_222	1680	351	244
ST_1_227	838	500	361
ST_1_247	2499		
ST_1_249	1224	979	217
ST_1_261	500	500	500
ST_1_046	1653	500	347
ST_1_R1N	2500		

*2500 can be the highest number of poses possible

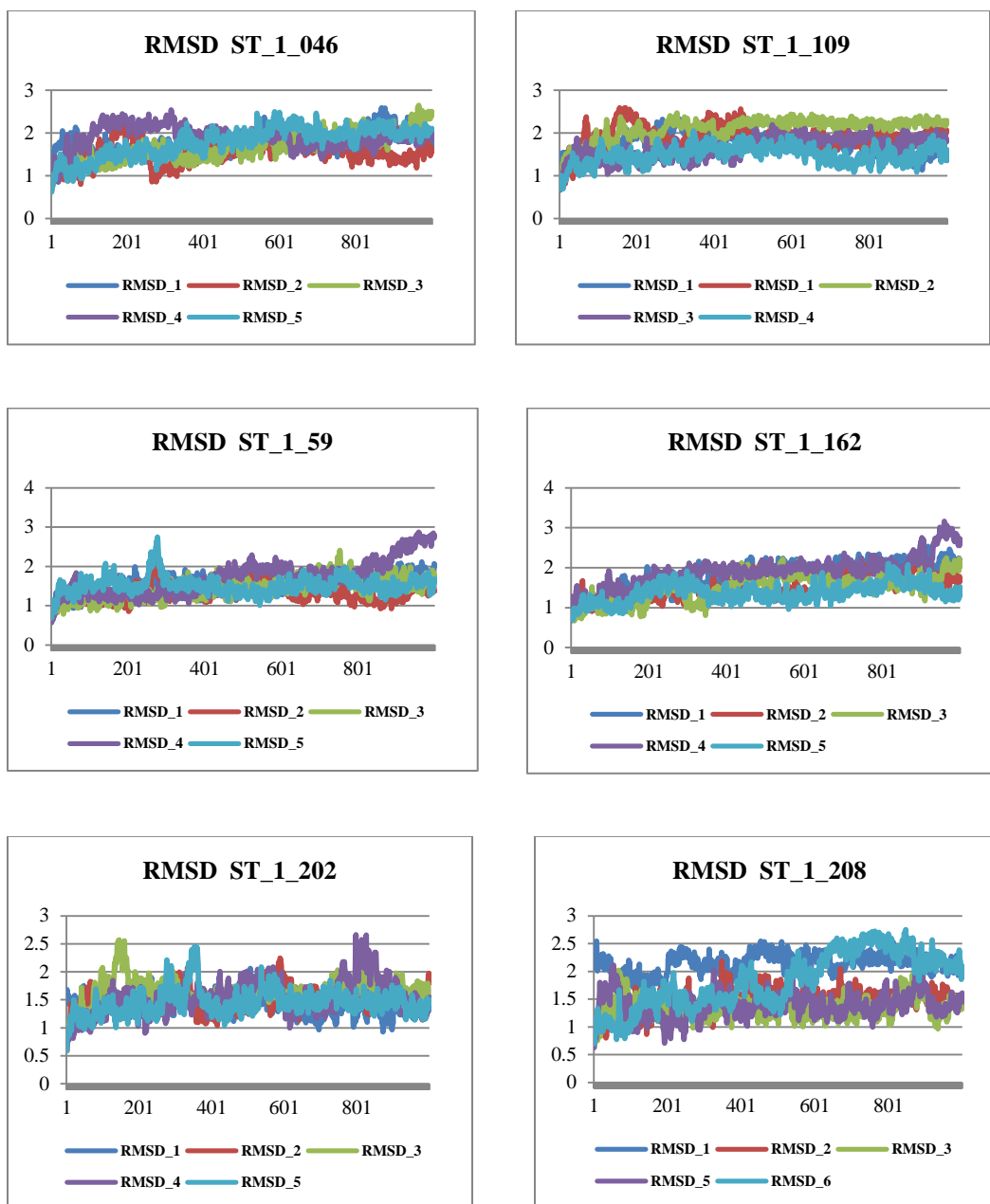


Figure8.1 RMSD comparison of the backbone atoms between five trajectories

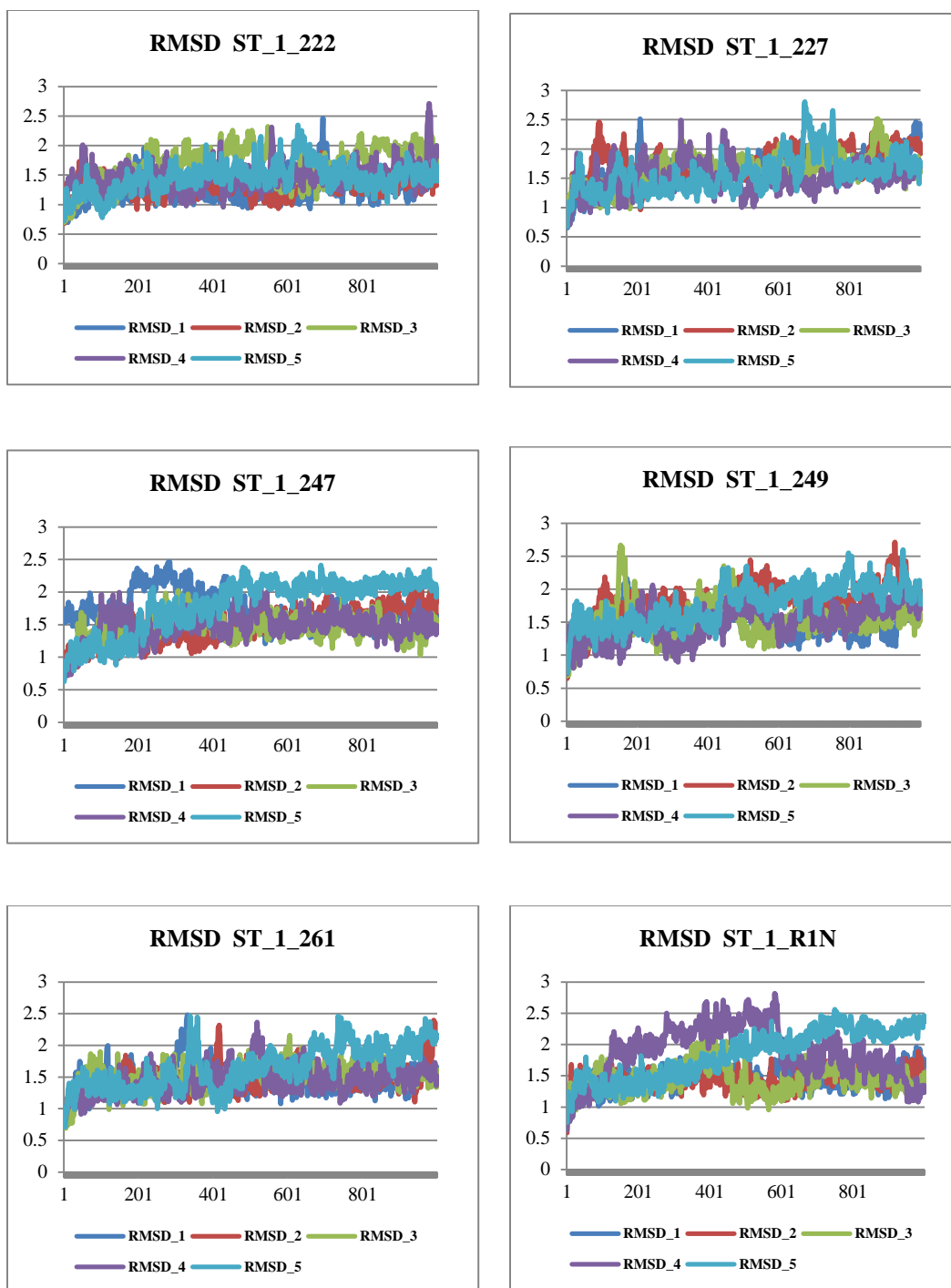


Figure8.2 RMSD comparison of the backbone atoms between five trajectories continued.

8.3.2 Binding free energy calculation

Table 8.2 shows the average binding free energy (Avg_GBSA) calculated from 5 trajectories, which has a Pearson correlation coefficient of 0.2693 with IC₅₀. However, after discarding values >100 from calculation value of the Pearson correlation coefficient increased to 0.655. Avg_GBSA was not able to clearly rank the compounds according to activity, as ST_1_046 was ranked third according to Avg_GBSA (Avg) even though it is the strongest inhibitor.

Compounds	Traj1	Traj2	Traj3	Traj4	Traj5	Avg	IC ₅₀
ST_1_109	-31.8544	-31.1679	-23.6406	-30.6913	-32.0964	-29.8901	22.31
R1N	-29.7772	-28.3109	-30.349	-28.7014	-27.5232	-28.9323	17.84
ST_1_046	-27.7594	-29.812	-30.0837	-27.0106	-29.8793	-28.909	10.64
ST_1_202	-27.8631	-26.548	-27.0877	-27.3295	-27.0592	-27.1775	58.5
ST_1_222	-22.8788	-28.9803	-30.9283	-23.6671	-27.8489	-26.8607	57.68
ST_1_247	-28.5871	-26.5147	-25.968	-25.4701	-27.1897	-26.7459	38.46
St_1_159	-25.1196	-23.1986	-26.5482	-27.881	-24.0857	-25.3666	100
ST_1_162	-16.9434	-26.4017	-28.1194	-27.177	-27.8438	-25.2971	100
ST_1_249	-28.1785	-24.4087	-28.2246	-22.3472	-19.6618	-24.5642	100
ST_1_208	-18.313	-24.0395	-25.2114	-23.5915	-28.7542	-23.9819	38.97
ST_1_227	-23.0639	-24.9602	-29.8268	-27.8299	-12.1799	-23.5721	45.9
ST_1_261	-23.3401	-16.0467	-30.067	-13.7237	-27.7501	-22.1855	57.07

8.3.3 Interactions

8.3.3.1 ST_1_046

Clustering of pose for 25NsT reveals that ST_1_046 occupies P1, P2 and P3 pocket of the MCL-1 binding groove (**Figure8.3**). The 3,4 dimethoxy group

faces the base of the P1 cavity and forms van der Waals interaction with Leu246 and Val249, while phenyl ring forms π - π and π -alkyl interactions with Met231 and Phe270. Pyridine ring forms π -interactions with Val253 and Met231. In addition, Arg263 is involved in formation of salt bridge with carboxylic group and Van der Waal's interaction with exocyclic oxygen present in rhodanine ring. Exocyclic oxygen is involved in formation of Van der Waal's interaction with Thr266.

The visualization of the longer 45ns simulation and clustering of the pose for 45NsT indicates that ST_1_046 flips by 180° with exocyclic sulphur of rhodanine ring facing the base of P2 pocket (**Figure8.3**). This shifts the terminal phenyl and carboxylic group near P3 pocket, pyridine ring at P1 pocket while 3,4 dimethoxy phenyl ring shifts towards C-terminal of the natural peptide.

8.3.3.2 *ST_1_109*

The pose predicted by the clustering of 25NsT shows that P2 pocket is occupied by terminal phenyl ring and rhodanine. However, the rhodanine ring was not found deep inside the P2 cavity. Pyridine ring was found close to P1 pocket (**Figure8.3**). The inward movement of α 3 and outwards movement of α 4 loop resulted in bigger P2 pocket and smaller P3 pocket. Similar pose was observed from the clustering of longer 45NsT.

8.3.3.3 *ST_1_R1N*

The clustering of poses from 25NsT resulted in a single cluster indicating stability of predicted pose. The naphthalene ring was found to be parallel to P1

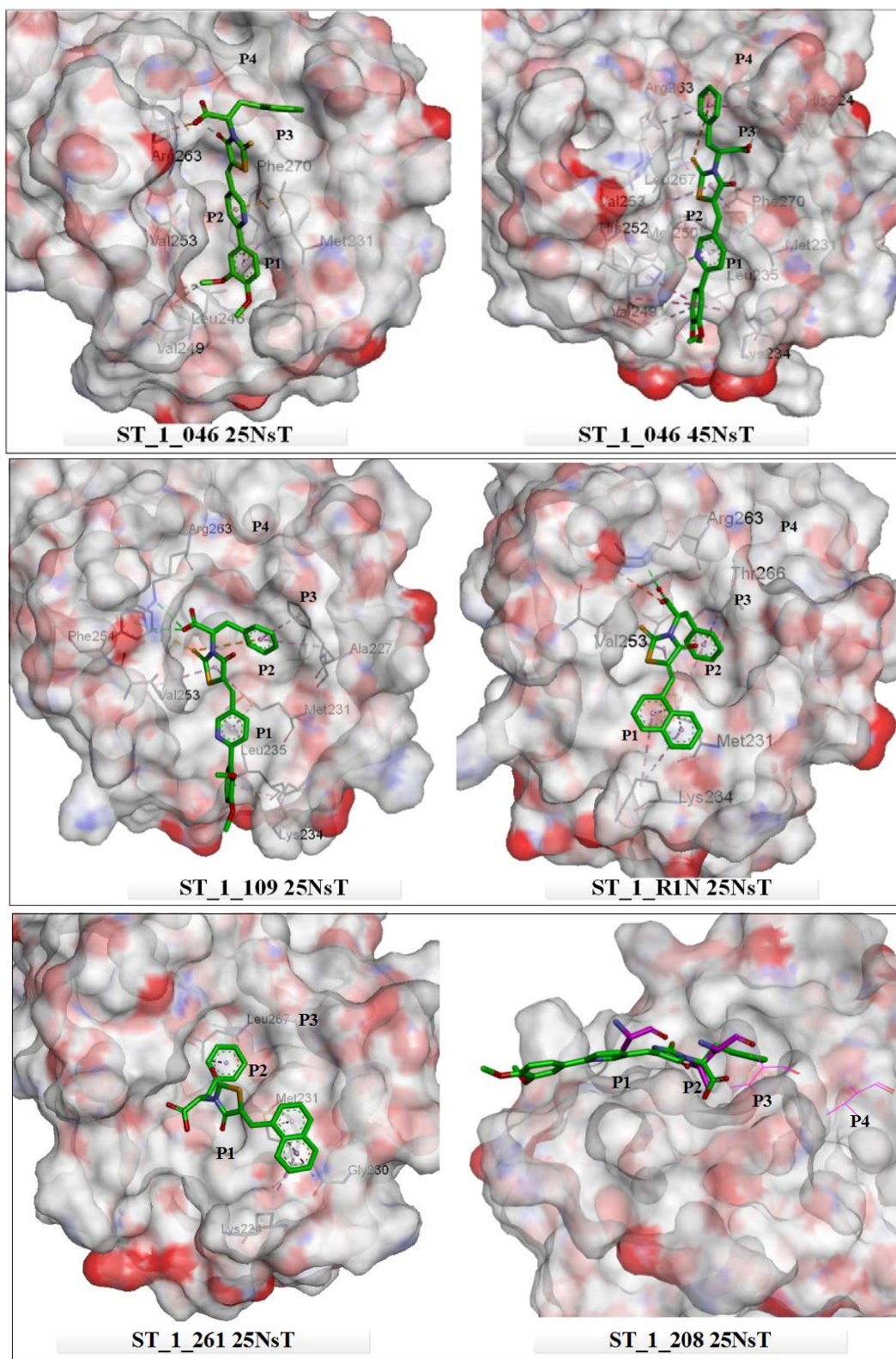


Figure8.3 Orientation of ST_1_046, ST_1_109, ST_1_R1N, ST_1_261, ST_1_208 25NsT and 45NsT belongs to 25ns and 45ns trajectory

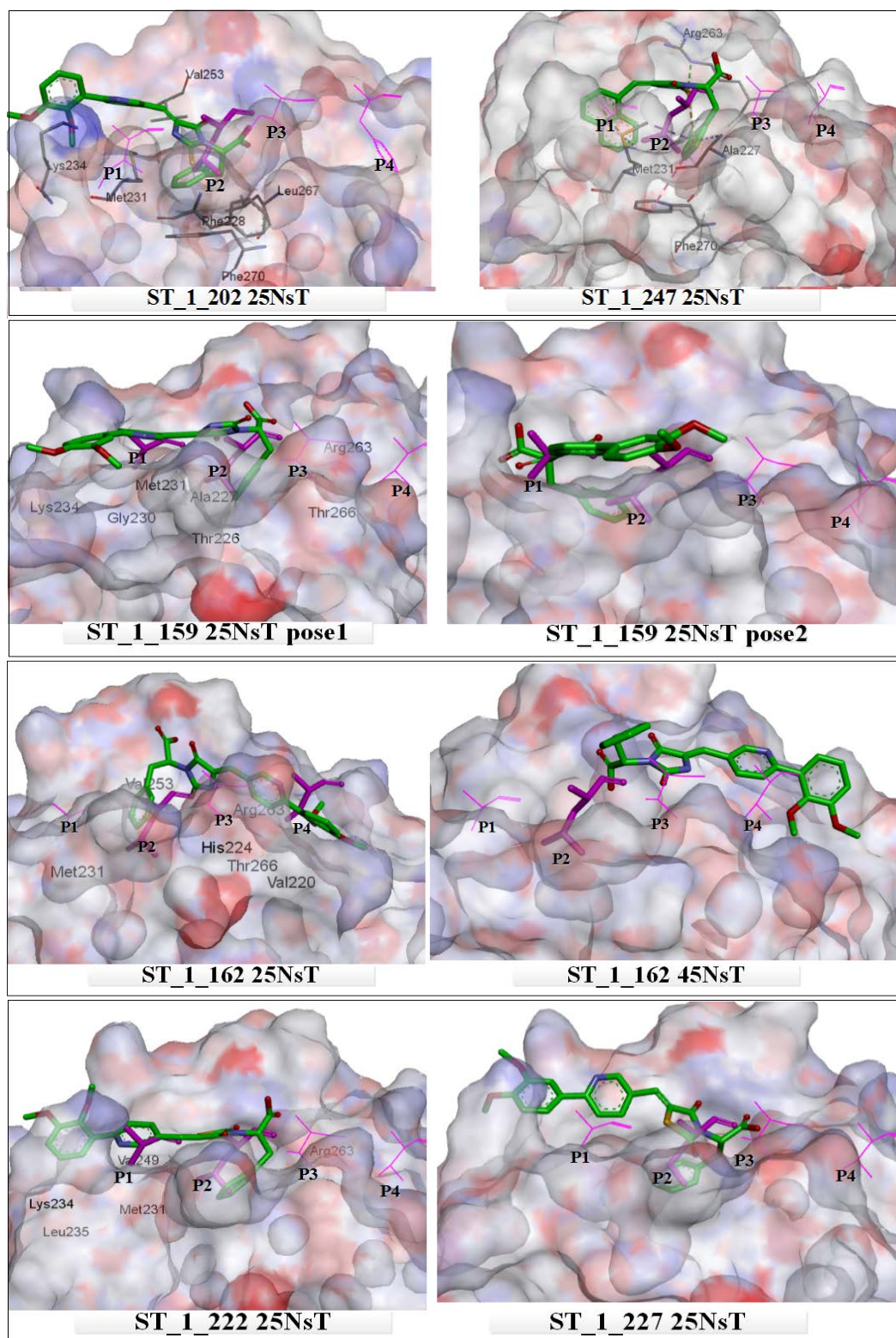


Figure 8.4 Orientation of ST_1_202, ST_1_227, ST_1_159, ST_1_162, ST_1_222 and ST_1_227 Pose1 and pose 2 belongs to cluster 1 and cluster 2, 25NsT and 45NsT belongs to 25ns and 45ns trajectory

pocket and forms π -alkyl interactions with Met231 and Lys234 (**Figure8.3**). The orientation of naphthalene ring causes rhodanine ring to move out of P2 pocket resulting in loss of interaction between exocyclic sulphur and phenyl ring of Phe254 in P2 pocket.

8.3.3.4 *ST_1_208*

The pose predicted as a result of clustering indicates that the ligand does not align along the MCL-1 binding groove but only interacts with the P2 pocket with N3 benzyl substituent (**Figure8.3**). However, longer simulation 45NsT indicates that *ST_1_208* flips and N3 benzyl ring moves out of the pocket.

8.3.3.5 *ST_1_247*

Similar pose was predicted by clustering of 25NsT and 45NsT. As can be seen in that thiohydantoin ring and N3 benzyl substituent occupies P2 pocket (**Figure8.4**). In addition, naphthalene ring stays close to P1. The trajectory analysis of 45NsT reveals that *ST_1_247* stays close to P2 pocket for complete 45ns utilizing its more non-polar groups.

8.3.3.6 *ST_1_202*

As can be seen from **Figure8.4** the thiohydantoin ring and N3 benzyl substituent occupies P2 pocket. However, the ligand was not found to be aligned with hydrophobic groove. Similar observation was found for longer 45NsT.

8.3.3.7 *ST_1_159*

The size of cluster_1 and cluster_2 was found to be almost similar in case of *ST_1_159*. In both poses, i.e. predicted by cluster_1 and cluster_2 (**Figure8.4**), its N3 benzyl substituent interacts with the pocket P2.

8.3.3.8 *ST_1_249*

The difference in the number of poses predicted for cluster_1 and cluster_2 of *ST_1_249* is not large. Moreover, superimposition of the pose from 45_NsT showed a third possible orientation for ligand. Analysis of three orientations revealed that, similar to pose predicted by cluster_1, *ST_1_249* utilizes its benzyl substituent to interact with the P2 pocket (**Figure8.5**). On inspection of 45NsT trajectory it was observed that after 20ns of simulation the compound starts flipping and after 40ns the carboxylic group faces P2 pocket.

8.3.3.9 *ST_1_162*

The predicted orientation indicates that *ST_1_162* interacts with the P2 pocket with its N3 benzyl substituent (**Figure8.4**). However, the pose predicted by 45NsT was found to be outside pocket P2. On visualization of the 45NsT trajectory, it was observed that after initial stay in the pocket *ST_1_162* moves out and is not able to interact with P2 pocket again.

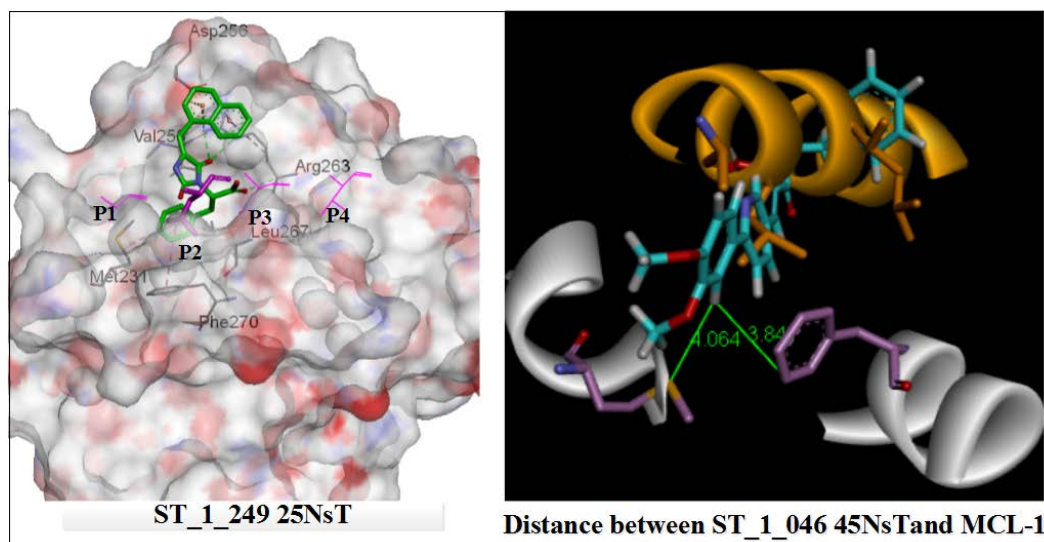


Figure8.5 Orientation of ST_1_249 and the distance between the pocket residues and closet atom of the pose ST_1_046 in 45NsT

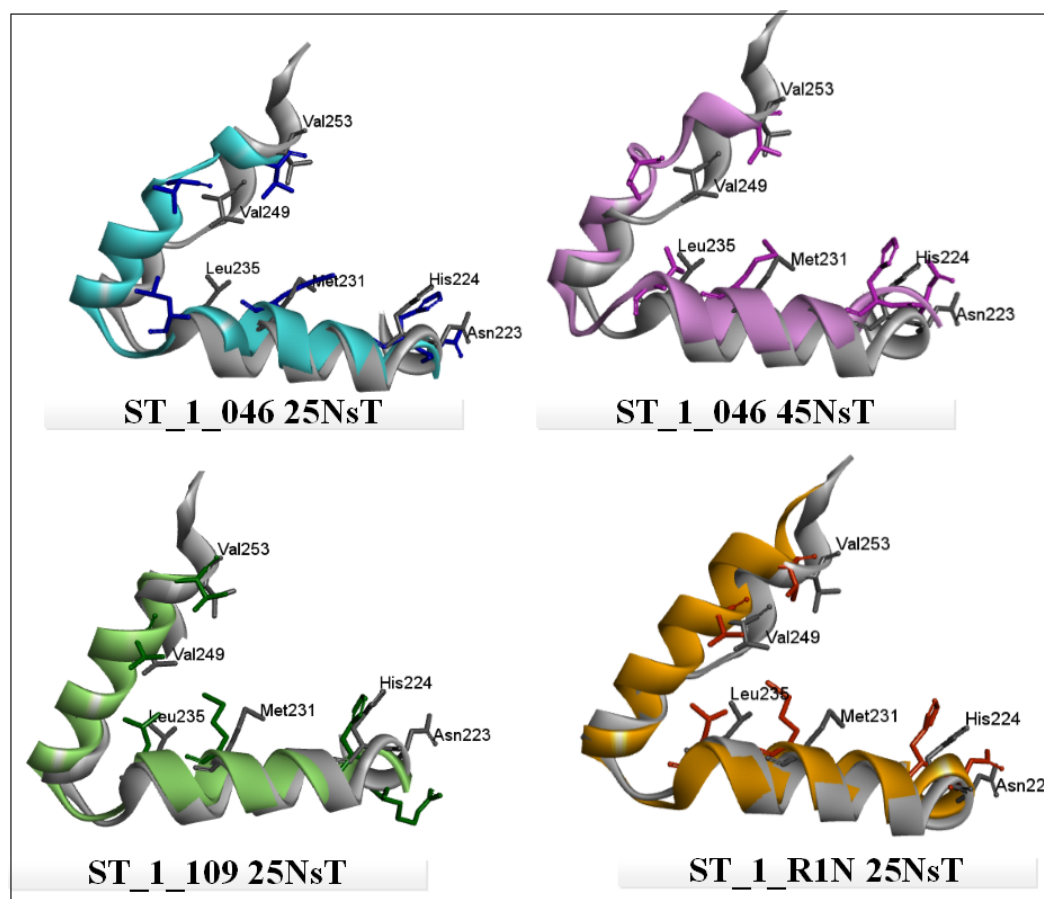


Figure8.6 Comparison of the residues of α_3 and α_4 and loop α_2 - α_4 loop for ST_1_046 25NsT, ST_1_046 45NsT, ST_1_109 25NsT and ST_1_R1N 25NsT

8.3.3.10 *ST_1_227 and ST_1_222*

The pose predicted by clustering of 25NsT and 45NsT for both *ST_1_227* and *ST_1_222* (**Figure8.4**) shows that its N3 benzyl substituent interacts with the pocket P2 while its biaryl ring attempts to interact with the P1 pocket. The 45NsT simulation of *ST_1_227* and *ST_1_222* reveals that thiazolidinedione ring is not able to enter P2 pocket and subsequently flips out of the pocket.

8.3.3.11 *ST_1_261*

The similar cluster size of best 3 clusters reveals that *ST_1_261* was not able to form stable interactions. The pose predicted by cluster_1 (**Figure8.3**) indicates that N3 benzyl substituent occupies the P2 pocket of MCL-1.

8.3.4 Conformation of the residues

The conformations of residues were analyzed to understand the interactions and flexibility of $\alpha 3$ and $\alpha 4$ loop (Figure8.6). The conformations of receptor bound to active compounds i.e. *ST_1_046*, *ST_1_R1N*, *ST_1_109* were analyzed. The residues at the $\alpha 2$ - $\alpha 3$ and $\alpha 3$ - $\alpha 4$ loop were found in different conformations. Among them, residues such as His224 and Asn223 seem to influence the P3 pocket. For *ST_1_109*, His224 shifts inside the pocket by 3.11Å thus shrinking the P3 pocket while not a significant change was observed for *ST_1_R1N*. The position of His224 was observed to be shifted slightly in opposite direction for 25NsT of *ST_1_046*. However, after a longer simulation,

the position of His224 was found close to the one observed in case of ST_1_109 and ST_1_R1N.

The position of Leu235 which is the last residue of $\alpha 3$ loop was found to be shifted by 3.9 Å for shorter ST_1_046 simulation. Val249 was found to shift outwards for all the poses except ST_1_R1N where it was found to move inside the cavity. Met231 was observed to possess three conformations i.e. facing towards $\alpha 2$ - $\alpha 3$ loop, towards the P2 pocket and facing towards $\alpha 3$ - $\alpha 4$ loop i.e. P1 pocket. For both ST_1_109 and ST_1_R1N, Met231 was facing $\alpha 3$ - $\alpha 4$ loop. For ST_1_R1N, it initially faces the $\alpha 2$ - $\alpha 3$ loop but then shifts towards P2 pocket.

8.3.5 Comparison between different scaffolds

8.3.5.1 Rhodanine

The compounds with rhodanine ring, i.e. ST_1_046, ST_1_109 and ST_1_R1N, possess both endo and exocyclic sulphur. All three compounds form stable orientation by going deep into P2 pocket as well as by interacting with residues of other pockets. This can be correlated with better activity in comparison to compounds with different scaffold. The difference in ST_1_046 and ST_1_109 lies in the position of methoxy group, i.e. ST_1_046 possess 3,4 dimethoxy in comparison to 2,3 dimethoxy present in ST_1_109. As shown by poses predicted for ST_1_046, the methoxy group at meta position (meta-p) interacts with Val249 residue comprising P1 pocket, and either Leu246 or Lys234. Moreover, it was observed that the distance between the residues in proximity of the hydrogen atom present at metaposition is close to 4Å

(**Figure8.5**) which is not big enough for a bigger methoxy group attached to the ortho position. This results in ST_1_109 occupying P1 and P2 pocket while ST_1_046 occupies P1, P2 and P3 pockets. Similarly, ST_1_R1N due to large naphthalene ring was not able to align with the binding groove.

The presence of endo and exocyclic sulphur along with the exocyclic oxygen in ST_1_046 goes well with the electrostatics of the MCL-1 binding pocket. The binding groove of the MCL-1 is hydrophobic in nature which was utilized by the rhodanine ring while the electropositive surface interacted with the methoxy and carboxylic group flanking the rhodanine ring.

With the increase in polar character of the central ring, interactions got weaker. Out of the endo and exocyclic sulphur of rhodanine ring, exocyclic sulphur had major contribution as can be observed from the difference in activity of compounds with thiohydantoin and thiazolidinedione ring.

8.3.5.2 *Thiohydantoin*

The thiohydantoin ring, possess exocyclic sulphur while endocyclic sulphur is replaced by amino group. Exocyclic sulphur was able to interact with the residue occupying hydrophobic groove. However, due to increase in polar character, central ring was not able to align with the hydrophobic groove. This resulted in loss of key interactions leading to poor activity of ST_202, ST_208 and ST_1_247.

8.3.5.3 *Hydantoin*

The hydantoin ring, possess endocyclic sulphur while exocyclic sulphur is replaced by oxygen. Due to presence of polar groups like endocyclic oxygen and carboxylic group in proximity, a strong interaction with the binding grove was not established. Hence, ST_222, ST_227 and ST_1_261 were not able to form stable interactions.

8.3.5.4 *Thiazolidinedione*

Thiazolidinedione ring lacks both endocyclic and exocyclic sulphur. The presence of polar groups hindered the alignment with hydrophobic grove of MCL-1. Thus, ST_159, ST_162 and ST_1_249 were not able to stay in hydrophobic grove for long duration resulting in least activity of compounds. In all four classes of compounds, presence of larger naphthalene ring did not improve the activity. The concluding remarks of this work have been described in Chapter 9.

Chapter 9

Conclusions

The last chapter summarizes the contributions (9.1) of this work towards the development of inhibitors against neuraminidase and MCL-1. Limitations and future work are described in 9.2 and Future work sections.

9.1 Contributions

In Chapter 2, 4 and 7, applications of computational methods like QSAR, docking and molecular dynamics was shown to develop drugs against diseases like influenza and cancer, i.e against neuraminidase and MCL-1.

The data utilized to train the QSAR models was collected from a number of publications on different strains of neuraminidase and with different mutations. This made dataset quite diverse in terms of scaffold. In contrast, most of the previous studies have used a limited a number of compounds effective against a particular strain. Dataset was processed and filtered thereby removing unwanted components. AD is a much desired property for any QSAR model to stop the prediction of the compounds not belonging to chemical space used to train the model. This ensured that prediction by the models in our work was accurate with low FPR and high accuracy. The base models had low FP and low false negative rate with Matthew's correlation coefficient, denoting accuracy of the model,

greater than 95%. The consensus model, combining the strengths of top 10 base models, increased the accuracy of the model to 98% with low FPR of 0.8% and low FDR of 6.3%. The QSAR model developed in our work to screen PNI had better prediction accuracy than most of the previous models developed. This shows greater impact of our model in screening PNI. Moreover, it highlights the importance of diverse and clean dataset for building the model. This along with use of consensus modeling and AD can help to improve the prediction of the model considerably. It was observed that in comparison to overall accuracy, MCC dropped for validation set. Thus in future work, with such imbalanced dataset, MCC should be used as main criteria for evaluation. In addition other modeling methods can also be implemented with diverse parameters can be used to build the base models. This will help to achieve the full potential of consensus model.

As the aim of this work is to discover pan neuraminidase inhibitors, extensive docking study was performed to reduce the compounds obtained by screening of ZINC library to most potent compounds. None of the docking studies to my knowledge have used both neuraminidase group I and group II along with multiple strains for the discovery of neuraminidase inhibitors. The docking performed against group I and group II neuraminidase in open and closed form helped to cover two most important conformations of the neuraminidase. Various mutations have been reported in neuraminidase making it resistance against oseltamivir and zanamivir. Among the mutations reported, some have been found by reverse genetics while others are found in clinical samples. The viability of the viruses with mutations discovered by reverse genetics is yet not established.

Hence, the mutations established in clinical samples i.e. H274Y, N294S and R292K were selected for docking studies. These mutations also belong to different groups of neuraminidase with H274Y and N294S belonging to group I neuraminidase and R292K to group II neuraminidase. In addition, these mutations also cover resistance to all three recommended drugs against influenza i.e. oseltamivir, zanamivir and laninamivir. This ensured that the predicted inhibitors can act as pan neuraminidase inhibitors. Though this study uses three most important mutations, the probable inhibitors can be tested against other resistant strains to achieve real pan neuraminidase inhibitors.

It was found that all top 5 compounds have similar orientation of the central ring, carboxylic, guanidino and acetamide group as compared to the existing inhibitors i.e. oseltamivir, zanamivir and laninamivir. The carboxylic group interacted with Arg118, Arg292, Tyr347 and Arg371 while guanidino group interacted with Glu119, Asp151, Trp178 and Glu227 and acetamide group interacted with Arg152. The side chain in compounds ZN88, ZN33 and ZN35 was bifurcated at 1st carbon providing it extra flexibility to negate the constriction in hydrophobic cavity. This flexibility allowed the side chain to be pushed backwards or rotate in H274Y and R292K mutation respectively. The shorter side chain was found to be facing the base of the cavity while the longer side chain was found close to the face of the cavity. Comparison of ZN33 and ZN35 indicate that maximum length of the side chain facing the base of cavity can be equal to propyl group. The longer side chain was found to be facing away from Glu276 and closer to Ile222, Arg224, Ala246 (N1)/Ser246 (N9). Except N1_N294S,

ZN88 formed hydrogen bond between the amino group of side chain and Glu276, Glu277 and Asp151 in all systems. However, slight movement in Asp151 can lead to formation of hydrogen bond in N1_N294S. The guanidino group provided extra stability by binding to residues comprising 150 loop. Thus, it can be concluded that inhibitors with guanidino group, flexible side chain and amino group in side chain can act as pan neuraminidase inhibitors. The low SD was observed for of ZN43, ZN88, ZN35 and ZN46 along with high average binding free energy. These compounds can be used to develop pharmacophore which later can used to develop potent pan neuraminidase inhibitors. Though this work covered most important conformations and resistant strains but these compounds can be tested against other resistant strains to check their validity as pan neuraminidase inhibitors.

It is a long desired dream to develop drugs to cure cancer. Many drugs have been developed that target anti-apoptotic proteins like BCL-XL. However, these drugs fail due to overexpression of MCL-1 and the drugs ineffectiveness against MCL-1. The second phase of our work concentrated on deducing the cause behind the activity of drugs against MCL-1. Poses clustered from the multiple trajectories indicate that the P2 pocket is important for interaction. Though inhibitors against MCL-1 have been designed but most of them are not selective to MCL-1. Moreover, they compete with BH3-only peptides by binding to one or two pockets. Thus the binding cavity is not utilized fully. In our work, it was found that compounds spanning through majority of binding grove increases its activity.

The analysis of the orientation of different scaffolds reveals that non-polar nature of the rhodanine facilitates the occupancy of the hydrophobic groove. As the polar character of the ring increases, the interactions get weaker leading to poor activity. This caused the drop in activity of compounds with other scaffolds i.e. thiohydantoin, hydantoin and thiazolidinedione. ST_1_046 was found to span the hydrophobic groove thus occupying P1, P2 and P3 pockets. The alignment with the groove was assisted by the polar groups flanking the non-polar rhodanine ring. Hence, the electrostatics of the ligand can determine the interaction with the binding groove. It was also observed that as the ST_1_046 goes deeper into the hydrophobic groove resulting in better activity. However, not every pocket can accommodate larger ligands and it can be observed from the fact that larger naphthalene of ST_1_R1N caused a drop in activity.

The analysis of residues reveals that $\alpha 3$ and $\alpha 4$ loop play important role in opening and closing of the pocket. The movement of $\alpha 3$ and $\alpha 4$ is due to greater flexibility of the loops linking them to adjacent helices. Val249 and Met231 seem to be crucial for the interaction and plasticity of the P2 pocket. Met231 by adopting different conformation can interact with inhibitors in both P1 and P2 pocket. These findings can be expanded to design more potent and selective MCL-1 inhibitors.

Five compounds had been synthesized by our collaborators, taking the computational results into consideration. Four compounds had a longer backbone with extension at N3 benzyl. In fifth compound i.e. ST_1_345, 3,4 dimethoxy is replaced by 3,4 diol. All five compounds had activity less than 20 μM while the

best compound had activity of about 13 μM . The compounds without extended structure had activity of about 15 μM . The extension of the structure did not improve activity. However, the central group can be made bulkier to occupy P2 pocket to greater depth. This can further help to improve the activity.

9.2 Limitations

The limitation of any QSAR model is its dependence on the dataset. Though the model was trained on most elaborate dataset but the dataset could not be referred to as representative dataset of all the compounds. Inadequate representation of the compounds could affect the prediction accuracy.

Docking was used to discover pan neuraminidase inhibitors. However, docking has limited accuracy in ranking the poses of the compounds. Energy minimization and pose rescoring was performed to overcome it but the error cannot be completely eliminated. There can be a chance of some important compound being overlooked in process of filtration.

The docking against broad hydrophobic groove at surface of MCL-1 was quite challenging. Docking was not able to produce sufficient results to deduce the poses for inhibitors in order to establish a trend. Molecular dynamics was used to overcome those limitations and binding free energy was calculated. Though molecular dynamics was able to explain the results on basis of poses but binding free energy was not able to correlate with experimental binding activity.

9.3 Future work

With time, the dataset used to train our neuraminidase QSAR model will no longer be sufficient. Thus, the QSAR model should be trained with new dataset for a better prediction. The QSAR model to predict inhibitors against influenza should be trained again with updated dataset including new inhibitors. The models can be tested using Akaike information criterion (AIC) which seems to be less prone to overfitting.

The neuraminidase inhibitors predicted by our work could be analyzed by molecular dynamics study against important mutant strains. This will help to get more insights about the mutations and predict novel neuraminidase inhibitors. The compounds can be further verified by *in-vivo* and *in-vitro* studies to develop novel potent pan neuraminidase inhibitors. This model can be extended to other viral diseases facing resistance against existing inhibitors. In our present work non-inhibitors were only used to build QSAR models and were not used in docking study. However, in future these compounds can be checked for possible role as neuraminidase inhibitor. Two probable inhibitors predicted in our study ZN33 and ZN35 come close to I222 and S246. Recently, it has been shown that both of these residues play role in resistance. Hence, in future more detailed study of can be performed involving mutations caused by I222 and S246. Compounds shown to be less effected by the mutant strains i.e. ZN43, ZN88, ZN35 and ZN46 will be used to develop pharmacophore model. This model will be used to develop novel and potent pan neuraminidase inhibitors. Recent discovered strains of neuraminidase i.e. N10 and N11 will be included for the screening and testing.

The effect of the compounds on the human sialidase will also be tested in the to check for any possible side effects.

The results of MCL-1 indicate that inhibitors occupying a large surface of the hydrophobic groove can increase potency. However, occupancy of only three pockets, i.e. P1, P2 and P3, can lead to a potent inhibitor. The application the electrostatic properties of the ligand along with the selective nature of the pockets can be utilized to develop novel, potent and selective inhibitors against MCL-1. Inhibitors with bulkier non-polar central scaffold with electronegative ends can be developed. This approach can be extended to other members of BCL-2 protein family to develop selective inhibitors. Docking production for surface groove is poor. Hence, the algorithms should be trained to solve such hurdles. Also there is a need to establish a relation between number and duration of trajectories required for a molecular dynamics experiment, in order to attain a good correlation between predicted binding energy and experimental activity. Molecular dynamics will be performed to study the minimum duration and number of trajectories required to explain biological activity.

Bibliography

- Abagyan, R., Totrov, M., & Kuznetsov, D. (1994). ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *Journal of computational chemistry*, *15*(5), 488-506.
- Abdellatif, K. R. A., Belal, A., & Omar, H. A. (2013). Design, synthesis and biological evaluation of novel triaryl (Z)-olefins as tamoxifen analogues. *Bioorganic & medicinal chemistry letters*, *23*(17), 4960-4963.
- Abed, Y., Baz, M., & Boivin, G. (2006). Impact of neuraminidase mutations conferring influenza resistance to neuraminidase inhibitors in the N1 and N2 genetic backgrounds. *Antivir Ther*, *11*(8), 971-976.
- Abed, Y., Nehme, B., Baz, M., & Boivin, G. (2008). Activity of the neuraminidase inhibitor A-315675 against oseltamivir-resistant influenza neuraminidases of N1 and N2 subtypes. *Antiviral research*, *77*(2), 163-166.
- Adler, M., & Beroza, P. (2013). Improved Ligand Binding Energies Derived from Molecular Dynamics: Replicate Sampling Enhances the Search of Conformational Space. *Journal of chemical information and modeling*, *53*(8), 2065-2072.
- Air, G. M. (2012). Influenza neuraminidase. *Influenza and Other Respiratory Viruses*, *6*(4), 245-256.
- Air, G. M., & Laver, W. G. (1989). The neuraminidase of influenza virus. *Proteins: Structure, Function, and Bioinformatics*, *6*(4), 341-356.
- Akgul, C. (2009). Mcl-1 is a potential therapeutic target in multiple types of cancer. *Cellular and Molecular Life Sciences*, *66*(8), 1326-1336.
- Allen, C. E., Chow, C. L., Caldwell, J. J., Westwood, I. M., M. van Montfort, R. L., & Collins, I. (2013). Synthesis and evaluation of heteroaryl substituted diazspiropyrans as scaffolds to probe the ATP-binding site of protein kinases. *Bioorganic & medicinal chemistry*, *21*(18), 5707-5724.
- Allison, M. (2012). Reinventing clinical trials. *Nat Biotech*, *30*(1), 41-49.
- Amaro, R. E., Minh, D. D. L., Cheng, L. S., Lindstrom Jr, W. M., Olson, A. J., Lin, J. H., . . . McCammon, J. A. (2007). Remarkable loop flexibility in avian influenza N1 and its implications for antiviral drug design. *Journal of the American Chemical Society*, *129*(25), 7764-7765.
- Andrade, C. H., Pasqualoto, K. F., Ferreira, E. I., & Hopfinger, A. J. (2010). 4D-QSAR: perspectives in drug design. *Molecules*, *15*(5), 3281-3294.
- Anthony, C. S., Masuyer, G., Sturrock, E. D., & Acharya, K. R. (2012). Structure based drug design of angiotensin-I converting enzyme inhibitors. *Curr Med Chem*, *19*(6), 845-855.
- Apisarnthanarak, A., Kitphati, R., Thongphubeth, K., Patoomanunt, P., Anthanont, P., Auwanit, W., . . . Fraser, V. J. (2004). Atypical avian influenza (H5N1). *Emerg Infect Dis*, *10*(7), 1321-1324.
- Arakawa, M., Hasegawa, K., & Funatsu, K. (2007). The Recent Trend in QSAR Modeling - Variable Selection and 3D-QSAR Methods. *Current Computer - Aided Drug Design*, *3*(4), 254-262.

- Ari, H. (1992). Unpacking the incoming influenza virus. *Cell*, 69(4), 577-578.
- Bajorath, J. (2002). Integration of virtual and high-throughput screening. *Nature reviews. Drug discovery*, 1(11), 882-894.
- Bakheet, T. M., & Doig, A. J. (2009). Properties and identification of human protein drug targets. *Bioinformatics*, 25(4), 451-457.
- Beck, D. A. C., & Daggett, V. (2004). Methods for molecular dynamics simulations of protein folding/unfolding in solution. *Methods*, 34(1), 112-120.
- Bekkar, M., Djemaa, H. K., & Alitouche, T. A. (2013). Evaluation Measures for Models Assessment over Imbalanced Data Sets *Journal of Information Engineering and Applications*, 3(10).
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., . . . Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235-242.
- Bernardo, P. H., Sivaraman, T., Wan, K.-F., Xu, J., Krishnamoorthy, J., Song, C. M., . . . Chai, C. L. L. (2010). Structural Insights into the Design of Small Molecule Inhibitors That Selectively Antagonize Mcl-1. *Journal of medicinal chemistry*, 53(5), 2314-2318.
- Bian, X., Dong, W., Zhao, Y., Sun, R., Kong, W., & Li, Y. (2014). Definition of the binding mode of phosphoinositide 3-kinase alpha-selective inhibitor A-66S through molecular dynamics simulation. *Journal of molecular modeling*, 20(4), 2166.
- Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., . . . Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*.
- Birchall, K., Gillet, V. J., Harper, G., & Pickett, S. D. (2008). Evolving Interpretable Structure–Activity Relationship Models. 2. Using Multiobjective Optimization To Derive Multiple Models. *Journal of chemical information and modeling*, 48(8), 1558-1570.
- Broach, J. R., & Thorner, J. (1996). High-throughput screening for drug discovery. *Nature*, 384(6604 Suppl), 14-16.
- Brotin, E., Meryet-Figuiera, M., Simonin, K., Duval, R. E., Villedieu, M., Leroy-Dudal, J., . . . Poulain, L. (2010). Bcl-XL and MCL-1 constitute pertinent targets in ovarian carcinoma and their concomitant inhibition is sufficient to induce apoptosis. *Int J Cancer*, 126(4), 885-895.
- Can005. (2011). A simplified depiction of how neuraminidase enzyme and its inhibitor functions; http://commons.wikimedia.org/wiki/File:Neuraminidasemechanism_with_with_out.jpg. Wikimedia.
- Cavarretta, I. T., Neuwirt, H., Untergasser, G., Moser, P. L., Zaki, M. H., Steiner, H., . . . Culig, Z. (2006). The antiapoptotic effect of IL-6 autocrine loop in a cellular model of advanced prostate cancer is mediated by Mcl-1. *Oncogene*, 26(20), 2822-2832.
- Cavasotto, C. N., & Phatak, S. S. (2009). Homology modeling in drug discovery: current trends and applications. *Drug Discov Today*, 14(13-14), 676-683.
- Chachra, R., & Rizzo, R. C. (2008). Origins of resistance conferred by the R292K neuraminidase mutation via molecular dynamics and free energy calculations. *Journal of Chemical Theory and Computation*, 4(9), 1526-1540.
- Chai, H. H., Lim, D., Lee, S. H., Chai, H. Y., & Jung, E. (2014). Homology modeling study of bovine mu-calpain inhibitor-binding domains. *International journal of molecular sciences*, 15(5), 7897-7938.

- Chaturvedi, P. R., Decker, C. J., & Odinecs, A. (2001). Prediction of pharmacokinetic properties using experimental approaches during early drug discovery. *Curr Opin Chem Biol*, 5(4), 452-463.
- ChemDraw. (2011). CambridgeSoft Desktop Software <http://www.cambridgesoft.com/> PerkinElmer Informatics.
- Chen, J., Wang, J., & Zhu, W. (2014). Binding Modes of Three Inhibitors 8CA, F8A and I4A to A-FABP Studied Based on Molecular Dynamics Simulation. *PLoS one*, 9(6), e99862.
- Chen, V. C. P. (2010). *Data mining: Special issue in annals of information systems* (Vol. 8): Springer.
- Chen, Y. Z., & Zhi, D. G. (2001). Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins*, 43(2), 217-226.
- Cheng, K. C., Korfmacher, W. A., White, R. E., & Njoroge, F. G. (2008). Lead Optimization in Discovery Drug Metabolism and Pharmacokinetics/Case study: The Hepatitis C Virus (HCV) Protease Inhibitor SCH 503034. *Perspect Medicin Chem*, 1, 1-9.
- Cheng, L. S., Amaro, R. E., Xu, D., Li, W. W., Arzberger, P. W., & McCammon, J. A. (2008). Ensemble-Based Virtual Screening Reveals Potential Novel Antiviral Compounds for Avian Influenza Neuraminidase. *Journal of medicinal chemistry*, 51(13), 3878-3894.
- Cheng, X., & Ivanov, I. (2012). Molecular Dynamics. In B. Reifeld & A. N. Mayeno (Eds.), *Computational Toxicology* (Vol. 929, pp. 243-285): Humana Press.
- Claveria, C., Giovinazzo, G., Sierra, R., & Torres, M. (2013). Myc-driven endogenous cell competition in the early mammalian embryo. *Nature*, 500(7460), 39-44.
- Collins, P. J., Haire, L. F., Lin, Y. P., Liu, J., Russell, R. J., Walker, P. A., . . . Hay, A. J. (2009). Structural basis for oseltamivir resistance of influenza viruses. *Vaccine*, 27(45), 6317-6323.
- Collins, P. J., Haire, L. F., Lin, Y. P., Liu, J., Russell, R. J., Walker, P. A., . . . Gamblin, S. J. (2008). Crystal structures of oseltamivir-resistant influenza virus neuraminidase mutants. *Nature*, 453(7199), 1258-1261.
- Colman, P. M. (1994). Influenza virus neuraminidase: Structure, antibodies, and inhibitors. *Protein Science*, 3(10), 1687-1696.
- Colman, P. M., Varghese, J. N., & Laver, W. G. (1983). Structure of the catalytic and antigenic sites in influenza virus neuraminidase. *Nature*, 303(5912), 41-44.
- Congreve, M., Dias, J. M., & Marshall, F. H. (2014). Structure-based drug design for G protein-coupled receptors. *Prog Med Chem*, 53, 1-63.
- Czabotar, P. E., Lee, E. F., van Delft, M. F., Day, C. L., Smith, B. J., Huang, D. C., . . . Colman, P. M. (2007). Structural insights into the degradation of Mcl-1 induced by BH3 domains. *Proceedings of the National Academy of Sciences of the United States of America*, 104(15), 6217-6222.
- Czabotar, P. E., Lessene, G., Strasser, A., & Adams, J. M. (2014). Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. *Nat Rev Mol Cell Biol*, 15(1), 49-63.
- D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Götz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui,

- V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman (2012), AMBER 12, University of California, San Francisco. (2012). AMBER 12. University of California, San Francisco.
- Dalkas, G. A., Vlachakis, D., Tsagkrasoulis, D., Kastania, A., & Kossida, S. (2013). State-of-the-art technology in modern computer-aided drug design. *Brief Bioinform*, 14(6), 745-752.
- Dao, T. T., Nguyen, P. H., Lee, H. S., Kim, E., Park, J., Lim, S. I., & Oh, W. K. (2011). Chalcones as novel influenza A (H1N1) neuraminidase inhibitors from *Glycyrrhiza inflata*. *Bioorganic & medicinal chemistry letters*, 21(1), 294-298.
- Dao, T. T., Tung, B. T., Nguyen, P. H., Thuong, P. T., Yoo, S. S., Kim, E. H., . . . Oh, W. K. (2010). C-methylated flavonoids from *cleistocalyx operculatus* and their inhibitory effects on novel influenza a (H1N1) neuraminidase. *Journal of natural products*, 73(10), 1636-1642.
- Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12), 10089-10092.
- Dave, K., & Lahiry, A. (2012). Conotoxins: review and docking studies to determine potentials of conotoxin as an anticancer drug molecule. *Curr Top Med Chem*, 12(8), 845-851.
- Day, C. L., Chen, L., Richardson, S. J., Harrison, P. J., Huang, D. C. S., & Hinds, M. G. (2005). Solution Structure of Prosurvival Mcl-1 and Characterization of Its Binding by Proapoptotic BH3-only Ligands. *Journal of Biological Chemistry*, 280(6), 4738-4744.
- Dearden, J. C., Cronin, M. T. D., & Kaiser, K. L. E. (2009). How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR). *SAR and QSAR in environmental research*, 20(3), 241 - 266.
- Debnath, A. K. (2005). Application of 3D-QSAR techniques in anti-HIV-1 drug design--an overview. *Curr Pharm Des*, 11(24), 3091-3110.
- Ding, Q., He, X., Xia, W., Hsu, J. M., Chen, C. T., Li, L. Y., . . . Hung, M. C. (2007). Myeloid cell leukemia-1 inversely correlates with glycogen synthase kinase-3beta activity and associates with poor prognosis in human breast cancer. *Cancer Res*, 67(10), 4564-4571.
- Dudek, A. Z., Arodz, T., & Gálvez, J. (2006). Computational methods in developing quantitative structure-activity relationships (QSAR): A review. *Combinatorial Chemistry and High Throughput Screening*, 9(3), 213-228.
- Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biol*, 9, 71.
- Dutta Gupta, S., Snigdha, D., Mazaira, G. I., Galigniana, M. D., Subrahmanyam, C. V., Gowrishankar, N. L., & Raghavendra, N. M. (2014). Molecular docking study, synthesis and biological evaluation of Schiff bases as Hsp90 inhibitors. *Biomed Pharmacother*, 68(3), 369-376.
- Dutta, S., Gullá, S., Chen, T. S., Fire, E., Grant, R. A., & Keating, A. E. (2010). Determinants of BH3 Binding Specificity for Mcl-1 versus Bcl-xL. *Journal of Molecular Biology*, 398(5), 747-762.
- Ekins, S., Mestres, J., & Testa, B. (2007). In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br J Pharmacol*, 152(1), 9-20.
- Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., & Mee, R. P. (1997). Empirical scoring functions: I. The development of a fast empirical scoring function to

- estimate the binding affinity of ligands in receptor complexes. *Journal of computer-aided molecular design*, 11(5), 425-445.
- Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., . . . Sali, A. (2007). Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci, Chapter 2*, Unit 2.9.
- Ewing, T. A., Makino, S., Skillman, A. G., & Kuntz, I. (2001). DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *Journal of computer-aided molecular design*, 15(5), 411-428.
- Favaloro, B., Allocati, N., Graziano, V., Di Ilio, C., & De Laurenzi, V. (2012). Role of apoptosis in disease. *Aging (Albany NY)*, 4(5), 330-349.
- Ferino, G., Vilar, S., Matos, M. J., Uriarte, E., & Cadoni, E. (2012). Monoamine oxidase inhibitors: ten years of docking studies. *Curr Top Med Chem*, 12(20), 2145-2162.
- Ferraris, O., & Lina, B. (2008). Mutations of neuraminidase implicated in neuraminidase inhibitors resistance. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*, 41(1), 13-19.
- Fire, E., Gullá, S. V., Grant, R. A., & Keating, A. E. (2010). Mcl-1–Bim complexes accommodate surprising point mutations via minor structural changes. *Protein Science*, 19(3), 507-519.
- Fodor, E., & Brownlee, G. G. (2002). Influenza virus replication. In C. W. Potter (Ed.), *Perspectives in Medical Virology* (Vol. Volume 7, pp. 1-29): Elsevier.
- Fourches, D., Muratov, E., & Tropsha, A. (2010). Trust, But Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research. *Journal of chemical information and modeling*, 50(7), 1189-1204.
- Fulda, S., & Debatin, K. M. (2006). Extrinsic versus intrinsic apoptosis pathways in anticancer chemotherapy. *Oncogene*, 25(34), 4798-4811.
- Gammerman, T. B. I. N. M. Y. A. (2014). Feature Selection. In V. B. S.-S. H. V. Vovk (Ed.), *Conformal Prediction for Reliable Machine Learning: Theory, Adaptations, and Applications*: Morgan Kaufmann.
- Garozzo, A., Timpanaro, R., Stivala, A., Bisignano, G., & Castro, A. (2011). Activity of Melaleuca alternifolia (tea tree) oil on Influenza virus A/PR/8: Study on the mechanism of action. *Antiviral research*, 89(1), 83-88.
- Ghosh, S., Nie, A., An, J., & Huang, Z. (2006). Structure-based virtual screening of chemical libraries for drug discovery. *Current Opinion in Chemical Biology*, 10(3), 194-202.
- Giorgio Fumera, Fabio Roli, & Giacinto, G. (2000). Reject option with multiple thresholds. *Pattern Recognition Society*, 33, 2099-2101.
- Golbraikh, A., & Tropsha, A. (2002). Beware of q²! *Journal of Molecular Graphics and Modelling*, 20(4), 269-276.
- Golstein, P. (1998). Cell Death in Us and Others. *Science*, 281(5381), 1283.
- Goncharenko-Khaider, N., Matte, I., Lane, D., Rancourt, C., & Piche, A. (2012). Ovarian cancer ascites increase Mcl-1 expression in tumor cells through ERK1/2-Elk-1 signaling to attenuate TRAIL-induced apoptosis. *Mol Cancer*, 11, 84.
- Gong, J., Xu, W., & Zhang, J. (2007). Structure and functions of influenza virus neuraminidase. *Current Medicinal Chemistry*, 14(1), 113-122.
- Goodarzi, M., Dejaegher, B., & Vander Heyden, Y. (2012). Feature selection methods in QSAR studies. *J AOAC Int*, 95(3), 636-651.

- Gores, G. J., & Kaufmann, S. H. (2012). Selectively targeting Mcl-1 for the treatment of acute myelogenous leukemia and solid tumors. *Genes Dev*, 26(4), 305-311. doi: 10.1101/gad.186189.111
- Gramatica, P. (2007). Principles of QSAR models validation: internal and external. *QSAR & Combinatorial Science*, 26(5), 694-701.
- Gramatica, P. (2010). Chemometric Methods and Theoretical Molecular Descriptors in Predictive QSAR Modeling of the Environmental Behavior of Organic Pollutants. In T. Puzyn, J. Leszczynski & M. T. Cronin (Eds.), *Recent Advances in QSAR Studies* (Vol. 8, pp. 327-366): Springer Netherlands.
- Grover, J., Kumar, V., Singh, V., Bairwa, K., Sobhia, M. E., & Jachak, S. M. (2014). Synthesis, biological evaluation, molecular docking and theoretical evaluation of ADMET properties of nepodin and chrysophanol derivatives as potential cyclooxygenase (COX-1, COX-2) inhibitors. *European journal of medicinal chemistry*, 80(0), 47-56.
- Guimarães, C. R. W., & Cardozo, M. (2008). MM-GB/SA Rescoring of Docking Poses in Structure-Based Lead Optimization. *Journal of chemical information and modeling*, 48(5), 958-970.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157-1182.
- Guzel, O., Innocenti, A., Vullo, D., Scozzafava, A., & Supuran, C. T. (2010). 3-phenyl-1H-indole-5-sulfonamides: structure-based drug design of a promising class of carbonic anhydrase inhibitors. *Curr Pharm Des*, 16(29), 3317-3326.
- Häggström, M. (2014). Main symptoms of influenza; Medical gallery of Mikael Häggström ;Wikiversity Journal of Medicine 1 (2). DOI:10.15347/wjm/2014.008.
- Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of medicinal chemistry*, 47(7), 1750-1759.
- Hanahan, D., & Weinberg, R. A. (2000). The Hallmarks of Cancer. *Cell*, 100(1), 57-70.
- Hanahan, D., & Weinberg, Robert A. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, 144(5), 646-674.
- Hannigan, K., Kulkarni, S. S., Bdzhola, V. G., Golub, A. G., Yarmoluk, S. M., & Talele, T. T. (2013). Identification of novel PARP-1 inhibitors by structure-based virtual screening. *Bioorganic & medicinal chemistry letters*, 23(21), 5790-5794.
- Hardwick, J. M., & Soane, L. (2013). Multiple functions of BCL-2 family proteins. *Cold Spring Harb Perspect Biol*, 5(2).
- Hay, A. J., Gregory, V., Douglas, A. R., & Lin, Y. P. (2001). The evolution of human influenza viruses. *Philos Trans R Soc Lond B Biol Sci*, 356(1416), 1861-1870.
- Hertzberg, R. P., & Pope, A. J. (2000). High-throughput screening: new technology for the 21st century. *Current Opinion in Chemical Biology*, 4(4), 445-451.
- Honda, T., Masuda, T., Yoshida, S., Arai, M., Kaneko, S., & Yamashita, M. (2002). Synthesis and anti-Influenza virus activity of 7-O-Alkylated derivatives related to zanamivir. *Bioorganic & medicinal chemistry letters*, 12(15), 1925-1928.
- Honda, T., Masuda, T., Yoshida, S., Arai, M., Kobayashi, Y., & Yamashita, M. (2002). Synthesis and anti-influenza virus activity of 4-guanidino-7-substituted Neu5Ac2en derivatives. *Bioorganic & medicinal chemistry letters*, 12(15), 1921-1924.

- Hou, T., Wang, J., Li, Y., & Wang, W. (2011). Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *Journal of chemical information and modeling*, *51*(1), 69-82.
- Hou, T., Wang, J., Li, Y., & Wang, W. (2011). Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *Journal of computational chemistry*, *32*(5), 866-877.
- Hu, C. M., Zhu, J., Guo, X. E., Chen, W., Qiu, X. L., Ngo, B., . . . Lee, W. H. (2014). Novel small molecules disrupting Hec1/Nek2 interaction ablate tumor progression by triggering Nek2 degradation through a death-trap mechanism. *Oncogene*.
- Huang, H.-J., Yu, H. W., Chen, C.-Y., Hsu, C.-H., Chen, H.-Y., Lee, K.-J., . . . Chen, C. Y.-C. (2010). Current developments of computer-aided drug design. *Journal of the Taiwan Institute of Chemical Engineers*, *41*(6), 623-635.
- Huang, S. Y., Grinter, S. Z., & Zou, X. (2010). Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys Chem Chem Phys*, *12*(40), 12899-12908.
- Huang, S. Y., & Zou, X. (2010). Advances and challenges in protein-ligand docking. *International journal of molecular sciences*, *11*(8), 3016-3034.
- Hug, S. (2013). Classical Molecular Dynamics in a Nutshell. In L. Monticelli & E. Salonen (Eds.), *Biomolecular Simulations* (Vol. 924, pp. 127-152): Humana Press.
- Hughes, J. P., Rees, S., Kalindjian, S. B., & Philpott, K. L. (2011). Principles of early drug discovery. *Br J Pharmacol*, *162*(6), 1239-1249.
- Hurt, A. C., Lowther, S., Middleton, D., & Barr, I. G. (2010). Assessing the development of oseltamivir and zanamivir resistance in A(H5N1) influenza viruses using a ferret model. *Antiviral research*, *87*(3), 361-366.
- Hurt, A. C., Selleck, P., Komadina, N., Shaw, R., Brown, L., & Barr, I. G. (2007). Susceptibility of highly pathogenic A(H5N1) avian influenza viruses to the neuraminidase inhibitors and adamantanes. *Antiviral research*, *73*(3), 228-231.
- Irwin, J. J., & Shoichet, B. K. (2005). ZINC--a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling*, *45*(1), 177-182.
- Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., & Coleman, R. G. (2012). ZINC: A Free Tool to Discover Chemistry for Biology. *Journal of chemical information and modeling*, *52*(7), 1757-1768.
- Jaworska, J., Aldenberg, T., & Nikolova, N. (2004). Review of methods for assessing the applicability domains of QSARs and QSARS. *Sponsor: The European Commission—Joint Research Ctr., Institute for Health and Consumer Protection—ECVAM, Italy*.
- Jaworska, J., Nikolova-Jeliazkova, N., & Aldenberg, T. (2005). QSAR applicability domain estimation by projection of the training set in descriptor space: A review. *ATLA. Alternatives to laboratory animals*, *33*(5), 445-459.
- Jayanthi, S., Kang, S. W., Bingham, D., Tessaro, B. A., Suresh Kumar, T. K., & Kuenzel, W. J. (2014). Identification of antagonists to the vasotocin receptor sub-type 4 (VT4R) involved in stress by molecular modelling and verification using anterior pituitary cells. *J Biomol Struct Dyn*, *32*(4), 648-660.

- Jones, G., Willett, P., & Glen, R. C. (1995). Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of Molecular Biology*, 245(1), 43-53.
- Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, 267(3), 727-748.
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2), 926-935.
- Juin, P., Geneste, O., Gautier, F., Depil, S., & Campone, M. (2013). Decoding and unlocking the BCL-2 dependency of cancer cells. *Nat Rev Cancer*, 13(7), 455-465.
- Kalyanamoorthy, S., & Chen, Y.-P. P. (2011). Structure-based drug design to augment hit discovery. *Drug Discovery Today*, 16(17-18), 831-839.
- Kandil, S., Biondaro, S., Vlachakis, D., Cummins, A. C., Coluccia, A., Berry, C., . . . Brancale, A. (2009). Discovery of a novel HCV helicase inhibitor by a de novo drug design approach. *Bioorganic & medicinal chemistry letters*, 19(11), 2935-2937.
- Kapetanovic, I. M. (2008). Computer-aided drug discovery and development (CADD): in silico-chemico-biological approach. *Chem Biol Interact*, 171(2), 165-176.
- Karthick, V., & Ramanathan, K. (2014). Insight into the oseltamivir resistance R292K mutation in H5N1 influenza virus: a molecular docking and molecular dynamics approach. *Cell Biochem Biophys*, 68(2), 291-299.
- Kerr, J. F., Wyllie, A. H., & Currie, A. R. (1972). Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *British journal of cancer*, 26(4), 239-257.
- Kerrigan, J. E. (2013). Molecular dynamics simulations in drug design. *Methods Mol Biol*, 993, 95-113.
- Khalid, S., & Paul, S. (2014). Identifying a C-terminal ATP binding sites-based novel Hsp90-Inhibitor in silico: A plausible therapeutic approach in Alzheimer's disease. *Med Hypotheses*, 83(1), 39-46.
- Kim, C. U., Lew, W., Williams, M. A., Liu, H., Zhang, L., Swaminathan, S., . . . Stevens, R. C. (1997). Influenza Neuraminidase Inhibitors Possessing a Novel Hydrophobic Interaction in the Enzyme Active Site: Design, Synthesis, and Structural Analysis of Carbocyclic Sialic Acid Analogues with Potent Anti-Influenza Activity. *Journal of the American Chemical Society*, 119(4), 681-690.
- Kim, C. U., Lew, W., Williams, M. A., Wu, H., Zhang, L., Chen, X., . . . Stevens, R. C. (1998). Structure-Activity Relationship Studies of Novel Carbocyclic Influenza Neuraminidase Inhibitors. *Journal of medicinal chemistry*, 41(14), 2451-2460.
- Kim, J., Kim, B.-S., & Savarese, S. (2012). *Comparing image classification methods: K-nearest-neighbor and support-vector-machines*. Paper presented at the Proceedings of the 6th WSEAS international conference on Computer Engineering and Applications, and Proceedings of the 2012 American conference on Applied Mathematics, Harvard, Cambridge.
- Kimble, B., Nieto, G. R., & Perez, D. R. (2010). Characterization of influenza virus sialic acid receptors in minor poultry species. *Virology journal*, 7, 365.
- Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature reviews. Drug discovery*, 3(11), 935-949.

- Klopman, G., Stefan, L. R., & Saiakhov, R. D. (2002). ADME evaluation: 2. A computer model for the prediction of intestinal absorption in humans. *European Journal of Pharmaceutical Sciences*, 17(4–5), 253-263.
- Kopp, J., & Schwede, T. (2004). Automated protein structure homology modeling: a progress report. *Pharmacogenomics*, 5(4), 405-416. doi: 10.1517/14622416.5.4.405
- Korfmacher, W. A. (2003). Lead optimization strategies as part of a drug metabolism environment. *Curr Opin Drug Discov Devel*, 6(4), 481-485.
- Korošec, B., Sova, M., Turk, S., Kraševc, N., Novak, M., Lah, L., . . . Komel, R. (2014). Antifungal activity of cinnamic acid derivatives involves inhibition of benzoate 4-hydroxylase (CYP53). *Journal of Applied Microbiology*, 116(4), 955-966.
- Kozakov, D., Clodfelter, K. H., Vajda, S., & Camacho, C. J. (2005). Optimal Clustering for Detecting Near-Native Conformations in Protein Docking. *Biophysical Journal*, 89(2), 867-875.
- Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., . . . Karplus, K. (2009). Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins*, 77 Suppl 9, 114-122.
- Krieger, E., Nabuurs, S. B., & Vriend, G. (2005). Homology Modeling *Structural Bioinformatics* (pp. 509-523): John Wiley & Sons, Inc.
- Kubinyi, H. (1997). QSAR and 3D QSAR in drug design Part 1: methodology. *Drug Discovery Today*, 2(11), 457-467.
- Lamb, R. A., & King, R. M. (2001). *Orthomyxoviridae: the viruses and their replication*: Lippincott, Williams and Wilkins.
- Landon, M. R., Amaro, R. E., Baron, R., Ngan, C. H., Ozonoff, D., McCammon, J. A., & Vajda, S. (2008). Novel druggable hot spots in avian influenza neuraminidase H5N1 revealed by computational solvent mapping of a reduced and representative receptor ensemble. *Chemical biology & drug design*, 71(2), 106-116.
- Lange, J. H. M., Coolen, H. K. A. C., van der Neut, M. A. W., Borst, A. J. M., Stork, B., Verveer, P. C., & Kruse, C. G. (2010). Design, Synthesis, Biological Properties, and Molecular Modeling Investigations of Novel Tacrine Derivatives with a Combination of Acetylcholinesterase Inhibition and Cannabinoid CB1 Receptor Antagonism. *Journal of medicinal chemistry*, 53(3), 1338-1346.
- Lawrenz, M., Wereszczynski, J., Amaro, R., Walker, R., Roitberg, A., & McCammon, J. A. (2010). Impact of calcium on N1 influenza neuraminidase dynamics and binding free energy. *Proteins: Structure, Function, and Bioinformatics*, 78(11), 2523-2532.
- Le Gouill, S., Podar, K., Harousseau, J. L., & Anderson, K. C. (2004). Mcl-1 regulation and its role in multiple myeloma. *Cell Cycle*, 3(10), 1259-1262.
- Leari, R. (2007). Genetic algorithms in chemistry. *Journal of Chromatography A*, 1158(1–2), 226-233.
- Lee, E. F., Czabotar, P. E., Smith, B. J., Deshayes, K., Zobel, K., Colman, P. M., & Fairlie, W. D. (2007). Crystal structure of ABT-737 complexed with Bcl-xL: implications for selectivity of antagonists of the Bcl-2 family. *Cell Death Differ*, 14(9), 1711-1713.
- Lei Yu, & Liu, H. (2003). *Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution*. Paper presented at the Proceedings of the Twentieth International Conference on Machine Learning, Washington DC.

- Lessene, G. C., Peter E.; Sleebs, Brad E.; Zobel, Kerry; Lowes, Kym N.; Adams, Jerry M.; Baell, Jonathan B.; Colman, Peter M.; Deshayes, Kurt; Fairbrother, Wayne J.; Flygare, John A.; Gibbons, Paul; Kersten, Wilhelmus J. A.; Kulasegaram. (2013). Structure-guided design of a selective BCL-XL inhibitor. *Nat Chem Biol*, 9(6), 390-397.
- Lew, W., Wu, H., Chen, X., Graves, B. J., Escarpe, P. A., MacArthur, H. L., . . . Kim, C. U. (2000). Carbocyclic influenza neuraminidase inhibitors possessing a C3-cyclic amine side chain: synthesis and inhibitory activity. *Bioorganic & medicinal chemistry letters*, 10(11), 1257-1260.
- Lew, W., Wu, H., Mendel, D. B., Escarpe, P. A., Chen, X., Laver, W. G., . . . Kim, C. U. (1998). A new series of C3-aza carbocyclic influenza neuraminidase inhibitors: synthesis and inhibitory activity. *Bioorganic & medicinal chemistry letters*, 8(23), 3321-3324.
- Li, Z., Jia, L., Wang, J., Wu, X., Shi, G., Lu, C., & Shen, Y. (2014). Discovery of novel 17-phenylethylaminegeldanamycin derivatives as potent Hsp90 inhibitors. *Chemical biology & drug design*.
- Lindstrom, A., Edvinsson, L., Johansson, A., Andersson, C. D., Andersson, I. E., Raubacher, F., & Linusson, A. (2011). Postprocessing of docked protein-ligand complexes using implicit solvation models. *Journal of chemical information and modeling*, 51(2), 267-282.
- Liu, A. L., Wang, H. D., Lee, S. M., Wang, Y. T., & Du, G. H. (2008). Structure-activity relationship of flavonoids as influenza virus neuraminidase inhibitors and their in vitro anti-viral activities. *Bioorganic & medicinal chemistry*, 16(15), 7141-7147.
- Liu, Q., Liu, D.-y., & Yang, Z.-q. (2013). Characteristics of human infection with avian influenza viruses and development of new antiviral agents. *Acta Pharmacol Sin*, 34(10), 1257-1269.
- Liu, Y., Jing, F., Xu, Y., Xie, Y., Shi, F., Fang, H., . . . Xu, W. (2011). Design, synthesis and biological activity of thiazolidine-4-carboxylic acid derivatives as novel influenza neuraminidase inhibitors. *Bioorganic & Medicinal Chemistry*, 19(7), 2342-2348.
- Lyne, P. D. (2002). Structure-based virtual screening: an overview. *Drug Discovery Today*, 7(20), 1047-1055.
- Ma, Z., Liu, H., & Wu, B. (2014). Structure-based drug design of catechol-O-methyltransferase inhibitors for CNS disorders. *Br J Clin Pharmacol*, 77(3), 410-420.
- Mackay, I. M. Structure of an influenzavirus virion VDU's blog; www.virologydownunder.blogspot.com.au .
- Malaisree, M., Rungrotmongkol, T., Nunthaboot, N., Aruksakunwong, O., Intharathep, P., Decha, P., . . . Hannongbua, S. (2009). Source of oseltamivir resistance in avian influenza H5N1 virus with the H274Y mutation. *Amino Acids*, 37(4), 725-732.
- Marjuki, H., Mishin, V. P., Chesnokov, A. P., Jones, J., De La Cruz, J. A., Sleeman, K., . . . Gubareva, L. V. (2014). Characterization of Drug-Resistant Influenza A(H7N9) Variants Isolated From an Oseltamivir-Treated Patient in Taiwan. *J Infect Dis*.
- Marsoni, S., & Damia, G. (2004). Molecular targeting: new therapeutic strategies to improve tumour apoptosis. *Ann Oncol*, 15 Suppl 4, iv229-231.

- Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F., & Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*, 29, 291-325.
- Martis E A, R. R., Badve R R. (2011). High-Throughput Screening: The Hits and Leads of Drug Discovery- An Overview. *Journal of Applied Pharmaceutical Science*, 1(1).
- Masuda, T. (2003). Synthesis and anti-influenza evaluation of orally active bicyclic ether derivatives related to zanamivir. *Bioorganic & medicinal chemistry letters*, 13(4), 669-673.
- McGann, M. (2011). FRED Pose Prediction and Virtual Screening Accuracy. *Journal of chemical information and modeling*, 51(3), 578-596.
- Meier, P., Finch, A., & Evan, G. (2000). Apoptosis in development. *Nature*, 407(6805), 796-801.
- Merino, D., Khaw, S. L., Glaser, S. P., Anderson, D. J., Belmont, L. D., Wong, C., . . . Bouillet, P. (2012). Bcl-2, Bcl-x(L), and Bcl-w are not equivalent targets of ABT-737 and navitoclax (ABT-263) in lymphoid and leukemic cells. *Blood*, 119(24), 5807-5816.
- Michels, J., Johnson, P. W., & Packham, G. (2005). Mcl-1. *Int J Biochem Cell Biol*, 37(2), 267-271.
- Mihajlovic, M. L., & Mitrasinovic, P. M. (2008). Another look at the molecular mechanism of the resistance of H5N1 influenza A virus neuraminidase (NA) to oseltamivir (OTV). *Biophysical chemistry*, 136(2-3), 152-158.
- Mitchell, C., Yacoub, A., Hossein, H., Martin, A. P., Bareford, M. D., Eulitt, P., . . . Dent, P. (2010). Inhibition of MCL-1 in breast cancer cells promotes cell death in vitro and in vivo. *Cancer Biol Ther*, 10(9), 903-917.
- MOE. <http://www.chemcomp.com/software.htm>. Montreal, QC, Canada: Molecular Operating Environment, Chemical Computing Group.
- Mohan Sahoo, B., Chandra Dinda, S., Ravi Kumar, B. V. V., Panda, J., & S Brahmksatriya, P. (2014). Design, Green Synthesis, and Anti-Inflammatory Activity of Schiff Base of 1,3,4-oxadiazole Analogues. *Letters in Drug Design & Discovery*, 11(1), 82-89.
- Mohan, V., Gibbs, A. C., Cummings, M. D., Jaeger, E. P., & DesJarlais, R. L. (2005). Docking: Successes and Challenges. *Current Pharmaceutical Design*, 11(3), 323-333.
- Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of computational chemistry*, 19(14), 1639-1662.
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*, 30(16), 2785-2791.
- Moscona, A. (2008). Medical Management of Influenza Infection. *Annual Review of Medicine*, 59(1), 397-413.
- Muegge, I. (2002). A knowledge-based scoring function for protein-ligand interactions: Probing the reference state. In G. Klebe (Ed.), *Virtual Screening: An Alternative or Complement to High Throughput Screening?* (Vol. 20, pp. 99-114): Springer Netherlands.
- Naik, P., Santoshi, S., & Joshi, H. (2012). Noscapiroids with anti-cancer activity against human acute lymphoblastic leukemia cells (CEM): a three dimensional chemical

- space pharmacophore modeling and electronic feature analysis. *Journal of molecular modeling*, 18(1), 307-318.
- Nelson, M. I., & Holmes, E. C. (2007). The evolution of epidemic influenza. *Nat Rev Genet*, 8(3), 196-205.
- Nematollahi, A., Church, W. B., Nadvi, N. A., Gorrell, M. D., & Sun, G. (2014). Homology Modeling of Human Kynurenine Aminotransferase III and Observations on Inhibitor Binding Using Molecular Docking. *Cent Nerv Syst Agents Med Chem*.
- Netzeva, T. I., Worth, A. P., Aldenberg, T., Benigni, R., Cronin, M. T. D., Gramatica, P., . . . Marchant, C. A. (2005). Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. *Alternatives to Laboratory Animals*, 33(2), 1-19.
- Ni, S., Yuan, Y., Huang, J., Mao, X., Lv, M., Zhu, J., . . . Li, J. (2009). Discovering potent small molecule inhibitors of cyclophilin A using de novo drug design approach. *Journal of medicinal chemistry*, 52(17), 5295-5298.
- NIAID. (2011). Antigenic shift; . National Institute of Allergy and Infectious Diseases; <http://www.niaid.nih.gov/topics/flu/research/basic/pages/antigenicshiftillustration.aspx>.
- Nicholson, K. G. (1992). Clinical features of influenza. *Semin Respir Infect*, 7(1), 26-37.
- Nikolova, N., & Jaworska, J. (2003). Approaches to Measure Chemical Similarity – a Review. *QSAR & Combinatorial Science*, 22(9-10), 1006-1026.
- Oltersdorf, T., Elmore, S. W., Shoemaker, A. R., Armstrong, R. C., Augeri, D. J., Belli, B. A., . . . Rosenberg, S. H. (2005). An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature*, 435(7042), 677-681.
- Onufriev, A., Bashford, D., & Case, D. A. (2000). Modification of the Generalized Born Model Suitable for Macromolecules. *The Journal of Physical Chemistry B*, 104(15), 3712-3720.
- Onufriev, A., Bashford, D., & Case, D. A. (2004). Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics*, 55(2), 383-394.
- Ooms, F. (2000). Molecular modeling and computer aided drug design. Examples of their applications in medicinal chemistry. *Curr Med Chem*, 7(2), 141-158.
- Oprea, T. I., Davis, A. M., Teague, S. J., & Leeson, P. D. (2001). Is There a Difference between Leads and Drugs? A Historical Perspective. *Journal of Chemical Information and Computer Sciences*, 41(5), 1308-1315.
- Ou-Yang, S.-s., Lu, J.-y., Kong, X.-q., Liang, Z.-j., Luo, C., & Jiang, H. (2012). Computational drug discovery. *Acta Pharmacol Sin*, 33(9), 1131-1140.
- Park, H., Lee, S., Lee, S., & Hong, S. (2014). Structure-based de novo design and identification of D816V mutant-selective c-KIT inhibitors. *Org Biomol Chem*, 12(26), 4644-4655.
- Park, J. W., & Jo, W. H. (2009). Infiltration of water molecules into the oseltamivir-binding site of H274Y neuraminidase mutant causes resistance to oseltamivir. *Journal of chemical information and modeling*, 49(12), 2735-2741.
- Pielak, R. M., & Chou, J. J. (2011). Influenza M2 proton channels. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1808(2), 522-529.
- Pinto, L. H., & Lamb, R. A. (2006). The M2 Proton Channels of Influenza A and B Viruses. *Journal of Biological Chemistry*, 281(14), 8997-9000.
- Pipeline. (2011). Pipeline Pilot <http://accelrys.com/products/pipeline-pilot/>.

- Price, D. J., & Brooks, C. L. (2002). Modern protein force fields behave comparably in molecular dynamics simulations. *Journal of computational chemistry*, 23(11), 1045-1057.
- Quinn, B. A., Dash, R., Azab, B., Sarkar, S., Das, S. K., Kumar, S., . . . Fisher, P. B. (2011). Targeting Mcl-1 for the therapy of cancer. *Expert Opin Investig Drugs*, 20(10), 1397-1411.
- Rahman, M. M., Karim, M. R., Ahsan, M. Q., Khalipha, A. B. R., Chowdhury, M. R., & Saifuzzaman, M. (2012). *Use of computer in drug design and drug discovery: A review* (Vol. 1).
- Rao, J. R., Jha, A. K., Rawal, R. K., Sharon, A., Day, C. W., Barnard, D. L., . . . Chu, C. K. (2010). (-)-Carbodine: enantiomeric synthesis and in vitro antiviral activity against various strains of influenza virus including H5N1 (avian influenza) and novel 2009 H1N1 (swine flu). *Bioorganic & medicinal chemistry letters*, 20(8), 2601-2604.
- Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *J Mol Biol*, 261(3), 470-489. doi: 10.1006/jmbi.1996.0477
- Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A Fast Flexible Docking Method using an Incremental Construction Algorithm. *Journal of Molecular Biology*, 261(3), 470-489.
- Rathmell, J. C., & Thompson, C. B. (2002). Pathways of Apoptosis in Lymphocyte Development, Homeostasis, and Disease. *Cell*, 109(2), S97-S107.
- Richard, M., Deleage, C., Barthelemy, M., Lin, Y. P., Hay, A., Lina, B., & Ferraris, O. (2008). Impact of influenza A virus neuraminidase mutations on the stability, activity, and sensibility of the neuraminidase to neuraminidase inhibitors. *J Clin Virol*, 41(1), 20-24.
- Roberts, A. W., Seymour, J. F., Brown, J. R., Wierda, W. G., Kipps, T. J., Khaw, S. L., . . . Humerickhouse, R. (2012). Substantial susceptibility of chronic lymphocytic leukemia to BCL2 inhibition: results of a phase I study of navitoclax in patients with relapsed or refractory disease. *J Clin Oncol*, 30(5), 488-496.
- Rockey, W. M., & Elcock, A. H. (2006). Rapid computational identification of the targets of protein kinase inhibitors. *Curr Opin Drug Discov Devel*, 9(3), 326-331.
- Rodrigues, T., Roudnicky, F., Koch, C. P., Kudoh, T., Reker, D., Detmar, M., & Schneider, G. (2013). De novo design and optimization of Aurora A kinase inhibitors. *Chemical Science*, 4(3), 1229-1233.
- Rognan, D. (2006). Development and virtual screening of target libraries. *Journal of Physiology-Paris*, 99(2-3), 232-244.
- Russell, R. J., Haire, L. F., Stevens, D. J., Collins, P. J., Lin, Y. P., Blackburn, G. M., . . . Skehel, J. J. (2006). The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature*, 443(7107), 45-49.
- Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3), 327-341.
- Saeed, A., Khan, M. S., Rafique, H., Shahid, M., & Iqbal, J. (2014). Design, synthesis, molecular docking studies and in vitro screening of ethyl 4-(3-benzoylthioureido) benzoates as urease inhibitors. *Bioorganic Chemistry*, 52(0), 1-7.

- Saeyns, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, *23*(19), 2507-2517.
- Schlessinger, A., Khuri, N., Giacomini, K. M., & Sali, A. (2013). Molecular modeling and ligand docking for solute carrier (SLC) transporters. *Curr Top Med Chem*, *13*(7), 843-856.
- Shahlaei, M. (2013). Descriptor Selection Methods in Quantitative Structure–Activity Relationship Studies: A Review Study. *Chem Rev*, *113*(10), 8093-8103.
- Shan, Y., Ma, Y., Wang, M., & Dong, Y. (2012). Recent advances in the structure-based design of neuraminidase inhibitors as antiinfluenza agents. *Curr Med Chem*, *19*(34), 5885-5894.
- Sharma, H., Landau, M. J., Sullivan, T. J., Kumar, V. P., Dahlgren, M. K., Jorgensen, W. L., & Anderson, K. S. (2014). Virtual screening reveals allosteric inhibitors of the *Toxoplasma gondii* thymidylate synthase–dihydrofolate reductase. *Bioorganic & medicinal chemistry letters*, *24*(4), 1232-1235.
- Sharma, N., & Yap, C. (2012). Consensus QSAR model for identifying novel H5N1 inhibitors. *Molecular Diversity*, *16*(3), 513-524.
- Shie, J. J., Fang, J. M., Wang, S. Y., Tsai, K. C., Cheng, Y. S. E., Yang, A. S., . . . Wong, C. H. (2007). Synthesis of tamiflu and its phosphonate congeners possessing potent anti-influenza activity. *Journal of the American Chemical Society*, *129*(39), 11892-11893.
- Shinde, S., Mol, M., Jamdar, V., & Singh, S. (2014). Molecular modeling and molecular dynamics simulations of GPI 14 in *Leishmania major*: Insight into the catalytic site for active site directed drug design. *Journal of Theoretical Biology*, *351*(0), 37-46.
- Shoichet, B. K. (2004). Virtual screening of chemical libraries. *Nature*, *432*(7019), 862-865.
- Shtyrya, Y. A. M., L.V.;Bovinc,N.V. (2009). Influenza Virus Neuraminidase: Structure and Function. *Acta Naturae*, *1*(2), 26–32.
- Shu, M., Lin, Z., Zhang, Y., Wu, Y., Mei, H., & Jiang, Y. (2011). Molecular dynamics simulation of oseltamivir resistance in neuraminidase of avian influenza H5N1 virus. *Journal of molecular modeling*, *17*(3), 587-592.
- Smith, P. W., Sollis, S. L., Howes, P. D., Cherry, P. C., Starkey, I. D., Cobley, K. N., . . . Beresford, A. (1998). Dihydropyranocarboxamides Related to Zanamivir: A New Series of Inhibitors of Influenza Virus Sialidases. 1. Discovery, Synthesis, Biological Activity, and Structure–Activity Relationships of 4-Guanidino- and 4-Amino-4H-pyran-6-carboxamides. *Journal of medicinal chemistry*, *41*(6), 787-797.
- Sollis, S. L., Smith, P. W., Howes, P. D., Cherry, P. C., & Bethell, R. C. (1996). Novel inhibitors of influenza sialidase related to GG167 Synthesis of 4-amino and guanidino-4H-pyran-2-carboxylic acid-6-propylamides; selective inhibitors of influenza a virus sialidase. *Bioorganic & medicinal chemistry letters*, *6*(15), 1805-1808.
- Song, C. M., Lim, S. J., & Tong, J. C. (2009). Recent advances in computer-aided drug design. *Brief Bioinform*, *10*(5), 579-591.
- Song, L., Coppola, D., Livingston, S., Cress, D., & Haura, E. B. (2005). Mcl-1 regulates survival and sensitivity to diverse apoptotic stimuli in human non-small cell lung cancer cells. *Cancer Biol Ther*, *4*(3), 267-276.

- Song, Y., Huang, J., Zhou, D., Zha, H., & Giles, C. (2007). IKNN: Informative K-Nearest Neighbor Pattern Classification Knowledge Discovery in Databases: PKDD 2007. In J. Kok, J. Koronacki, R. Lopez de Mantaras, S. Matwin, D. Mladenic & A. Skowron (Eds.), (Vol. 4702, pp. 248-264): Springer Berlin / Heidelberg.
- Soto, A., Cecchini, R., Vazquez, G., & Ponzoni, I. (2008). A Wrapper-Based Feature Selection Method for ADMET Prediction Using Evolutionary Computing. In E. Marchiori & J. Moore (Eds.), *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics* (Vol. 4973, pp. 188-199): Springer Berlin Heidelberg.
- Souers, A. J., Levenson, J. D., Boghaert, E. R., Ackler, S. L., Catron, N. D., Chen, J., . . . Elmore, S. W. (2013). ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat Med*, *19*(2), 202-208.
- Sousa, S. F., Fernandes, P. A., & Ramos, M. J. (2006). Protein–ligand docking: Current status and future challenges. *Proteins: Structure, Function, and Bioinformatics*, *65*(1), 15-26.
- Srivastava, H. K., & Sastry, G. N. (2012). Molecular Dynamics Investigation on a Series of HIV Protease Inhibitors: Assessing the Performance of MM-PBSA and MM-GBSA Approaches. *Journal of chemical information and modeling*, *52*(11), 3088-3098.
- Subramaniam, S., Mehrotra, M., & Gupta, D. (2008). Virtual high throughput screening (vHTS)--a perspective. *Bioinformation*, *3*(1), 14-17.
- Sun, C., Zhang, X., Huang, H., & Zhou, P. (2006). Synthesis and evaluation of a new series of substituted acyl(thio)urea and thiadiazolo [2,3-a] pyrimidine derivatives as potent inhibitors of influenza virus neuraminidase. *Bioorganic & medicinal chemistry*, *14*(24), 8574-8581.
- Swayne, D. E. (2008). *Avian influenza* Ames, Iowa Carlton, Vic :: Wiley-Blackwell.
- SYBYL-X. (Version 1.3). 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA: Tripos International.
- Tait, S. W. G., & Green, D. R. (2010). Mitochondria and cell death: outer membrane permeabilization and beyond. *Nat Rev Mol Cell Biol*, *11*(9), 621-632.
- Tan, W. N., Khairuddean, M., Wong, K. C., Khaw, K. Y., & Vikneswaran, M. (2014). New cholinesterase inhibitors from *Garcinia atroviridis*. *Fitoterapia*.
- Tang, Y., Zhu, W., Chen, K., & Jiang, H. (2006). New technologies in computer-aided drug design: Toward target identification and new chemical entity discovery. *Drug Discovery Today: Technologies*, *3*(3), 307-313. doi: <http://dx.doi.org/10.1016/j.ddtec.2006.09.004>
- Tang, Y., Zhu, W., Chen, K., & Jiang, H. (2006). New technologies in computer-aided drug design: Toward target identification and new chemical entity discovery. *Drug Discov Today Technol*, *3*(3), 307-313.
- Tanwar, O., Tanwar, L., Shaquiquzzaman, M., Alam, M. M., & Akhter, M. (2014). Structure based virtual screening of MDPI database: Discovery of structurally diverse and novel DPP-IV inhibitors. *Bioorganic & medicinal chemistry letters*.
- Taubenberger, J. K., & Morens, D. M. (2008). The Pathology of Influenza Virus Infections. *Annual Review of Pathology: Mechanisms of Disease*, *3*(1), 499-522.
- Thede, S. M. (2004). An introduction to genetic algorithms. *J. Comput. Small Coll.*, *20*(1), 115-123.
- Thomas, L. W., Lam, C., & Edwards, S. W. (2010). Mcl-1; the molecular regulation of protein function. *FEBS Lett*, *584*(14), 2981-2989. doi: 10.1016/j.febslet.2010.05.061

- Thomas, T., McLean, K. C., McRobb, F. M., Manallack, D. T., Chalmers, D. K., & Yuriev, E. (2014). Homology modeling of human muscarinic acetylcholine receptors. *Journal of chemical information and modeling*, 54(1), 243-253.
- Times, Y. (2007). A diagram of influenza viral cell invasion and replication; . http://en.wikipedia.org/wiki/File:Virus_Replication_large.svg: Wikipedia.
- Todeschini, R., & Consonni, V. (2000). *Handbook of molecular descriptors* (Vol. 11): Wiley-VCH.
- Todeschini, R., & Consonni, V. (2009). *Molecular descriptors for chemoinformatics (2 volumes)* (Vol. 41): Wiley-VCH.
- Tomasic, T., & Peterlin Masic, L. (2012). Rhodanine as a scaffold in drug discovery: a critical review of its biological activities and mechanisms of target modulation. *Expert Opin Drug Discov*, 7(7), 549-560.
- Tong, W., Xie, Q., Hong, H., Shi, L., Fang, H., & Perkins, R. (2004). Assessment of Prediction Confidence and Domain Extrapolation of Two Structure-Activity Relationship Models for Predicting Estrogen Receptor Binding Activity. *Environ Health Perspect*, 112(12).
- Tripathi, S. K., & Singh, S. K. (2014). Insights into the structural basis of 3,5-diaminoindazoles as CDK2 inhibitors: prediction of binding modes and potency by QM-MM interaction, MESP and MD simulation. *Mol Biosyst*.
- Tropsha, A. (2010). Best Practices for QSAR Model Development, Validation, and Exploitation. *Molecular Informatics*, 29(6-7), 476-488.
- Tropsha, A., & Golbraikh, A. (2007). Predictive QSAR Modeling Workflow, Model Applicability Domains, and Virtual Screening. *Current Pharmaceutical Design*, 13(34), 3494-3504.
- Tropsha, A., Gramatica, P., & Gombar, V. K. (2003). The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR & Combinatorial Science*, 22(1), 69-77.
- Tropsha, A., & Wang, S. X. (2006). QSAR modeling of GPCR ligands: methodologies and examples of applications. *Ernst Schering Found Symp Proc*(2), 49-73.
- Tse, C. S., Alexander R.; Adickes, Jessica; Anderson, Mark G.; Chen, Jun; Jin, Sha; Johnson, Eric F.; Marsh, Kennan C.; Mitten, Michael J.; Nimmer, Paul; Roberts, Lisa; Tahir, Stephen K.; Xiao, Yu; Yang, Xiufen; Zhang, Haichao; . (2008). ABT-263: A Potent and Orally Bioavailable Bcl-2 Family Inhibitor. *Cancer Res*, 68(9), 3421-3428.
- Tsujimoto, Y., Ikegaki, N., & Croce, C. M. (1987). Characterization of the protein product of bcl-2, the gene involved in human follicular lymphoma. *Oncogene*, 2(1), 3-7.
- Tuccinardi, T. (2009). Docking-Based Virtual Screening: Recent Developments. *Combinatorial chemistry & high throughput screening*, 12(3), 303-314.
- Uddin, R., & Saeed, K. (2014). An exhaustive yet simple virtual screening campaign against Sortase A from multiple drug resistant Staphylococcus aureus. *Mol Biol Rep*.
- Varghese, J. N., Chandana Epa, V., & Colman, P. M. (1995). Three-dimensional structure of the complex of 4-guanidino-Neu5Ac2en and influenza virus neuraminidase. *Protein Science*, 4(6), 1081-1087.
- Varghese, J. N., Laver, W. G., & Colman, P. M. (1983). Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 [angst] resolution. *Nature*, 303(5912), 35-40.

- Varghese, J. N., McKimm-Breschkin, J. L., Caldwell, J. B., Kortt, A. A., & Colman, P. M. (1992). The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. *Proteins*, *14*(3), 327-332.
- Vergara-Jaque, A., Poblete, H., Lee, E. H., Schulten, K., Gonzalez-Nilo, F., & Chipot, C. (2012). Molecular basis of drug resistance in A/H1N1 virus. *Journal of chemical information and modeling*, *52*(10), 2650-2656.
- Verlinde, C. L. M. J., & Hol, W. G. J. (1994). Structure-based drug design: progress, results and challenges. *Structure*, *2*(7), 577-587.
- Verma, J., Khedkar, V. M., & Coutinho, E. C. (2010). 3D-QSAR in drug design - A review. *Current Topics in Medicinal Chemistry*, *10*(1), 95-115.
- Vihinen, M. (2012). How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis. *BMC Genomics*, *13 Suppl 4*, S2.
- von Itzstein, M. (2007). The war against influenza: discovery and development of sialidase inhibitors. *Nature reviews. Drug discovery*, *6*(12), 967-974.
- von Itzstein, M., & Thomson, R. (2009). Anti-influenza drugs: the development of sialidase inhibitors. *Handb Exp Pharmacol*(189), 111-154.
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of computational chemistry*, *25*(9), 1157-1174.
- Wang, M., Qi, J., Liu, Y., Vavricka, C. J., Wu, Y., Li, Q., & Gao, G. F. (2011). Influenza A Virus N5 Neuraminidase Has an Extended 150-Cavity. *Journal of virology*, *85*(16), 8431-8435.
- Warr, M. R., & Shore, G. C. (2008). Unique biology of Mcl-1: therapeutic opportunities in cancer. *Curr Mol Med*, *8*(2), 138-147.
- Warr, W. A. (1997). Combinatorial Chemistry and Molecular Diversity. An Overview†. *Journal of Chemical Information and Computer Sciences*, *37*(1), 134-140.
- Weaver, S., & Gleeson, M. P. (2008). The importance of the domain of applicability in QSAR modeling. *Journal of Molecular Graphics and Modelling*, *26*(8), 1315-1326.
- Wei, G., Margolin, A. A., Haery, L., Brown, E., Cucolo, L., Julian, B., . . . Golub, T. R. (2012). Chemical genomics identifies small-molecule MCL1 repressors and BCL-xL as a predictor of MCL1 dependency. *Cancer Cell*, *21*(4), 547-562.
- Wei, S. H., Dong, K., Lin, F., Wang, X., Li, B., Shen, J. J., . . . Zhang, H. Z. (2008). Inducing apoptosis and enhancing chemosensitivity to gemcitabine via RNA interference targeting Mcl-1 gene in pancreatic carcinoma cell. *Cancer Chemother Pharmacol*, *62*(6), 1055-1064.
- Weiser, J., Shenkin, P. S., & Still, W. C. (1999). Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *Journal of computational chemistry*, *20*(2), 217-230.
- Wen, W. H., Wang, S. Y., Tsai, K. C., Cheng, Y. S., Yang, A. S., Fang, J. M., & Wong, C. H. (2010). Analogs of zanamivir with modified C4-substituents as the inhibitors against the group-1 neuraminidases of influenza viruses. *Bioorganic & medicinal chemistry*, *18*(11), 4074-4084. doi: 10.1016/j.bmc.2010.04.010
- Weyhenmeyer, B., Murphy, A. C., Prehn, J. H., & Murphy, B. M. (2012). Targeting the anti-apoptotic Bcl-2 family members for the treatment of cancer. *Exp Oncol*, *34*(3), 192-199.
- Wiley, D. C., & Skehel, J. J. (1987). The Structure and Function of the Hemagglutinin Membrane Glycoprotein of Influenza Virus. *Annual Review of Biochemistry*, *56*(1), 365-394.

- Williams, M. A., Lew, W., Mendel, D. B., Tai, C. Y., Escarpe, P. A., Laver, W. G., . . . Kim, C. U. (1997). Structure-activity relationships of carbocyclic influenza neuraminidase inhibitors. *Bioorganic & medicinal chemistry letters*, 7(14), 1837-1842.
- Winkler, D. A. (2002). The role of quantitative structure-activity relationships (QSAR) in biomolecular discovery. *Brief Bioinform*, 3(1), 73-86.
- Wlodawer, A., & Vondrasek, J. (1998). INHIBITORS OF HIV-1 PROTEASE: A Major Success of Structure-Assisted Drug Design. *Annu Rev Biophys Biomol Struct*, 27(1), 249-284.
- Wong, C. F., & Bairy, S. (2013). Drug design for protein kinases and phosphatases: flexible-receptor docking, binding affinity and specificity, and drug-binding kinetics. *Curr Pharm Des*, 19(26), 4739-4754.
- Wong, K. Y., Duchowicz, P. R., Mercader, A. G., & Castro, E. A. (2012). QSAR applications during last decade on inhibitors of acetylcholinesterase in Alzheimer's disease. *Mini Rev Med Chem*, 12(10), 936-946.
- Woods, C. J., Malaisree, M., Long, B., McIntosh-Smith, S., & Mulholland, A. J. (2013). Computational Assay of H7N9 Influenza Neuraminidase Reveals R292K Mutation Reduces Drug Binding Affinity. *Sci. Rep.*, 3. doi: 10.1038/srep03561
- Woods, C. J., Malaisree, M., Pattarapongdilok, N., Sompornpisut, P., Hannongbua, S., & Mulholland, A. J. (2012). Long Time Scale GPU Dynamics Reveal the Mechanism of Drug Resistance of the Dual Mutant I223R/H275Y Neuraminidase from H1N1-2009 Influenza Virus. *Biochemistry*, 51(21), 4364-4375.
- Wu, Y., Bi, Y., Vavricka, C. J., Sun, X., Zhang, Y., Gao, F., . . . Gao, G. F. (2013). Characterization of two distinct neuraminidases from avian-origin human-infecting H7N9 influenza viruses. *Cell Res*, 23(12), 1347-1355.
- Wu, Y., Qin, G., Gao, F., Liu, Y., Vavricka, C. J., Qi, J., . . . Gao, G. F. (2013). Induced opening of influenza virus neuraminidase N2 150-loop suggests an important role in inhibitor binding. *Sci. Rep.*, 3.
- Xu, H.-Y., Zou, J.-W., Yu, Q.-S., Wang, Y.-H., Zhang, J.-Y., & Jin, H.-X. (2007). QSPR/QSAR models for prediction of the physicochemical properties and biological activity of polybrominated diphenyl ethers. *Chemosphere*, 66(10), 1998-2010.
- Xue, L., & Bajorath, J. (2000). Molecular descriptors in chemoinformatics, computational combinatorial chemistry, and virtual screening. *Comb Chem High Throughput Screen*, 3(5), 363-372.
- Yamashita, M. (2010). Laninamivir and its prodrug, CS-8958: long-acting neuraminidase inhibitors for the treatment of influenza. *Antivir Chem Chemother*, 21(2), 71-84.
- Yamashita, M., Tomozawa, T., Kakuta, M., Tokumitsu, A., Nasu, H., & Kubo, S. (2009). CS-8958, a prodrug of the new neuraminidase inhibitor R-125489, shows long-acting anti-influenza virus activity. *Antimicrobial agents and chemotherapy*, 53(1), 186-192.
- Yang, C.-Y., & Wang, S. (2012). Analysis of Flexibility and Hotspots in Bcl-xL and Mcl-1 Proteins for the Design of Selective Small-Molecule Inhibitors. *ACS Medicinal Chemistry Letters*, 3(4), 308-312.
- Yang, G. F., & Huang, X. (2006). Development of quantitative structure-activity relationships and its application in rational drug design. *Curr Pharm Des*, 12(35), 4601-4611.
- Yap, C. W. (2011). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of computational chemistry*, 32(7), 1466-1474.

- Yarnitzky, T., Levit, A., & Niv, M. Y. (2010). Homology modeling of G-protein-coupled receptors with X-ray structures on the rise. *Curr Opin Drug Discov Devel*, 13(3), 317-325.
- Yecies, D., Carlson, N. E., Deng, J., & Letai, A. (2010). Acquired resistance to ABT-737 in lymphoma cells that up-regulate MCL-1 and BFL-1. *Blood*, 115(16), 3304-3313.
- Yee, L. C., & Wei, Y. C. (2012). Current Modeling Methods Used in QSAR/QSPR *Statistical Modelling of Molecular Descriptors in QSAR/QSPR* (pp. 1-31): Wiley-VCH Verlag GmbH & Co. KGaA.
- Yip, K. W., & Reed, J. C. (2008). Bcl-2 family proteins and cancer. *Oncogene*, 27(50), 6398-6406.
- Youle, R. J., & Strasser, A. (2008). The BCL-2 protein family: opposing activities that mediate cell death. *Nat Rev Mol Cell Biol*, 9(1), 47-59.
- Yuen, K. Y., Chan, P. K. S., Peiris, M., Tsang, D. N. C., Que, T. L., Shortridge, K. F., . . . Cheng, A. F. B. (1998). Clinical features and rapid viral diagnosis of human disease associated with avian influenza A H5N1 virus. *The Lancet*, 351(9101), 467-471.
- Yuriev, E., & Ramsland, P. A. (2013). Latest developments in molecular docking: 2010–2011 in review. *Journal of Molecular Recognition*, 26(5), 215-239.
- Zhang, H., Guttikonda, S., Roberts, L., Uziel, T., Semizarov, D., Elmore, S. W., . . . Lam, L. T. (2011). Mcl-1 is critical for survival in a subgroup of non-small-cell lung cancer cell lines. *Oncogene*, 30(16), 1963-1968.
- Zhang, S. (2011). Computer-Aided Drug Discovery and Development. In S. D. Satyanarayanajois (Ed.), *Drug Des Discov* (Vol. 716, pp. 23-38): Humana Press.
- Zheng, M., Yu, K., Liu, H., Luo, X., Chen, K., Zhu, W., & Jiang, H. (2006). QSAR analyses on avian influenza virus neuraminidase inhibitors using CoMFA, CoMSIA, and HQSAR. *Journal of computer-aided molecular design*, 20(9), 549-566.
- Zhou, T., Li, G., Cao, B., Liu, L., Cheng, Q., Kong, H., . . . Gao, N. (2013). Downregulation of Mcl-1 through inhibition of translation contributes to benzyl isothiocyanate-induced cell cycle arrest and apoptosis in human leukemia cells. *Cell Death Dis*, 4, e515.
- Zoete, V., Grosdidier, A., & Michielin, O. (2009). Docking, virtual high throughput screening and in silico fragment-based drug design. *Journal of Cellular and Molecular Medicine*, 13(2), 238-248.
- Zuzarte-Luis, V., & Hurle, J. M. (2002). Programmed cell death in the developing limb. *Int J Dev Biol*, 46(7), 871-876.