

**VIRTUAL SCREENING FOR JNK INHIBITORS AND  
PREDICTION OF PXR ACTIVITY USING  
COMPUTATIONAL MODELS**

**WOO SZE KWANG**

*(B.Sc (Hons), UB)*

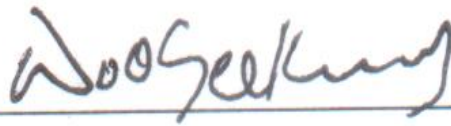
**A THESIS SUBMITTED  
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY  
DEPARTMENT OF PHARMACY  
NATIONAL UNIVERSITY OF SINGAPORE**

**2015**

## 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.

A handwritten signature in black ink, appearing to read 'Woo Sze Kwang', is written above a horizontal line.

Woo Sze Kwang

29 December 2014

## **Acknowledgements**

I would like to express my most sincere and deepest appreciation to my supervisor Asst. Prof. Yap Chun Wei for all the support, patience, encouragement, guidance, advice and for providing the opportunities to do multiple collaborations throughout my Ph.D. study.

I would like to thank all current and past members of the Pharmaceutical Data Exploration Laboratory, collaborators Asst. Prof. Ho Han Kiat, Ho Jia Pei, Prof. Thomas Chang, Asst. Prof. Lau Aik Jiang, Assoc. Prof. Eric Chan Chun Yong, Dr. Chng Hui Ting, my thesis committee members, members of the Science Training in Action (SCIENTIA) program from NUS High School of Mathematics and Science and members of the Polytechnic Science Research Program (PSRP) from Ngee Ann Polytechnic during my term of candidature for their invaluable discussions and input into this current project.

I would also like to thank the staff and members of the NUS Centre for Computational Science and Engineering, NUS High Performance Computing Centre and Medicinal Chemistry Research Laboratory for their assistance in providing the high-performance server clusters in which to carry out the various data mining experiments, and in the additional modeling and docking software.

To my family members, thank you very much for all the support and understanding throughout my study.

Sze Kwang

# Table of contents

<b>Declaration</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iii</b>
<b>Table of contents</b>	<b>iv</b>
<b>Summary</b>	<b>ix</b>
<b>List of Tables</b>	<b>xiii</b>
<b>List of Figures</b>	<b>xv</b>
<b>1. Introduction</b> .....	<b>1</b>
1.1 Physiology of JNK .....	3
1.2 Inhibition of JNK .....	4
1.2.1 ATP binding site .....	6
1.2.2 JIP binding site .....	7
1.2.3 DFG binding site .....	8
1.3 Physiology of PXR .....	11
1.4 Current drug discovery efforts to identify JNK inhibitors and PXR activators .....	12
1.5 Computational methods to screen and validate JNK inhibitors and PXR activators .....	14
1.5.1 Ligand-based methods .....	15
1.5.2 Structure-based methods .....	18
1.6 Objectives .....	23
1.6.1 Objective 1: Identify potent JNK inhibitors that target JIP binding site and DFG binding site .....	23
1.6.2 Objective 2: Form an <i>in silico</i> model for human and zebrafish PXR activators .....	25

<b>2.</b>	<b>Phase 1: Development of QSAR model for JIP binding site inhibitors</b>	<b>28</b>
2.1	Summary	28
2.2	Methods	29
2.2.1	Dataset	29
2.2.2	Calculation of descriptors	32
2.2.3	QSAR model development	33
2.2.4	Defining the applicability domain with multiple thresholds	33
2.2.5	Consensus model	34
2.2.6	QSAR model validation	35
2.3	Results and discussion	36
2.3.1	Validation of the consensus model	36
2.3.2	Screening of ZINC chemical library using QSAR model	37
2.3.3	Structural similarity and distribution	38
2.3.4	Molecular descriptors	41
2.4	Conclusion	43
<b>3.</b>	<b>Phase 2: Development of structure-based model for JIP binding site inhibitors</b>	<b>44</b>
3.1	Summary	44
3.2	Methods	44
3.2.1	Preparation of JNK1 structure	44
3.2.2	Virtual screening for potential JNK1-JIP1 inhibitors	46
3.2.3	Analysis of interaction profiles of identified JNK1-JIP1 inhibitors	50

3.3	Results and discussion	51
3.3.1	Initial virtual screening using docking with pharmacophore constraints	51
3.3.2	Refinement of virtual screening using molecular dynamics	54
3.3.3	Clustering	59
3.3.4	Analysis of interaction profiles	60
3.4	Conclusion	76
<b>4.</b>	<b>Development of structure-based model for DFG binding site inhibitors</b>	<b>77</b>
4.1	Summary	77
4.2	Methods	77
4.2.1	Ranking of DFG inhibitors based on scoring function	77
4.2.2	Analysis of interaction profiles of identified DFG site inhibitors	79
4.2.3	Preparation of compounds for <i>in vitro</i> testing	79
4.3	Results and discussion	80
4.3.1	Consensus scoring functions using different docking software	80
4.3.2	KINOMEScan Assay	83
4.4	Conclusion	85
<b>5.</b>	<b>Docking study of human Pregnane X Receptor (PXR) activators</b>	<b>86</b>
5.1	Summary	86
5.2	Structures of human PXR	86
5.3	Materials and methods	88

5.3.1	Biological activity	88
5.3.2	Initial docking of compounds	89
5.3.3	Analysis of interactions	90
5.4	Results and discussion	91
5.4.1	Initial docking results	91
5.4.2	Analysis of interactions	93
5.5	Conclusion	102
<b>6.</b>	<b>Structure-based study on human and zebrafish PXR activators</b>	<b>103</b>
6.1	Summary	103
6.2	Introduction to zebrafish PXR	103
6.3	Materials and methods	104
6.3.1	Biological activity	104
6.3.2	Homology modeling of zebrafish PXR	105
6.3.3	Structure-based study of compounds	108
6.4	Results and discussion	109
6.4.1	Structures of human and zebrafish PXR	109
6.4.2	Initial docking results	111
6.4.3	Analysis of interactions	113
6.4.4	Molecular dynamics as a filter tool	147
6.5	Conclusion	149
<b>7.</b>	<b>Conclusion</b>	<b>150</b>
7.1	Major findings	150
7.1.1	Ligand-based study	150

7.1.2	Structure-based study of ZINC database compounds on JIP binding site	151
7.1.3	Structure-based study of ZINC database compounds on DFG binding site	152
7.1.4	<i>in silico</i> model of human and zebrafish PXR activators	152
7.2	Limitations	153
7.3	Future work	154
7.4	Dual-targeting approach and polypharmacology	156
	<b>Bibliography</b>	<b>158</b>



## Summary

As drug development is time consuming and costly, compounds that are likely to fail should be weeded out early through the use of assays and toxicity screens. Computational methods are favourable complementary techniques. Nevertheless, it is not exploited to its full potential due to: models that were built from small data sets, a lack of applicability domain (AD), not being readily available for use, not following the Organisation for Economic Co-operation and Development(OECD) quantitative structure-activity relationship(QSAR) validation guidelines in developing a QSAR model, and due to static nature of the protein structure obtained from X-ray crystallography, difficulties in consistency and reproducibility of each docking process and result, and the absence of a strong correlation between the predicted binding modes and the actual biological activity of the compound in *in vitro* experiments in docking studies.

This thesis attempts to address these problems with the following strategies. First, the data augmentation approach using putative negatives was used to increase the information content of data sets. Second, ensemble methods were investigated as the approach to improve accuracies of QSAR models. Third, predictive models are to be built from data sets as large as possible, with the application of AD to define the usability of these models. Fourth, the QSAR models were built according to the guidance set out by the OECD. Fifth, the

models were packaged into a free software to facilitate independent evaluation and comparison of QSAR models.

The usefulness of these strategies were evaluated using structure activity relationship(SAR) data set in c-Jun N-terminal kinase (JNK) inhibitors. The compounds were used in QSAR modeling, docking, pharmacophore mapping and molecular dynamics simulations to refine the list of inhibitors. To the best of our knowledge, the models are from significantly larger training data with the effects of increased AD and reduced false positive hits.

In structure-based approaches, the compounds obtained from the consensus models obtained were docked by using a variety of rigid and flexible docking methods, pharmacophore mapping and molecular dynamics simulations as well as various modeling algorithms and different docking software to prepare the structure of the enzyme and to obtain a consensus of different docking results to improve on the correlation between the prediction made by the docking software and the actual biological activity of the compound in the *in vitro* setting. The various approaches are useful, to varying extents, for improving the virtual screening of potential drug leads for specific pharmacodynamic properties.

The results of the virtual screening project on JNK inhibitors prioritized 1 compound targeting the JNK-Interacting Protein 1(JIP1) binding site of JNK1. A set of compounds that target the DFG binding site of JNK2 were also tested for

inhibitory activity after a docking study. One compound showed 20% inhibition on JNK1, and one compound showed 19% inhibition on JNK2 at a concentration of 40  $\mu$ M.

To further refine and improve on structure-based methods used in this project, an *in silico* study was carried out on a set of approved drugs and one chemical compound available at PubChem. The aims of this study are to form a prediction model for both human and zebrafish pregnane X receptor (PXR) activators and non-activators due to the lower cost and shorter waiting period to obtain CYP3A4 gene expression data in zebrafish as decided and advised by collaborators. A docking study with simple energy minimization and homology modeling techniques was initially carried out. Docked poses would be identified using hierarchical clustering. The binding orientation of each compound would be identified using this method.

This study would eventually include pharmacophore mapping and molecular dynamics simulations to refine the model and provide an additional filter mechanism. Through molecular dynamics simulations, the stability of the trajectory, protein movement that is predicted to occur in the complex between PXR and retinoid X receptor (RXR), the duration of the ligand inside the ligand binding domain (LBD) of PXR and the length and duration of hydrogen bonds and pi interactions that are formed between the ligand and the receptor are obtained in this study. A more accurate prediction model was developed in this

phase of the project where zebrafish PXR activators are predicted to form interactions with at least one conserved residue that is also present in human PXR, and compounds which activate PXR and have similar chemical groups are identified.

## List of Tables

2.1	Performance of consensus model on each validation set and on the entire dataset .....	36
2.2	Virtual screening of ZINC library using consensus model .....	38
2.3	Structural diversity of compounds .....	39
2.4	A selection of ZINC compounds not reported as JNK inhibitors that are the most dissimilar from the positive compound BI-78D3 .....	40
2.5	Descriptors used in the consensus model .....	42
3.1	List of pharmacophore points selected in model .....	46
3.2	Docking results using Molecular Operating Environment .....	51
3.3	Chemical structures of validation compounds .....	56
3.4	Chemical structures and docking scores for molecular dynamics trajectories of 3 ns duration .....	58
3.5	Cluster size and total population of each compound .....	59
3.6	Distance between geometric center of aromatic ring in compound and Arg 127 .....	70
3.7	Distance between large atoms in Arg 127 of JNK1 and compound .....	75
4.1	Results of docking scores on various ZINC compounds using different docking software .....	81
4.2	KINOMEScan assay results .....	84
5.1	List of docked scores .....	92
5.2	Amino acid residues involved in the binding of compounds to human PXR .....	94
5.3	Distance between first set of compounds and residue where potential hydrogen bond interactions might be present .....	102
6.1	List of amino acid residues that form the ligand-binding domain in human PXR and zebrafish PXR .....	110
6.2	List of docked scores .....	112

6.3	Amino acid residues involved in the binding of compounds to human PXR .....	113
6.4	Amino acid residues involved in the binding of compounds to zebrafish PXR .....	114
6.5	Distance between second set of compounds and residue where potential hydrogen bond interactions might be present .....	145
6.6	Distance between geometric center of aromatic rings in second set of compounds and residue where potential pi interactions might be present .....	146

## List of Figures

1.1	Ribbon diagram of JNK with stick and surface representation of ligand in each binding pocket .....	10
2.1	Flowchart of steps taken to obtain the training and validation set .....	31
3.1	Ribbon structure of JNK1 with truncated JIP1 .....	45
3.2	Close-up of truncated JIP1 in binding pocket of JNK1 .....	45
3.3	Pharmacophore model developed using MOE 2012.10 .....	47
3.4	Flowchart of steps taken in this virtual screening workflow .....	55
3.5	Pose of Compound 19.....	62
3.6	Graph of distance between geometric center of aromatic ring in Compound 19 and Arg 127 against instance in trajectory .....	62
3.7	Graph of distance between geometric center of aromatic ring in Compound 25 and Arg 127 against instance in trajectory .....	63
3.8	Pose of Compound 19 over ribbon diagram of JNK1 .....	63
3.9	Pose of Compound 19 over ribbon diagram of JNK1 with polar residues .....	64
3.10	Pose of Compound 19 over ribbon diagram of JNK1 with hydrophobic residues .....	64
3.11	Pose of Compound BI-98A10 .....	65
3.12	Pose of Compound BI-90H8 .....	65
3.13	Graph of distance between geometric center of aromatic ring in Compound BI-98A10 and Arg 127 against instance in trajectory .....	66
3.14	Graph of distance between geometric center of aromatic ring in Compound BI-90H8 and Arg 127 against instance in trajectory .....	66
3.15	Pose of Compound BI-98A10 over ribbon diagram of JNK1 with polar residues .....	67
3.16	Pose of Compound BI-90H8 over ribbon diagram of JNK1 with polar residues .....	67

3.17	Pose of Compound BI-98A10 over ribbon diagram of JNK1 with hydrophobic residues .....	68
3.18	Pose of Compound BI-90H8 over ribbon diagram of JNK1 with hydrophobic residues .....	68
3.19	Graph of distance between geometric center of aromatic ring in Compound BI-83B3 and Arg 127 against instance in trajectory .....	69
3.20	Graph of distance between geometric center of aromatic ring in Compound 3 and Arg 127 against instance in trajectory .....	69
3.21	Graph of distance between geometric center of aromatic ring in Compound 10 and Arg 127 against instance in trajectory .....	70
3.22	Pose of Compound BI-90H8 .....	71
3.23	Graph of distance between large atom in Compound BI-90H8 and Arg 127 against instance in trajectory .....	71
3.24	Graph of distance between large atom in Compound BI-90E2 and Arg 127 against instance in trajectory .....	72
3.25	Graph of distance between large atom in Compound BI-90E7 and Arg 127 against instance in trajectory .....	72
3.26	Pose of ZINC03405497 .....	73
3.27	Graph of distance between large atom in ZINC03405497 and Arg 127 against instance in trajectory .....	74
3.28	Pose of ZINC03405497 over ribbon diagram of JNK1 with charged residues .....	74
3.29	Pose of ZINC03405497 over ribbon diagram of JNK1 with hydrophobic residues .....	75
5.1	Correlation curve for MM/GBVI binding free energy against fold increase in human PXR activity over vehicle control .....	93
5.2	Correlation curve for molecular weight against fold increase in human PXR activity over vehicle control .....	93
5.3	Pose of etravirine .....	95
5.4	Graph of distance between large atom in etravirine and Gln 285 against instance in trajectory .....	95



5.5	Pose of etravirine over ribbon diagram of human PXR with hydrophobic residues .....	96
5.6	Pose of rilpivirine .....	97
5.7	Graph of distance between large atom in rilpivirine and Gln 285 against instance in trajectory .....	97
5.8	Pose of rilpivirine over ribbon diagram of human PXR with hydrophobic residues .....	98
5.9	Pose of efavirenz .....	98
5.10	Graph of distance between large atom in efavirenz and His 407 against instance in trajectory .....	99
5.11	Pose of efavirenz over ribbon diagram of human PXR with hydrophobic residues .....	99
5.12	Graph of distance between large atom in efavirenz and His 407 against instance in trajectory .....	100
5.13	Graph of distance between large atom in PCN and His 407 against instance in trajectory .....	101
5.14	Graph of distance between large atom in nevirapine and His 407 against instance in trajectory .....	101
6.1	Amino acid sequence for homology model of zebrafish PXR and human PXR using PDB structure 1SKX as template .....	106
6.2	Ramachandran plot of X-ray structure of human PXR .....	107
6.3	Ramachandran plot of homology model of zebrafish PXR .....	108
6.4	Graph of distance between large atom in rifampicin and Ser 247 against instance in trajectory .....	116
6.5	Graph of distance between large atom in rifampicin and Thr 247 against instance in trajectory .....	116
6.6	Pose of phenytoin in human PXR .....	117
6.7	Graph of distance between large atom in phenytoin and Gln 285 against instance in trajectory .....	117

6.8	Pose of phenytoin over ribbon diagram of human PXR with hydrophobic residues .....	118
6.9	Pose of phenytoin in zebrafish PXR .....	118
6.10	Graph of distance between geometric center of aromatic ring in phenytoin and Phe 288 against instance in trajectory .....	119
6.11	Pose of phenytoin over ribbon diagram of zebrafish PXR with hydrophobic residues .....	119
6.12	Pose of carbamazepine in human PXR .....	120
6.13	Graph of distance between large atom in carbamazepine and Ser 247 against instance in trajectory .....	121
6.14	Pose of carbamazepine over ribbon diagram of human PXR with hydrophobic residues .....	121
6.15	Pose of carbamazepine in zebrafish PXR .....	122
6.16	Graph of distance between large atom in carbamazepine and Phe 288 against instance in trajectory .....	122
6.17	Pose of carbamazepine over ribbon diagram of zebrafish PXR with hydrophobic residues .....	123
6.18	Pose of dexamethasone in human PXR .....	124
6.19	Graph of distance between large atom in dexamethasone and Gln 285 against instance in trajectory .....	124
6.20	Pose of dexamethasone over ribbon diagram of human PXR with hydrophobic residues .....	125
6.21	Pose of dexamethasone in zebrafish PXR .....	125
6.22	Graph of distance between large atom in dexamethasone and Met 323 against instance in trajectory .....	126
6.23	Pose of dexamethasone over ribbon diagram of zebrafish PXR with hydrophobic residues .....	126
6.24	Pose of nafcillin in human PXR .....	127
6.25	Graph of distance between large atom in nafcillin and Gln 285 against instance in trajectory .....	128

6.26	Pose of nafcillin over ribbon diagram of human PXR with hydrophobic residues .....	128
6.27	Pose of nafcillin in zebrafish PXR .....	129
6.28	Graph of distance between large atom in nafcillin and Met 323 against instance in trajectory .....	129
6.29	Pose of nafcillin over ribbon diagram of zebrafish PXR with hydrophobic residues .....	130
6.30	Pose of efavirenz in zebrafish PXR .....	131
6.31	Graph of distance between large atom in efavirenz and Met 323 against instance in trajectory .....	131
6.32	Pose of efavirenz over ribbon diagram of zebrafish PXR .....	132
6.33	Pose of pioglitazone in zebrafish PXR .....	133
6.34	Graph of distance between geometric center of aromatic ring in pioglitazone and Trp 299 in zebrafish PXR against instance in trajectory .....	133
6.35	Pose of pioglitazone over ribbon diagram of zebrafish PXR with hydrophobic residues .....	134
6.36	Graph of distance between geometric center of aromatic ring in pioglitazone and Trp 299 in human PXR against instance in trajectory .....	134
6.37	Pose of prednisone in human PXR .....	135
6.38	Graph of distance between large atom in prednisone and Gln 285 against instance in trajectory .....	136
6.39	Pose of prednisone over ribbon diagram of human PXR with hydrophobic residues .....	136
6.40	Pose of prednisone in zebrafish PXR .....	137
6.41	Graph of distance between large atom in prednisone and Met 323 against instance in trajectory .....	137
6.42	Pose of prednisone over ribbon diagram of zebrafish PXR with hydrophobic residues .....	138

6.43	Graph of distance between large atom in rufinamide and Ser 247 against instance in trajectory .....	139
6.44	Graph of distance between large atom in rufinamide and Gln 285 against instance in trajectory .....	139
6.45	Graph of distance between large atom in rufinamide and His 407 against instance in trajectory .....	140
6.46	Graph of distance between geometric center of aromatic ring in rufinamide and Phe 288 in human PXR against instance in trajectory .....	140
6.47	Graph of distance between large atom in rufinamide and Met 323 against instance in trajectory .....	141
6.48	Graph of distance between geometric center of aromatic ring in rufinamide and Phe 288 in zebrafish PXR against instance in trajectory .....	141
6.49	Graph of distance between geometric center of aromatic ring in rufinamide and Trp 299 in zebrafish PXR against instance in trajectory .....	142
6.50	Graph of distance between large atom in acetaminophen and Ser 247 against instance in trajectory .....	142
6.51	Graph of distance between large atom in acetaminophen and Gln 285 against instance in trajectory .....	143
6.52	Graph of distance between large atom in acetaminophen and His 407 against instance in trajectory .....	143
6.53	Graph of distance between geometric center of aromatic ring in acetaminophen and Phe 288 in human PXR against instance in trajectory .....	144
6.54	Graph of distance between large atom in acetaminophen and Met 323 against instance in trajectory .....	144
6.55	Graph of distance between geometric center of aromatic ring in acetaminophen and Phe 288 in zebrafish PXR against instance in trajectory .....	145

## **Chapter 1**

### **Introduction**

Eukaryotic cells undergo a wide range of metabolic processes in order to function effectively. For many of these metabolic processes that occur in mammalian and human cells, enzymes are involved to catalyze a wide range of chemical reactions that are necessary to bring about changes to the cell. Important enzyme families that are currently being studied includes the protein kinase and the cytochrome monooxygenase family.

The protein kinase family consists of 518 kinases[1] and due to their critical role in phosphorylation of proteins, they played important roles in many aspects of cellular physiology which include cell division, proliferation, differentiation and apoptosis. Therefore, the design of drugs which inhibit protein kinases have been carried out by the pharmaceutical industry for over 20 years. There are currently 14 Food and Drug Administration (FDA) approved small molecule drugs which target protein kinases on the market, in addition to many more kinase inhibitors currently in clinical trials. Furthermore, it is estimated that over one quarter of all pharmaceutical drug design targets are protein kinases[2,3].

One class of protein kinase is mitogen-activated protein kinase (MAPK) enzymes that are activated in response to external environmental stimuli such as stress. Previously named as stress-activated protein kinase (SAPK), MAPK enzymes consist of c-Jun N-terminal kinase (JNK) enzymes, p38 enzymes and extracellular signal regulating kinase (ERK) enzymes where they catalyze the reaction of adenosine triphosphate (ATP) and a protein to adenosine diphosphate (ADP) and a phosphoprotein[4-9]. Among these, JNK is particularly interesting as it has been linked to several human diseases such as stroke, diabetes, atherosclerosis, neurodegenerative diseases, autoimmune and inflammatory diseases, multiple sclerosis, asthma, inflammatory bowel disease, psoriasis and cancer[10-12].

The cytochrome P450 monooxygenase family consists of 57 enzymes and may be found in all tissues[13]. Cytochrome P450 enzymes have an important role in metabolism, activation and detoxification of various chemical compounds, drugs and xenobiotics that are taken in by the human body. Therefore, cytochrome P450 induction and inhibition data are required to obtain regulatory approval for various drugs and chemical compounds. Human liver cells may be used as a biomarker to investigate cytochrome P450 activity[14]. Cytochrome P450 enzymes are also present in other species and could potentially be used as an alternative biomarker[15,16].

Expression of the gene that encodes for the cytochrome P450 family 3 subfamily A polypeptide 4 (CYP3A4) monooxygenase enzyme is strongly correlated with activation of Pregnane X Receptor (PXR) and the heterodimer being formed between PXR and Retinoid X Receptor (RXR)[17-21]. PXR and RXR are nuclear receptors which are part of a superfamily of 48 transcription factors[22]. PXR activity has important roles in regulating drug metabolism and drug transportation, along with other completely different processes such as karyopherin-mediated nuclear import. PXR also has the capacity to accept and bind to a very large variety of different drugs and chemical compounds of various molecular weights[23]. Therefore, the study of PXR activity is interesting as expression of CYP3A4 gene could potentially be used as a screening mechanism to predict and reduce drug and xenobiotic metabolism and toxicity.

### **1.1 Physiology of JNK**

JNKs are serine/threonine protein kinases which fall within the CMGC group of protein kinases and consist of three isoforms, JNK1, JNK2 and JNK3[24]. JNK1 and JNK2 are found in all parts of the human body[25,26], and JNK3 are found predominantly in the brain, heart and testes[27-29]. JNK enzymes catalyze the transfer of a phosphate group from adenosine triphosphate (ATP) that is present throughout the human body to a protein to form a phosphoprotein and adenosine diphosphate (ADP). This protein consists of a c-Jun component that is present in activator protein-1 (AP-1) which gives JNK its

name. c-Jun is the name for the messenger RNA codon which codes for the residue and structure of the protein that contains the c-Jun component. The c-Jun component of AP-1 has two serine residues which will bind to Thr183, Pro184 and Tyr185 of JNK (PDB ID: 3ELJ). ATP will bind at the ATP binding site of JNK, where a phosphate group from ATP will be transferred over to the c-Jun component of AP-1 and similar proteins with a c-Jun component. ATP will become ADP, and the protein will become a phosphoprotein. The phosphoprotein will then proceed further along the metabolic pathway, where if the phosphoprotein is produced constantly in response to stress, it will lead to a wide range of metabolic disorders, inflammatory and immune reactions, and neurodegenerative disorders.

## **1.2 Inhibition of JNK**

Since JNK was discovered in the 1990s, there have been studies to test the effects of inhibiting the enzyme as well as removal of one or more JNK isoform through gene knockout on animal studies and on cell studies. The results show increased insulin sensitivity on insulin receptors[30], improved arthritic scores[31], and improved neural function[32] in mice exposed to high lipid diet, inflammatory disease and neurotoxins. In cancer cell lines, there is also a reduction in cancer cell growth and size when JNK is inhibited in the *in vitro* setting[33,34]. Inhibition of JNK will cause ATP to bind less easily to the ATP binding site. The entire protein will eventually be broken down and removed by



an ubiquitin pathway. Hence, by inhibiting this part of the metabolic pathway, it could provide benefits in treating and preventing the various human diseases that are caused due to activation of the JNK pathway[29]. These diseases include stroke[35], type-2 diabetes, atherosclerosis, neurodegenerative diseases including Alzheimer's, Parkinson's and Huntington's diseases, autoimmune and inflammatory diseases including rheumatoid arthritis, multiple sclerosis, asthma, inflammatory bowel disease, psoriasis and cancer[10-12].

JNK can be inhibited by both competitive and non-competitive methods to reduce the activity of the enzyme. A competitive inhibitor occupies the ATP binding site to block ATP from binding. A non-competitive inhibitor will bind at an allosteric site on the JNK enzyme which causes the tertiary structure of JNK to fold and reduce the space of the ATP binding pocket. This increases the amount of steric hindrance at the ATP binding pocket, making ATP less likely to bind to the ATP pocket. There are currently two known non-competitive sites, one of them is the JNK interacting protein (JIP) site where there is a protein that is very specific to bind only to JNK at this site, and the other is the DFG site which is named after the one letter symbol of the aspartic acid, phenylalanine and glycine residues that are present at that binding site in JNK. For all three JNK isoforms, the ATP pocket, JIP pocket and DFG pocket have high sequence identity.

X-ray crystal structures of JNK have shown that their binding site residues have a high degree of similarity between all three JNK isoforms. This leads to the

obstacle in designing relatively selective small molecule inhibitors where the electrostatics and shape of the binding site tend to be similar. As a result of the highly-conserved structures of functional motifs present in JNKs and kinases in general, this makes developing relatively selective inhibitors for JNKs difficult[36].

For competitive inhibition, designing inhibitors that target the ATP site are likely to have more potential selectivity and specificity problems since there are many kinase enzymes that use ATP as a substrate. This could lead to an increase in side effects if the other kinase enzymes that are affected do not contribute to the disease process. Designing inhibitors that target the JIP or DFG site are expected to be more novel and have a less saturated chemical space to explore, hence this research project will attempt to address this knowledge gap. Some existing inhibitors, such as the set of compounds published by Chen et al, and the drug, imatinib were found to bind to both the ATP binding site, as well as the JIP or DFG binding site[37].

### **1.2.1 ATP binding site**

The ATP binding site is the most widely studied among the three binding site as it was identified when JNK was discovered. The first small-molecule inhibitor targeting the ATP binding site was discovered in 2001[38], and more than 900 compounds that target this site have since been published and registered

on patent databases as a novel inhibitor of JNK. The majority of published studies which examine the effects of inhibition of JNK using *in silico*, *in vitro* and *in vivo* models also use ATP competitive inhibitors[37,39-54]. The ATP binding site consists of 9 hydrophobic residues (Val 40, Ala 53, Ile 86, Met 108, Leu 110, Met 111, Ala 113, Val 158 and Leu 168) and 2 polar and charged residues (Glu 109 and Asp 112) from X-ray crystal structure of SP600125 complex with JIP1 and JNK1 (PDB entry: 1UKI)[55-58]. Inhibitors which bind at this active site are referred to as Type I inhibitors and usually meet the pharmacophore model by Traxler[59] as well as by forming hydrogen bonds with Glu 109 and/or Met 111 at the hinge region.

### **1.2.2 JIP binding site**

The JNK-interacting protein (JIP) binding site has been discovered and published in detail where a study by Barr et al have discovered a small and truncated form of JIP consisting of only 11 amino acid residues of JIP to be sufficient to start the inhibition mechanism of JNK at the JIP active site[60]. JIP1 is a scaffold protein which binds to JNK1 and reduces the volume of the ATP-binding site. This produces an inhibitory effect where one unit of ATP cannot bind to one unit of JNK1 and activate JNK1. JIP1 was found to bind to JNK1 and JNK2, and did not bind to ERK and p38 enzymes. Hence, JIP1 inhibitors that target the JIP1 binding site of JNK1 are postulated to have greater selectivity and specificity and fewer side-effects[61]. Currently, there are at least 191 compounds

targeting the JIP1 binding site of JNK1 with published biological activity data since the first small molecule inhibitor targeting this site was published in 2008[62] and one X-ray crystal structure of the truncated JIP1 structure in complex with JNK1[57].

The JIP binding site consists of 9 hydrophobic residues (Ala 113, Leu 115, Val 118, Met 121, Leu 123, Leu 131, Val 159, Val 323 and Trp 324), 3 polar residues (Tyr 130, Ser 161 and Cys 163) and 3 polar and charged residues (Glu 126, Arg 127 and Glu 329). Inhibitors which bind at this site are Type III inhibitors and may bind at any subsite that JIP binds to on JNK which is away from the ATP binding site. The binding site is shallow and is exposed to solvent to one side. Due to the polar and charged residue Arg 127 having a potential effect on conferring selectivity to ligands[63] and the novelty of exploring compounds which can bind to this subsite, this research project will attempt to screen and identify potential compounds from the ZINC database that can bind to the selected subsite using ligand and structure-based methods.

### **1.2.3 DFG binding site**

The DFG binding site has been identified from studies of the p38 enzyme that has the same enzyme classification number of JNK and fulfills a different role in metabolic reactions. The DFG binding site consists of a hydrophobic pocket adjacent to the ATP binding site. This binding site has an aspartic acid,

phenylalanine and glycine amino acid residue which make up the DFG motif where the phenylalanine residue can adopt a DFG-in conformation and a DFG-out conformation. Different kinase enzymes have different capacity to accept and bind to DFG site inhibitors at the DFG-in or DFG-out conformation. For binding of small-molecule inhibitors to occur favourably, the DFG binding site in p38 and JNK has to be in the DFG-out conformation, where the phenylalanine residue partially occupies the ATP site, has one face that helps to shield the DFG-out inhibitor, while the other face is exposed to solvent. However, DFG adopts this conformation only rarely[64]. Furthermore, the X-ray crystal structure of the compound BIRB798 with the JNK enzyme was only produced in 2010 (PDB Entry 3NPC). Hence, these factors led to the relatively lack of known DFG inhibitors. Inhibitors that bind at this site will block the ATP binding site by inducing the DFG-out conformation. This form of partial inhibition of JNK is more desirable than the Type I inhibition mechanism by having greater selectivity and specificity and fewer side-effects. There is currently a large knowledge gap in this area as only one compound has been found to bind to this binding site in JNK2 and one X-ray crystal structure of the compound at the DFG binding site of JNK2[65].

The DFG binding site consists of 12 hydrophobic residues (Ile 32, Val 40, Ala 53, Leu 76, Leu 77, Ile 86, Met 108, Leu 110, Met 111, Leu 142, Leu 168 and Phe 170), 1 polar residue (Gln 37) and 6 polar and charged residues (Lys 55, Arg 69, Arg 72, Glu 73, Glu 109 and Asp 169). Inhibitors which bind at this site are

Type II inhibitors and occupies a deep hydrophobic pocket away from the ATP binding site to produce the DFG out conformation, as well as a section of the ATP binding site at the hinge region. Hence, this type of inhibition can be considered to be competitive by strict definition, and deep pocket binders due to the deep-pocket interactions where DFG adopts which is away from the hinge region which ATP occupies[66]. Figure 1.1 shows each binding site in JNK.

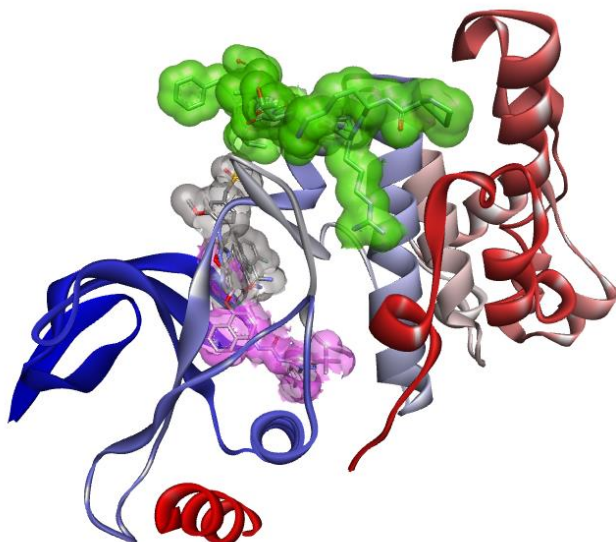


Figure 1.1: Ribbon diagram of JNK with stick and surface representation of ligand in each binding pocket. Amino terminal domains are coloured in blue and carboxyl terminal domains are coloured in red. Bis-anilino-pyrrolopyrimidine inhibitor in ATP binding pocket shown as grey stick and surface (PDB ID: 3ELJ). Truncated JIP in JIP binding pocket shown as green stick and surface (PDB ID: 1UKH). BIRB796 in DFG binding pocket shown as pink stick and surface (PDB ID: 3NPC).

### 1.3 Physiology of PXR

Human PXR was discovered in 1998 and was termed an orphan nuclear receptor[20] and nuclear receptor subfamily 1 group I member 2 (NR1I2)[67]. Found mainly in liver and intestinal cells, PXR have important roles in maintaining homeostasis and metabolism of bile acids, glucose, lipids, drugs, endocrine hormones and xenobiotics with another nuclear receptor Constitutive Androstane Receptor(CAR)[68]. Expression of PXR and CAR are activated by Glucocorticoid Receptor (GR)[69].

PXR forms a heterodimer complex with RXR through 21 amino acid residues[70]. Activation of PXR and formation of this complex with RXR regulates expression of genes that are involved in drug and xenobiotic sensing, metabolism and transport which include cytochrome P-450 family 3 subfamily A polypeptide 4 (CYP3A4) genes, where approximately 50% of marketed prescription drugs are metabolized by CYP3A4[71-73]. Hence, expression of CYP3A4 genes are regulated by GR, PXR, RXR and CAR directly or indirectly[74].

PXR activity consists of basal activity which is ligand-independent and ligand-dependent activity[75]. Ligand-dependent activity is influenced by the presence of ligands which bind to the ligand-binding domain (LBD) of PXR where different ligands may exert an agonist[74], antagonist[76] or partial

agonist[77] effect on the nuclear receptor. PXR is present in many species and may exhibit differences in selectivity of small molecule compounds in its ligand-binding domain[78].

#### **1.4 Current drug discovery efforts to identify JNK inhibitors and PXR activators**

Known JNK inhibitors have limited structural diversity. This is partly due to the fact that most are ATP-competitive inhibitors and the ATP binding pocket is highly conserved. Another reason for the low structural diversity is that most existing JNK inhibitors were discovered through high-throughput screening (HTS) followed by lead optimization using traditional medicinal chemistry approach. This is a time-consuming process and usually only evaluate a limited amount of structural diversity in each study. High-throughput screening (HTS) can improve the speed of searching for novel JNK inhibitors but is not practical for screening large chemical libraries with millions of compounds which would be more expensive. Furthermore, recent studies show that fewer drugs have entered the market with fewer compounds passing clinical trials, greater difficulty in obtaining approval from regulators and a high dropout rate[79,80].

Among the JNK inhibitors that were found to target the JIP binding site, they were discovered through HTS. In a HTS assay that was carried out to discover the first JIP-site inhibitor, it screened a total of 30000 compounds at a



rate of 20 compounds per assay. From the first lead compound BI-78D3 that was discovered as a hit, it was tested in murine models to assess safety and efficacy[62]. The subsequent lead and structural modifications would create a list of one to three classes of compounds[81-84].

Issues of compound selectivity and safety remain mainly for competitive inhibitors targeting the ATP binding site. This is because it is difficult to design small-molecule inhibitors that would only inhibit one isoform of JNK due to the highly conserved ATP binding site among all three JNK isoforms and in other protein kinases. Hence, the use of these JNK inhibitors beyond *in vitro* assays remain limited[10,11]. Inhibitors that target the JIP binding site and DFG binding site are less likely to bind to other kinase enzymes in a panel of related and unrelated kinase screening tests.

For PXR activity, there are studies being carried out to determine the type of ligand-dependent activity on the nuclear receptor as it could potentially be used as a model to assess and predict drug-drug interactions as well as effects of xenobiotics such as environmental pollutants and toxins occurring within human cells and the human body[17,76,85]. There are also studies being carried out to determine the differences in PXR shape and structure among different species, and how these differences affect interactions of PXR with various drug and chemical compounds[78,86].

## **1.5 Computational methods to screen and validate JNK inhibitors and PXR activators**

To screen large chemical libraries that contain millions of compounds to test for inhibitory activity on JNK in cell and animal models would be expensive and time-consuming. It is also expensive and time-consuming to conduct *in vitro* and *in vivo* tests for CYP3A4 and PXR activity. Computational methods are a fast and cost-effective method for complementing traditional drug development methods. With the introduction of computers with higher processing capacity, together with X-ray crystallography or nuclear magnetic resonance techniques that enable the 3D structures of entire proteins to be identified and recorded into a computer database, computational methods can identify the essential features required for activity by analyzing existing drug target binders and using this information to prioritize compounds for *in vitro* testing. This will help to improve the hit rate of HTS and to form better prediction models to improve safety profile of drugs, thereby reducing the cost of finding novel inhibitors or making useful lead modifications to existing drugs or chemical compounds.

There are two main approaches for computer-based methods: ligand-based and structure-based. Ligand-based methods can be used when there are at least a certain number of active and inactive compounds against a biological target. These compounds are used to derive a quantitative structure-activity relationship

(QSAR) model or a pharmacophore model, which can then be used for virtual screening of large chemical libraries for potential active compounds.

In structure-based methods, molecular docking and molecular modeling which includes building homology models, screening with target-based pharmacophore models and molecular dynamics simulations are used to dock and refine the binding orientation of the compound. A method to improve the screening efficiency of structure-based methods is to use a ligand-based method to perform a preliminary screening of large chemical libraries and reduce the list of potential inhibitors to a more manageable number for structure-based studies[87-92].

### **1.5.1 Ligand-based methods**

#### Quantitative structure-activity relationship (QSAR)

The QSAR approach attempts to identify and quantify various properties of different compounds using 1D, 2D, 3D chemical descriptors or fingerprints of the compounds to see whether any of these properties have any effect on biological activity. The underlying assumption in QSAR is that similar molecules should exhibit similar binding properties with respect to a given target. If there is a relationship or correlation between any of the descriptors and the compound's biological activity, an equation can be drawn up to quantify the relationship and

hence allow some level of prediction to be made for the biological activity of a new and structurally different drug compound within a chemical library or a related analogue. Since the first QSAR model was published[93], it have been used routinely in the early stages of drug design to screen large chemical libraries prior to *in vitro* testing. QSAR has the advantage of being able to screen through large chemical libraries very rapidly as it only looks for desirable 1D, 2D, 3D descriptors and structural fingerprints of compounds that have been reported in the literature to be a positive or negative compound. However, this method is not able to determine the binding conformation of the compound in the target site.

#### Ligand-based pharmacophore modeling

Ligand-based pharmacophore modeling is typically performed by extracting common chemical features from 3-dimensional structures of known ligands which show essential ligand-macromolecule interactions[94]. The common chemical features or pharmacophore elements that are usually used as types of the desired interactions are hydrogen-bond interactions, hydrophobic interactions and ionic interactions. A pharmacophore is based on the concept of similarity between ligands and is used in virtual screening to explore diverse chemical compounds in large chemical databases for the purpose of identifying novel structural hits.

## Current studies using ligand-based methods for JNK inhibitors and PXR activators

Ligand-based approaches have been used to develop 2-dimensional (2D) and 3-dimensional (3D) QSAR models for JNK inhibitors[95-98]. The models were used to propose modifications for improving the inhibitory activities of existing JNK inhibitors. There are several QSAR models that have been built on compounds that target the ATP-binding site, with an overall prediction accuracy that is greater than 90%[95], a Matthew's Correlation Coefficient that is greater than 0.7[96], a conventional or cross-validated  $r^2$  that is greater than 0.9[97,98]. The disadvantage in using these models are that they are used on compounds which target the ATP binding site, which may lead to potential selectivity and specificity problems. There is currently no QSAR model for inhibitors that bind to the JIP binding pocket.

Ligand-based approaches that have been used to predict human PXR activators are generally accurate with an overall prediction accuracy that is greater than 60%[99,100].

## 1.5.2 Structure-based methods

### Target-based pharmacophore modeling

Today, structure-based methods in pharmacophore modeling have gained significant interest in the past 5 years due to the increase in number of protein-ligand crystal structures. 3D pharmacophore modeling is a well-established *in silico* technique which provide benefits for early drug design. Pharmacophore models visualize chemical features that are supposed to be important for protein-ligand interactions and therefore for biological activity. They aim to be complementary to docking procedures, by providing information on possible docked poses and require less computational resources than random docking alone without pharmacophore models[88,91,101].

### Molecular docking

Docking is a structure-based method where computer software is used to assess the complementarity of a ligand to a defined binding site on the drug target. The docking program attempts to explore and obtain binding orientations and conformations of the ligand within the defined binding site to try to get the best fit.

Docking may also include calculation of all possible binding interactions that may occur within the ligand itself or specified amino acid residues within the entire protein, binding interactions between the ligand and the amino acid residues within the defined binding site and binding interactions between ligand or amino acid residues with X-ray crystallographic or free solvent molecules. Docking may also include optimization of the fit between the ligand and the defined binding site by modeling flexibility of the surrounding amino acid residues within the defined binding pocket and modeling movements of the ligand and solvent molecules within the docked pose.

Simpler methods of docking involve overlaying possible ligand conformers onto solvent accessible surfaces or Connolly surfaces within the defined binding pocket without calculating any possible steric clashes or binding interactions. More advanced and computationally demanding methods of docking involve finding as many possible poses of the ligand that is predicted to occur within the defined binding site and is assigned a docking score to each predicted pose that is parameterized using data derived from published X-ray structures. This scoring function is derived by assigning a certain number of weights and penalties to various different factors that are visible from the experimentally determined structures and uses this mathematical model to rank and predict the docked compound and its pose as having strong, weak or no binding affinity as a surrogate for the binding energy between the ligand and the defined target site.

There are two problems associated with molecular docking used in structure-based approaches. The first is the identification of accurate poses of the ligand in the active site of the protein. Generally, proteins with larger active sites, flexible and poorly defined binding sites due to hinge region and chemical compounds with a greater number of single bonds that are rotatable are more likely to lead to multiple predicted binding modes. This may lead to false predictions during virtual screening due to the compound having multiple favourable interaction profiles and multiple highly ranked poses.

Another problem is the assigned rank and docking score of compounds often have a poor correlation with the actual biological activity of the compound. This is due to multiple factors such as the simplistic sampling methods which are often based on a rigid frozen snapshot of the entire protein molecule and its active site, the use of an empirical scoring function, desolvation, conformational energy penalties, choice of force fields being used and removal of solvent molecules that may play an important role in the binding of the compound[102-105].

### Molecular dynamics (MD)

Molecular dynamics simulations, in which entire protein-ligand systems are allowed to interact for a period of time based on Newton's equations of motion, are widely used computational techniques for the study of protein macromolecules and small chemical compounds in solvent medium[106,107].



MD is very useful for understanding the dynamic behavior of the entire protein-ligand system that is modeled in a solvent medium at a certain temperature and pressure, which include internal motions, conformational changes or even protein-folding processes. Due to limitations from virtual screening methods in docking and pharmacophore modeling where the ranking correlation of the compounds with published biological activity is poor and inaccurate, molecular dynamics is carried out after docking with pharmacophore constraints where the reduced number of poses for each compound is more manageable for the molecular dynamics study.

The results from MD simulations can be used to refine models of protein-ligand systems obtained from docking. Such simulations display flexibility of the entire model of the protein-ligand system, and show its relative duration and likelihood of binding interactions with respect to time and thus coming closer to modeling induced-fit effects more accurately. Incorrectly docked structures and poses generated from docking have a higher chance of generating unstable MD trajectories leading to the disruption of the complex, which provides an additional mechanism to filter for false positives. MD simulations usually include explicit solvent molecules-often water with salts as counterions within the model, which is very important to understand the role of the particular solvent and its effect on the stability of the protein-ligand complexes[88]. In this research project, MD simulations will be carried out on selected poses obtained from docking with pharmacophore constraints.

## Current studies using structure-based methods for JNK inhibitors and PXR activators

Structure-based approaches are currently used on reported JNK inhibitors that bind strongly to JNK after HTS is carried out[65,81-84], and in virtual screening of chemical databases[108]. The models are used to explain the binding pose and the type of binding interactions that occur between the compound and the enzyme. These approaches test a very small subset of compounds from chemical databases, hence there is a knowledge gap in identifying JNK inhibitors where there is room for novelty by having a less saturated chemical space to explore.

There are 9 available X-ray crystal structures of human PXR[75,86,109-113] and due to the highly flexible nature of the ligand-binding domain in human PXR[75,111,112,114], as well as the capacity of the nuclear receptor to bind to a very large variety of different chemical compounds of various molecular weights, there is considerable interest in studying PXR activity through structure-based methods[72,74,85,99,100,115-118].

## **1.6 Objectives**

### **1.6.1 Objective 1: Identify potent JNK inhibitors that target JIP binding site and DFG binding site**

In this project, the JIP binding site and DFG binding site were selected as the target sites for designing potent JNK inhibitors. The JIP binding site was selected because there are known inhibitors that target this site and this will facilitate the development of a QSAR model for screening large chemical libraries[119]. This will reduce the number of compounds that need to be screened using structure-based drug design. A search through patent databases also revealed entire protein macromolecules as the only registered patents that target the JIP active site. The DFG binding site was selected because there is only one compound that is known to bind to this site currently. Thus there is great potential to discover novel potent inhibitors that target these two sites.

The ATP binding site was not selected because of the following reasons. Firstly, there are more than 900 compounds that have been published to bind at this site, with many new inhibitors being discovered and published every year. Thus, there is less novelty for such inhibitors. Secondly, the ATP site is highly conserved and is present in many other protein kinases. These factors reduce the appeal of discovering novel inhibitors that bind at this site. For these reasons,

designing ATP-competitive inhibitors are less attractive and will not be carried out in the project[120].

#### JNK inhibitors targeting the JIP binding site

This project aims to fill the current research gap by developing a QSAR model for the JIP binding site. The QSAR model will address the issue of having a limited diversity in the dataset by including an additional 1228 compounds from the FDA Orange Book as putative negative compounds to train the QSAR model. This QSAR model will also include an additional algorithm that will define multiple thresholds for positive and negative compounds to automatically reject compounds for which the computational model is not able to make a prediction with a certain degree of confidence. This will define the applicability domain of the model to ensure that it is not used inappropriately. This QSAR model will also include external five-fold cross validation to validate the model more rigorously and include a DivEnsemble algorithm to select models to use in each consensus model.

In this research project, a QSAR model will be developed to screen the ZINC chemical library and used in combination with structure-based methods to identify potent inhibitors that target the JIP1 binding site of JNK1.

## JNK inhibitors targeting the DFG binding site

Currently, there is only one compound known to bind to the DFG binding site of JNK2 and p38 $\alpha$ . Therefore, there is a large gap in this area. The Type II mechanism of inhibition at this active site suggests that inhibitors that target this binding site would also be less likely to inhibit other kinase enzymes. Therefore, the significance of this objective is that new compound classes that target this site could be discovered.

In this research project, docking will be carried out to screen the ZINC chemical library to identify inhibitors that target the DFG binding site of JNK2. Identified compounds would be tested for biological activity by determining their dissociation constant values on two JNK isoforms and one p38 subtype.

### **1.6.2 Objective 2: Form an *in silico* model for human and zebrafish PXR activators**

Since PXR was discovered, there are studies to determine the type of compounds that can bind and exert an agonist, antagonist or partial agonist effect on the receptor[74,76,77,121]. This is because PXR have been shown to regulate several genes which encode drug-metabolizing enzymes and drug transporters[23]. Hence, there is interest in studying the mechanism of binding and activation in detail as this nuclear receptor may provide valuable insights into

the possible metabolism and removal of various xenobiotics including various drugs and environmental toxins.

In order to study the biological effect of various approved drugs on PXR activation, the expression of the gene Cytochrome P-450 (CYP)3A4 was chosen as the marker to investigate the amount of biological activity and PXR activation by collaborators at University of British Columbia at Vancouver and NUS. This is because CYP3A4 gene expression was shown to be upregulated upon PXR activation[122-125]. CYP3A4 gene expression levels are obtained from Firefly/Renilla luciferase assay on human liver cells or from *in vivo* study on zebrafish *Danio rerio* as it is able to perform a high-throughput screen compared to other larger animal models, less expensive to perform a study on and takes a relatively shorter time to obtain results.

This study will attempt to investigate the mechanism of binding of various approved drugs on the ligand-binding domain (LBD) of human PXR and zebrafish PXR, attempt to correlate its degree of binding with the selected marker for biological activity and attempt to improve on existing structure-based and molecular modeling methods to obtain a better correlation.

There is currently a knowledge gap in this area where there are no available 3D structures of zebrafish PXR and there are no structure-based models that can correlate zebrafish PXR activity with human PXR activity in the

literature. The significance of this objective is that developing a structure-based model to predict PXR activity and CYP3A4 gene expression more accurately in human and zebrafish will help to reduce time and costs involved with assays which use human liver cells. In addition to availability of biological activity data on 14 approved drugs and 1 chemical compound from collaborators, there are site-directed mutagenesis studies on the ligand-binding domain of human PXR in the literature which could help to elucidate the mechanism of binding of various ligands in the X-ray crystal structure of human PXR and the homology model of zebrafish PXR[85,86,112]. All these factors make the development of structure-based models for human and zebrafish PXR activity attractive.

As rifampicin was chosen as the positive validation compound and it was found to show differences in CYP3A4 gene expression in human and zebrafish[126], the study of PXR agonist activity in both human and zebrafish PXR will be carried out first in this project. There are also fewer biological activity data on zebrafish in the literature and ligand-based methods to investigate PXR activity will not be carried out in this project.

## Chapter 2

### Phase 1: Development of QSAR model for JIP binding site inhibitors

#### 2.1 Summary

A truncated form of JNK interacting protein 1 (JIP1) consisting of 11 residues 153-163 (RPKRPTTLNLF) has been found to inhibit JNK1 by reducing the amount of space for ATP to bind to JNK1[57,60]. A dataset of 191 positive and negative compounds are obtained from published sources and are used to build a QSAR model. Building a QSAR model has its advantages in the early stages of screening large chemical libraries when 3D structural information of the protein system is not required, and the model when trained using putative negative compounds to augment the dataset, helps to increase its applicability domain and is more likely to make more accurate predictions and predict more compounds that may be structurally diverse. This QSAR model is then applied on the ZINC chemical library to identify a set of compounds that could potentially bind to the JIP1 binding site of JNK1.



## 2.2 Methods

### 2.2.1 Dataset

A total of 191 compounds and their reported  $IC_{50}$  in a LanthaScreen™, AlphaScreen™ or kinase assay test for JNK1 inhibition (JIP site binders, Dual-action ATP-JIP site binders) were collected from published studies[37,62,81-84,127]. The compounds were then categorized into positive (JNK inhibitors) and negative (JNK non-inhibitors) compounds using a cutoff value of  $IC_{50}$  at 10  $\mu$ M.

A single cutoff value was chosen because a binary classification model will have problems in predicting compounds that have biological activity that lie between two different  $IC_{50}$  cutoff values where most models do not have a valid prediction option to classify these compounds. Also, if these compounds were removed and not included in the dataset, it would reduce the number of compounds in the training set to train the model and might potentially impact the performance of the model.

An additional 1228 compounds from FDA Orange Book which were not known to inhibit JNK1 were added to the total dataset as putative negative compounds because this had been shown to reduce the number of false positives[128]. The total dataset of 1418 compounds were split into the training set data and external validation set data by a stratified sampling method in an 80

to 20 ratio to ensure that the number of positive compounds and negative compounds were adequately represented in both datasets. The training set was used to develop the QSAR model. The validation set was used to evaluate the performance of the QSAR model and was not used during the development of the model. To rigorously validate the final consensus model, we adopted the recommended rigorous validation approach[129]. This is done by repeating the entire model development process five times using different training and validation sets. The following flowchart in Figure 2.1 describes the gathering and splitting of the dataset to form the training set and validation set used in the study.

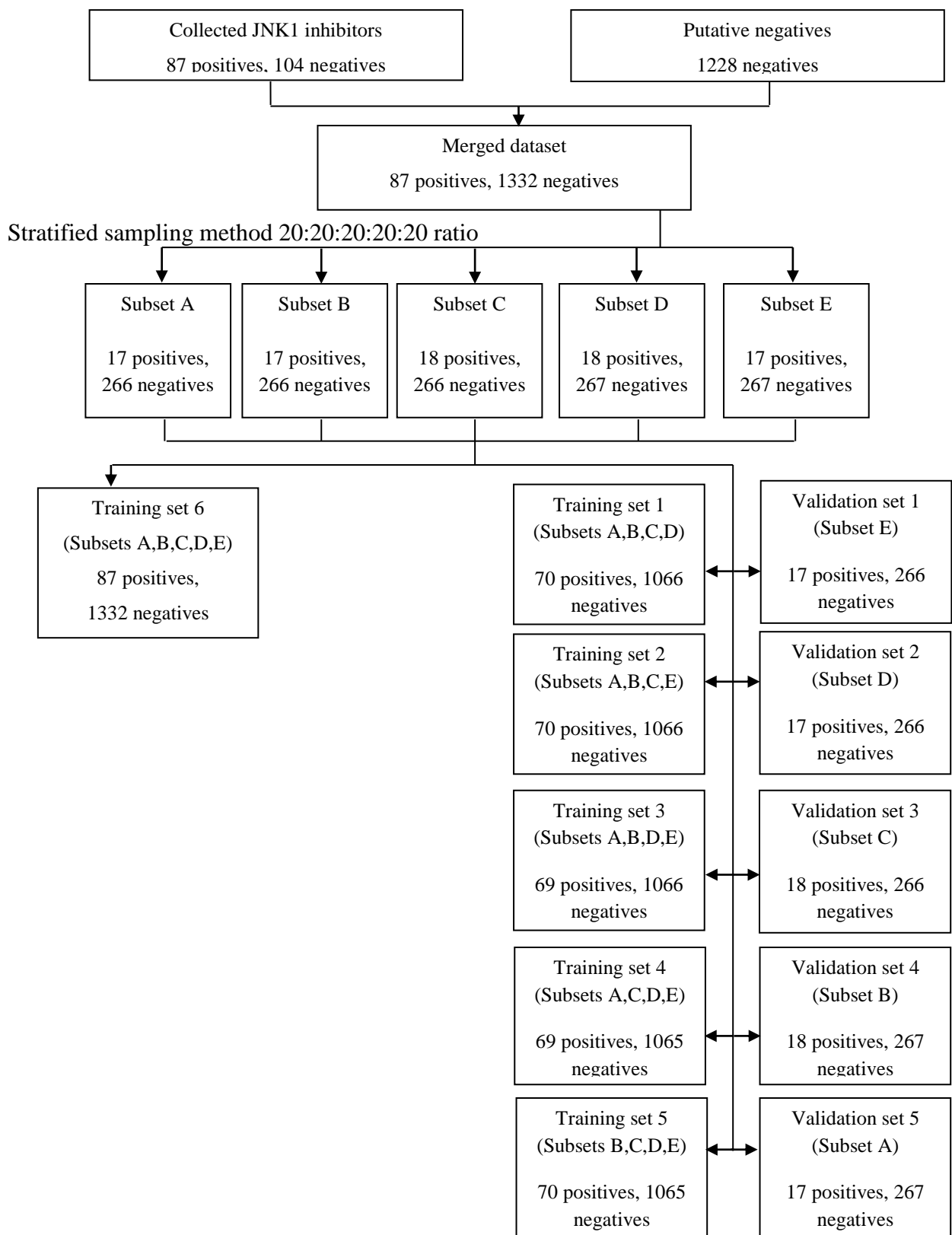


Figure 2.1: Flowchart of steps taken to obtain the training and validation set.

### 2.2.2 Calculation of descriptors

All the compounds were drawn using ChemDraw Pro 12.0[130]. A total of 729 1D and 2D descriptors were calculated by PaDEL-Descriptor 2.12[131]. 3D descriptors were not included in the calculation because in some previous QSAR studies, it has been shown that including 3D descriptors in a QSAR model did not necessarily improve the performance and prediction ability of such a QSAR model[132,133]. The descriptors were normalized using range transformation to a maximum value of 1 and a minimum value of 0. This is to ensure that all descriptors have equal potential to affect the QSAR model.

Irrelevant descriptors, such as constant descriptors were removed. Genetic algorithm (GA) was then used to remove the remaining irrelevant and redundant descriptors. A population size of 50 and number of generations of 100 was used for the GA descriptor selection process. During the GA descriptor selection process, QSAR models developed using different descriptor subsets were evaluated using five-fold cross-validation. The GA descriptor selection was repeated 30 times for each modeling method to generate 30 QSAR models for each modeling method.

### 2.2.3 QSAR model development

Two different modeling methods, Naïve Bayes (NB) [134,135] and support vector machine (SVM) [136,137] were used to develop the QSAR models. For SVM, the cost value is kept constant at a high value of 100000 to simulate a hard-margin SVM. The Gaussian radial basis function kernel, which is commonly used for QSAR studies, was used as the kernel function in this study since it is able to handle non-linear relationships between class labels and descriptor values. The optimum  $\gamma$  value was determined by experimenting with the empirical value of 1 divided by the number of examples in the training set and  $2^{-5}$ ,  $2^{-7}$ ,  $2^{-9}$ ..... $2^{-15}$ .

### 2.2.4 Defining the applicability domain with multiple thresholds

An important aspect of QSAR modeling is the definition of a domain of applicability for the model. In this study, we adopted the multiple threshold algorithm proposed by Fumera et al[138] to define the applicability domain for the QSAR models.

The algorithm uses a binary class-related reject threshold (CRT) rule where for a classification task with  $N$  data classes that are characterized by estimated a posteriori probabilities  $P(\omega_k | x)$ ,  $k = 1, \dots, N$ .

A pattern  $x$  is out of the applicability domain of the model if:

$$\max_{k=1, \dots, N} P(\omega_k | x) = P(\omega_k | x) < T_i \quad (1)$$

while it is within applicability domain of the model and assigned to the class  $\omega_i$ ,  
if

$$\max_{k=1, \dots, N} P(\omega_k | x) = P(\omega_k | x) \geq T_i \quad (2)$$

The CRTs take on values in the range [0,1]. The determination of the optimum threshold values was determined by minimizing the theoretical accuracy of the rejected compounds based on five-fold cross-validation[139,140].

### 2.2.5 Consensus model

Consensus modeling is carried out to build the final QSAR model. This is to enable the many different models, each with similar measures of performance, but different set of descriptors, to be used as an ensemble classifier to better predict unknown compounds during the actual screening process and to take into account random variations in each base model that might have certain descriptors being used, and other descriptors being left out of the model building process. This will help to improve the classification accuracy[141,142]. In this study, five models were selected using a DivEnsemble method to obtain the optimum

classification performance[139]. A compound is considered to be within the applicability domain of the consensus model if at least one base model considered the compound to be within its applicability domain and there are no ties in the predictions.

### 2.2.6 QSAR model validation

The performance of the QSAR models in the GA descriptor selection process were assessed by computing true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) from the cross-validated models, which were then used to calculate the Matthew's Correlation Coefficient (MCC)

$$= \frac{TP \times TN - FN \times FP}{\sqrt{(TP+FN)(TP+FP)(TN+FN)(TN+FP)}}.$$

The MCC value was used to compare the performance of the models developed using different descriptor subsets and to identify best descriptor subset for each GA descriptor selection run.

The consensus model developed using five selected performing models was assessed using the rigorous validation approach. Other than MCC, the

$$\text{sensitivity (SE)} = \frac{TP}{TP+FN} \times 100\% \quad , \quad \text{specificity (SP)} = \frac{TN}{TN+FP} \times 100\% \quad , \quad \text{overall}$$

$$\text{prediction accuracy (Q)} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad \text{and Geometric mean (GM)} =$$

$\sqrt{(SE)(SP)}$  were also calculated.

## 2.3 Results and discussion

### 2.3.1 Validation of the consensus model

The consensus model was formed from three SVM models and two NB models. The performance of the consensus model assessed using the rigorous validation approach are shown in Table 2.1.

Table 2.1: Performance of consensus model on each validation set and on the entire dataset.

Dataset	TP	FP	TN	FN	Outside AD	SE	SP	Q	GM	MCC
Validation set 1	9	3	260	5	3 positive, 3 negative	0.643	0.989	0.971	0.797	0.680
Validation set 2	11	6	260	5	1 positive	0.688	0.977	0.961	0.820	0.646
Validation set 3	9	8	249	6	3 positive, 9 negative	0.600	0.969	0.949	0.763	0.537
Validation set 4	9	9	256	5	4 positive, 2 negative	0.643	0.966	0.950	0.788	0.541
Validation set 5	10	3	263	3	4 positive, 1 negative	0.769	0.989	0.979	0.872	0.758
Average	N/A	N/A	N/A	N/A	N/A	0.669	0.978	0.962	0.808	0.632
Entire dataset	74	10	1319	11	2 positive, 3 negative	0.871	0.993	0.985	0.930	0.868

None of the 10 false positives were from the putative negative dataset. Although this may suggest that the consensus model may just be differentiating between the compounds with IC<sub>50</sub> values and putative negatives rather than



learning the actual difference between positive and negative compounds, we do not believe this is true. This is because the 10 false positives were just approximately 10% of the total true negatives. Thus the majority of true negatives were still accurately predicted by the model. In addition, as we will show using structural similarity and distribution below, when the consensus model was used on a large chemical library, the predicted actives include compounds that were structurally different from the positive training compounds. This suggests that the model is truly discriminating between positive and negative compound and not merely discriminating between compounds with  $IC_{50}$  values and putative negatives. Among the false negatives, they have  $IC_{50}$  values of between 0.4  $\mu M$  and 10  $\mu M$ . This indicates that the false negative predictions are mainly due to the compounds being difficult to classify as positive or negative compounds due to their  $IC_{50}$  values that are near to the defined threshold cutoff values.

### **2.3.2 Screening of ZINC chemical library using QSAR model**

The results from using the consensus model to screen the ZINC chemical library (Downloaded in March 2010) are shown in Table 2.2[143,144].

Table 2.2: Virtual screening of ZINC library using consensus model. Numbers in parenthesis refer to the percentage of compounds in their respective columns to the total number of compounds in the ZINC chemical library.

<b>Total number of compounds</b>	<b>Number of compounds predicted as positive</b>	<b>Number of compounds predicted as negative</b>	<b>Number of compounds outside AD</b>
17833934 (100.0)	45673 (0.256)	17409760 (97.622)	378501 (2.122)

The number of identified potential JIP1-binding inhibitors is manageable for the subsequent docking study.

### 2.3.3 Structural similarity and distribution

Structural similarity of the compounds in the dataset and in the ZINC compound database is calculated by a fingerprint, FP2 in Open Babel[145]. The fingerprints of positive compounds in the dataset, is compared to other compounds by using the Tanimoto coefficient, the number of bits in common divided by the union of the bits set.

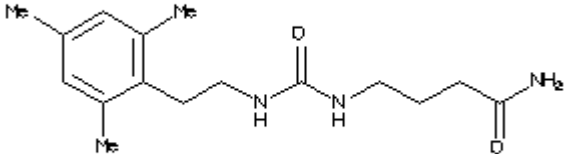
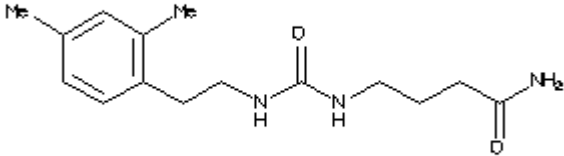
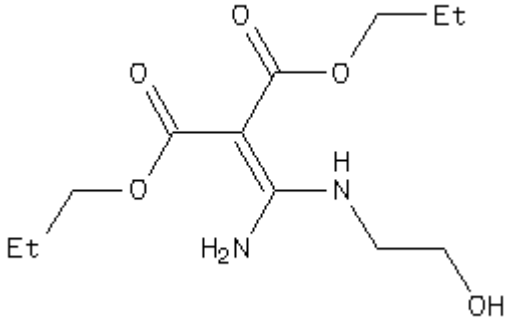
The data set is more diverse when the diversity index, which is the average value of the Tanimoto coefficients between pairs of compounds in the dataset, approaches 0. Table 2.3 below lists the Tanimoto coefficients and diversity index of each list of compounds.

Table 2.3: Structural diversity of compounds.

<b>Type of dataset</b>	<b>Lowest average Tanimoto coefficient</b>	<b>Highest average Tanimoto coefficient</b>	<b>Diversity Index</b>
JIP site compounds (n = 191)	0.0975	0.892	0.334
JIP site compounds and putative negatives (n = 1318)	0.0065	0.355	0.136
ZINC compounds picked out as hits (n = 45673)	0.0223	0.792	0.201

The average lowest pairwise Tanimoto coefficient for the ZINC compounds that were identified as potential actives compared with all the positive compounds used in training the model was 0.0223. This suggests that our consensus model presented in this work is able to identify novel potential inhibitors that are structurally different from currently known inhibitors. Three of the most dissimilar compounds are presented in Table 2.4.

Table 2.4: A selection of ZINC compounds not reported as JNK inhibitors that are the most dissimilar from the positive compound BI-78D3.

Compound name	Structure	Tanimoto coefficient
ZINC71819697		0.07
ZINC71831223		0.07
ZINC02854383		0.076

### 2.3.4 Molecular descriptors

A total of 168 descriptors, which are listed in Table 2.5, were included in the consensus model to screen through the ZINC chemical library. Details of the descriptors can be found in the journal article for PaDEL-Descriptor[131].

Statistical analysis of the molecular descriptors between inhibitors and non-inhibitors showed that 17 descriptors had statistically significant different mean values between inhibitors and non-inhibitors. Inhibitors have higher values in terms of hydrogen-bond donors and acceptors, and internal hydrogen bonds (nHBd, SHBa, nHBint4 and minHBInt6), molecular electrotopological variation (DELS), electronegativity of the extended topochemical atom (ETA\_Epsilon\_2) and Petitjean Number. On the other hand, non-inhibitors have higher Ghose-Crippen molar refractivity (AMR), Broto-Moreau autocorrelation of the topological structure weighted by polarizability (ATSp1 and ATSp3), valence chi path cluster (VPC-6), valence chi path (VP-3), sum of weak hydrogen bond acceptors (SwHBa), sum of the electronic state of aromatic bonds attached to a CH group (SaaCH), sum of the alpha values of all non-hydrogen vertices that is connected to only one other non-hydrogen vertex(ETA\_AlphaP) and number of six-membered rings (n6Ring and nT6Ring)[146-151]. Some of these features are consistent with the report by Heo et al[57].

Table 2.5: Descriptors used in the consensus model.

Descriptor type	Number	Descriptor
ALOGP	1	AMR
APol	1	apol
Atom count	5	nAtom, nHeavyAtom, nF, nP, nI
Autocorrelation (polarizability)	3	ATSp1, ATSp3, ATSp5
BCUT	2	BCUTc-1l, BCUTp-1h
Bond count	1	nBondsD2
Bpol	1	bpol
Carbon types	3	C2SP2, C3SP2, C4SP3
Chi chain	2	VCH-3, VCH-5
Chi path cluster	3	VPC-4, VPC-5, VPC-6
Chi path	7	SP-0, SP-3, SP-4, SP-6, VP-0, VP-1, VP-3
Crippen logP and MR	1	CrippenMR
Atom type electrotopological state	91	nHBd, nwHBd, nwHBa, nHBint4, nHBint5, nHBint7, nHsSH, nHssNH2p, nHsssNHp, nHdCH2, nHdsCH, nHCHnX, nHAvin, nsCH3, ndsCH, ndssC, naasC, nssssC, nsNH3p, nssNH2p, ndNH, nssNH, ntN, nsssn, nsOH, nsOm, nsSH, nsBr, SHBa, SwHBa, SHBint2, SHBint7, SHBint8, SHdNH, SHssNH, SHsNH3p, SHdsCH, SHCHnX, SHCsats, SHOther, SaaCH, StN, SdsN, SsssN, SaasN, SdO, SssO, SdS, SaaS, minHBa, minHBint4, minHBint6, minHsOH, minHsNH3p, minHsssNHp, minHtCH, minHdsCH, minHAvin, mintCH, mindsCH, mindssC, minsNH2, minssNH2p, minaaNH, mintN, minsssNHp, minssO, minsOm, minsF, minsSH, mindS, maxHBa, maxHBint8, maxHdNH, maxHsSH, maxHssNH, maxHaaNH, maxHtCH, maxHAvin, maxtCH, maxsNH2, maxdNH, maxtN, maxssssNp, maxsOm, maxsl, suml, gmin, MAXDP, DELS, MAXDN2
Extended topochemical atom	15	ETA_Alpha, ETA_AlphaP, ETA_Epsilon_1, ETA_Epsilon_2, ETA_Epsilon_3, ETA_Epsilon_5, ETA_dEpsilon_B, ETA_Shape_P, ETA_BetaP, ETA_Beta_ns, ETA_dBeta, ETA_Eta_R, ETA_Eta_F, ETA_EtaP_L, ETA_EtaP_B
Hbond acceptor count	1	nHBacc_Lipinski
Kappa shape indices	1	Kier1
Largest chain	1	nAtomLC
Largest pi system	1	nAtomP
Longest aliphatic chain	1	nAtomLAC
Molecular distance edge	7	MDEC-11, MDEC-13, MDEC-33, MDEC-44, MDEO-11, MDEN-11, MDEN-12
Molecular linear free energy relation	5	MLFER_BH, MLFER_BO, MLFER_E, MLFER_L
Petitjean number	1	PetitjeanNumber

Ring count	10	n3Ring, n6Ring, nG12Ring, nF7Ring, nF8Ring, nT5Ring, nT6Ring, nT7Ring, nT8Ring, nT12Ring
Weighted path	3	WTPT-1, WTPT-2, WTPT-5
Zagreb index	1	Zagreb

## 2.4 Conclusion

From a dataset of 191 JNK1-JIP1 binding site inhibitors and non-inhibitors and 1228 putative negative compounds, a consensus model that consists of three SVM models and two NB models was developed. This consensus model was used to screen the ZINC chemical library that contains 17833934 different compounds. A total of 45673 compounds were identified as positive compounds based on the consensus model. The identified positive compounds have a high degree of diversity from the tested positive compounds and hence there is more potential to identify novel compounds that are structurally different from tested compounds. The next phase of the research project will use structure-based methods to refine the model and prioritize compounds for *in vitro* testing.

## **Chapter 3**

### **Phase 2: Development of structure-based model for JIP binding site inhibitors**

#### **3.1 Summary**

The truncated form of JNK interacting protein 1(JIP1) has been found to inhibit JNK1 by reducing the amount of space for ATP to bind to JNK1[57,60]. A dataset of 45673 ZINC database compounds have been identified by the QSAR model to be possible inhibitors that target this binding site. The model is then refined using structure-based methods which include docking, pharmacophore modeling and molecular dynamics to prioritize compounds for *in vitro* testing.

#### **3.2 Methods**

##### **3.2.1 Preparation of JNK1 structure**

The X-ray crystal structure of JNK1 together with the truncated version of JIP1 which was discovered to be necessary for inhibition of JNK1 by Barr et al[60] was available in the RCSB Protein Databank by Heo et al(PDB ID: 1UKH)[56,57]. Figures 3.1 and 3.2 show JIP1 in complex with JNK1. The JNK1 structure was preprocessed by using the structure preparation wizard which



includes adding hydrogen atoms to the entire protein, assigning a cap at the terminal ends of the protein and assigning protons to the entire structure[152]. From the X-ray crystal structure, the following residues are thought to be important in binding and inhibitory activity: Arg 127 and Glu 329.

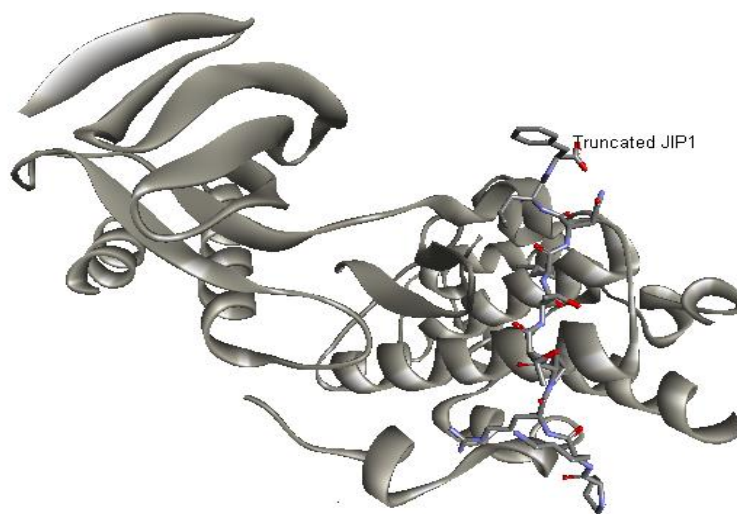


Figure 3.1: Ribbon structure of JNK1 with truncated JIP1 (PDB ID: 1UKH).

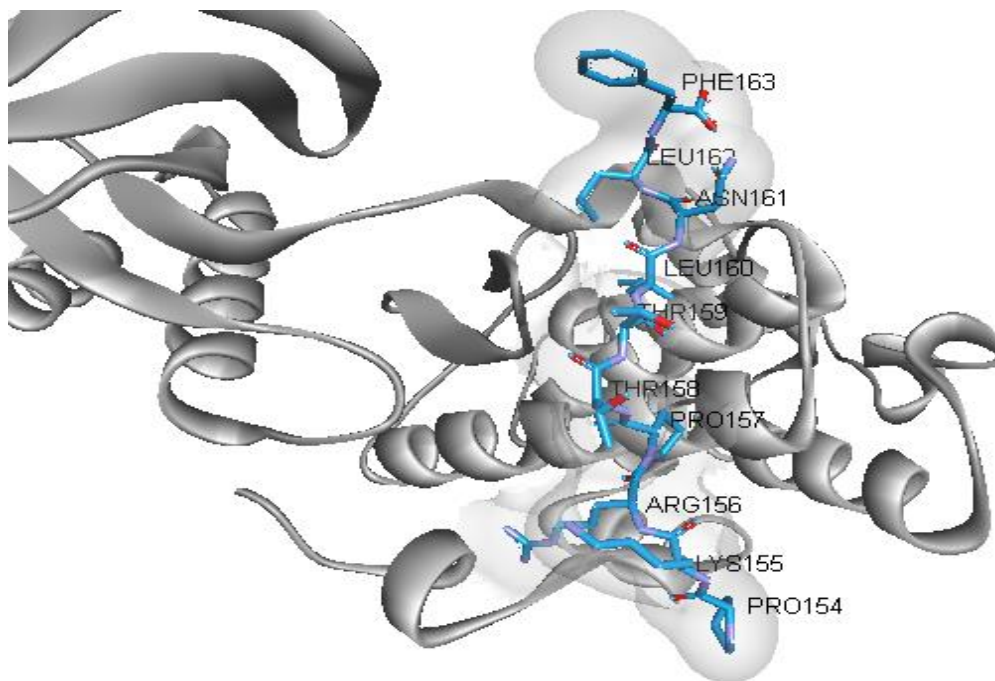


Fig 3.2: Close-up of truncated JIP1 in binding pocket of JNK1. Residues of JIP1 are labelled.

The JIP1 binding site consists of 9 hydrophobic residues (Ala 113, Leu 115, Val 118, Met 121, Leu 123, Leu 131, Val 159, Val 323 and Trp 324), 3 polar residues (Tyr 130, Ser 161 and Cys 163) and 3 polar and charged residues (Glu 126, Arg 127 and Glu 329) (Figures 3.3 and 3.4).

### 3.2.2 Virtual screening for potential JNK1-JIP1 inhibitors

Virtual screening of the 45673 compounds identified by the QSAR model was carried out using rigid docking methods in MOE 2012.10. A pharmacophore model was developed using MOE 2012.10[153] to improve the quality of the docked poses. Three interactions were chosen to include in the pharmacophore model in Table 3.1 and Figure 3.3.

Table 3.1: List of pharmacophore points selected in model.

<b>Name of residue(s) in JNK1</b>	<b>Name of residue in JIP1</b>	<b>Type of interaction</b>
Arg 127	Thr 159	Hydrogen bond
Ser 161	Leu 160	Hydrogen bond
Val 118, Leu 123, Leu 131, Val 159, Cys 163	Leu 160	Hydrophobic

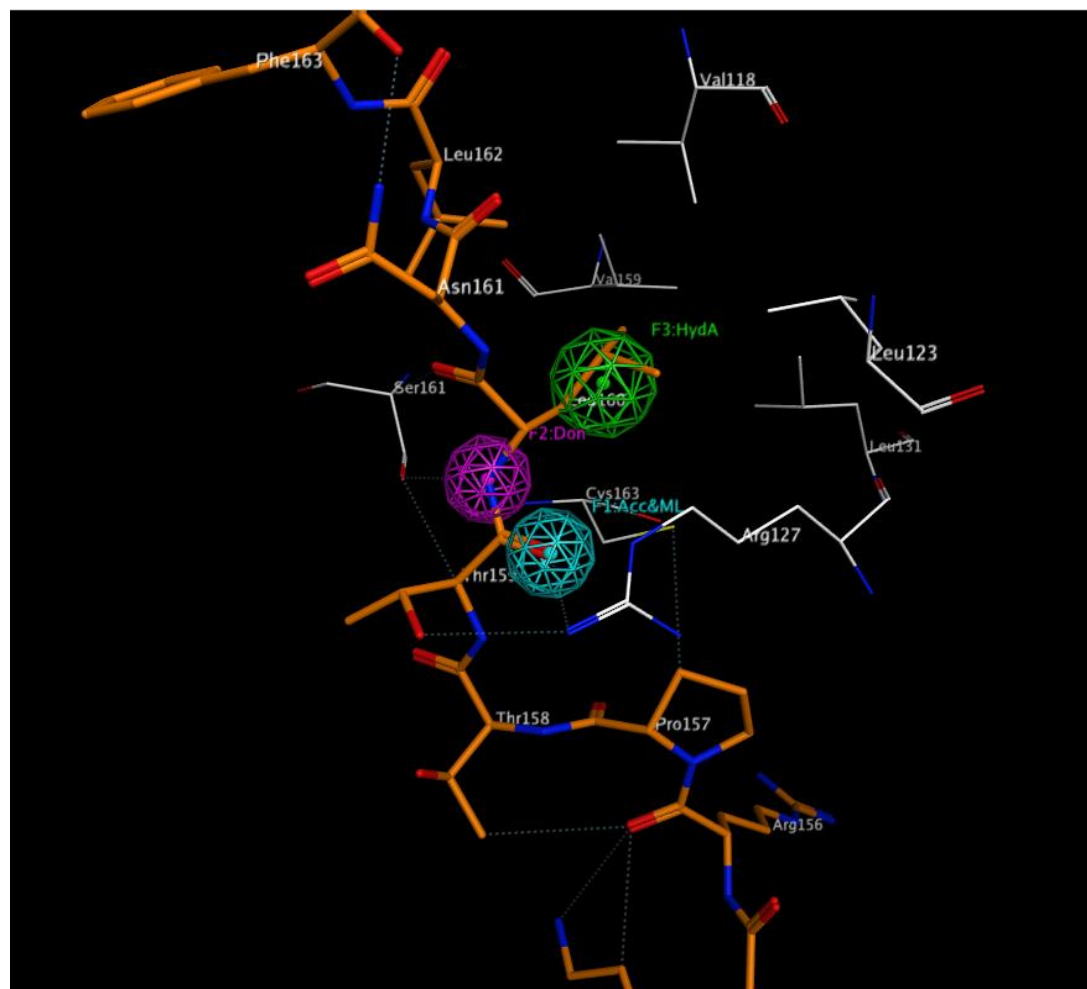


Figure 3.3: Pharmacophore model developed using MOE 2012.10. JIP1 is displayed as a stick diagram, JNK1 is displayed as a line diagram. Residues are labelled. Hydrogen bond donor and acceptor interactions are shown as magenta and cyan spheres respectively, and hydrophobic interactions are shown as green spheres. Carbon atoms are coloured orange or white, nitrogen atoms are coloured blue and oxygen atoms are coloured red.

During docking, energy minimization of the protein and ligand structures was carried out using the AMBER99 forcefield at a gradient of 0.1 kcal/mol/Å<sup>3</sup>[154]. The Triangle Matcher ligand placement[155], London dG scoring function, refinement using AMBER99 forcefield and Generalised Born

Volume Integral / Weighted Surface Area (GBVI/WSA) dG rescoring function were used in this docking with pharmacophore constraints process[156-158].

Molecular dynamics simulations were carried out on selected poses which fit alongside the JIP1 binding groove of JNK1, fit the pharmacophore model and had the best docking energy score. All topology and parameter files were created using the LEaP program included with AMBER. The entire system to be simulated consists of the protein surrounded by a truncated octahedron of water and sodium ions to maintain charge neutrality. An explicit solvent model was used with TIP3P water molecules filling 12.0 Å between the surface of the protein and the edge of the box[159].

The SANDER package within AMBER was used for 1000 steps of restrained energy minimization and non-restrained energy minimization. The amino acid residues which make up JNK1 were restrained in the first energy minimization process. In both energy minimization processes, steepest descent minimization was used for the first 500 steps. After 500 steps, the energy minimization process is switched to conjugate gradient. The PMEMD module from Amber 12.0[160] was then used to model the system to equilibrium by including 100 ps of constant volume conditions with heating from 0 to 310 K and positional restraints, followed by 100 ps at constant temperature conditions to maintain the density at equilibrium before beginning the molecular dynamics production run with the NPT ensemble. The MD simulations were carried out

with a 2 fs time step using the AMBER 2003 force field[161]. Hydrogen bonds were constrained using the SHAKE algorithm[162]. Each molecular dynamics production run has a total duration of 3 ns with a time step of 10 ps for every frame recorded in its trajectory.

Viewing of the trajectories and clustering of a trajectory was carried out using Chimera 1.8[163]. The CPPTRAJ module in Amber 12.0 was also used to carry out postprocessing of the trajectory which includes calculating the ligand root mean square deviation (RMSD) and an alternative hierarchical agglomerative clustering of the trajectory. Trajectories which show a large RMSD value where the bound ligand keeps changing conformation throughout the trajectory, moved out of the binding pocket or formed many clusters of very small and different populations were determined to have an unstable trajectory and the compound was removed from the dataset based on this additional filter. Trajectories which showed a relatively stable trajectory by remaining within the binding pocket with little change in conformation and hence low ligand RMSD values and the most populated cluster having the largest population compared to other clusters within the trajectory were retained for further analysis.

A redocking experiment was carried out using rigid docking methods in MOE 2013.10 where poses with interactions that are consistent with the molecular dynamics trajectory data are chosen. The ligand pose is energy minimized using the MMFF94 force field[164] and the electrostatics, polar and

hydrophobic contact surfaces surrounding the ligand pose are analysed using the surface analysis feature in MOE[165].

### **3.2.3 Analysis of interaction profiles of identified JNK1-JIP1 inhibitors**

After refining the list of compounds based on the results of the ligand-based and the structure-based studies, the list of compounds were further refined for further analysis by identifying poses which meet the pharmacophore modeling, molecular dynamics simulations and redocking criteria:

- 1) The compound fits into the binding pocket that JIP1 occupies in the X-ray crystal structure of JNK1.
- 2) The compound forms an electrostatic or polar interaction with Arg 127.
- 3) The compound forms hydrophobic non-bonded contacts with Val 118, Leu 123, Leu 131 and Val 159.

MGLTools[166], Ligplot[167], Discovery Studio Visualizer[55] and MOE were used to visually inspect selected binding pose on the protein target and the interactions that were formed. Chimera 1.8 was used to analyze each trajectory, calculate the root mean-square deviation (RMSD) of the ligand and to carry out clustering of the entire trajectory to select representative poses from the most populated cluster[163]. The CPPTRAJ module in Amber 12.0 was used to monitor the distance between a ligand atom and a residue atom where hydrogen bond could potentially be present. If pi interactions could potentially be present,

the distance between the geometric center of the aromatic ring of the ligand and the residue atom was monitored.

### 3.3 Results and Discussion

#### 3.3.1 Initial virtual screening using docking with pharmacophore constraints

The structure 1UKH which contains the truncated JIP1 in complex with JNK1 was used to obtain and select pharmacophore features. A less strict criteria was chosen for the pharmacophore map where the H-bond interaction with Arg 127 and the non-bonded contact with Cys 163 was chosen to enable a subset of the tested JNK1 inhibitors in the literature to meet this docking with pharmacophore constraints criteria. The results are briefly summarized in the Table 3.2.

Table 3.2: Docking results using Molecular Operating Environment.

Compound ID	Reference (PubMed ID)	Published IC <sub>50</sub> (μM)	Pharmacophore constraints filter	No. of poses	MM/GBVI (kcal/mol)
BI-78D3	18922779	0.28	Yes	0	-
			No	26	9.13 to -4.59
25	21458276	1.3	Yes	5	-3.96 to -4.65
			No	20	-0.37 to -4.42
19	21458276	1.8	Yes	2	-4.01 to -4.29
			No	23	-2.93 to -4.26

BI-98A10	20045647	3.0	Yes	3	-3.12 to -3.78
			No	24	7.09 to -4.29
1	19243309	3.6	Yes	0	-
			No	28	3.36 to -5.06
87D12	19282190	3.8	Yes	0	-
			No	19	3.80 to -3.82
BI-90H10	20045647	5.7	Yes	0	-
			No	27	8.82 to -4.72
BI-90H8	20045647	9.1	Yes	0	-
			No	24	-0.53 to -4.75
AV-7	19527717	10.0	Yes	0	-
			No	30	3.26 to -4.92
BI-90E2	20045647	> 25	Yes	3	-3.88 to -4.46
			No	24	6.52 to -4.56
BI-90E7	20045647	> 25	Yes	4	-3.76 to -4.73
			No	25	8.70 to -4.18
BI-83B3	18922779	> 100	Yes	0	-
			No	19	2.39 to -4.24
8e	19271755	> 100	Yes	0	-
			No	28	13.34 to -4.70
3	21458276	> 100	Yes	2	-3.45 to -3.64
			No	26	4.97 to -4.45
5a	21458276	> 100	Yes	1	-4.27
			No	28	-2.88 to -4.35
10	21458276	> 100	Yes	2	-3.86 to -4.28
			No	22	2.82 to -4.07



The results from the rigid docking of compounds show a large number of different predicted poses and a wide range of docked scores with a molecular mechanics generalized born/volume integral (MM/GBVI) binding free energy, with some poses having a calculated binding energy that is in the positive range and is less likely to bind and form interactions when the pharmacophore constraints filter was not used.

Including this pharmacophore constraints filter reduced the number of docked poses for each compound from 30 to 5 or less. This suggests that the pharmacophore constraints filter is potentially able to remove erroneous poses.

The docking process was repeated using the same pharmacophore filter on identified positive compounds from the QSAR model in Chapter 2 and reduced the number of possible compounds from 45673 to 17560. This additional step helps to reduce the time needed to dock each compound, and also helps to further refine and reduce the number of docked poses for molecular dynamics studies, as each molecular dynamics simulation run can take a very long time to complete and the starting pose of a compound appears to have an important role in producing a stable or unstable trajectory.

### **3.3.2 Refinement of virtual screening using molecular dynamics**

Molecular dynamics using the software package AMBER 12.0 was carried out on the top-ranked pose for each compound which meet the docking with pharmacophore constraints criteria. Due to the very high computation cost and time required to run each molecular dynamics simulation, only the top-ranked pose and the pose which fit alongside the JIP-binding pocket identified by the pharmacophore model were selected for molecular dynamics simulations. Poses which fit perpendicular across the JIP-binding pocket tend to show a less stable trajectory and were not selected for molecular dynamics simulations. A total of 9 validation compounds and 9 ZINC database compounds were selected for molecular dynamics simulations.

A flowchart of steps taken in this phase of the research project is shown in Figure 3.4.

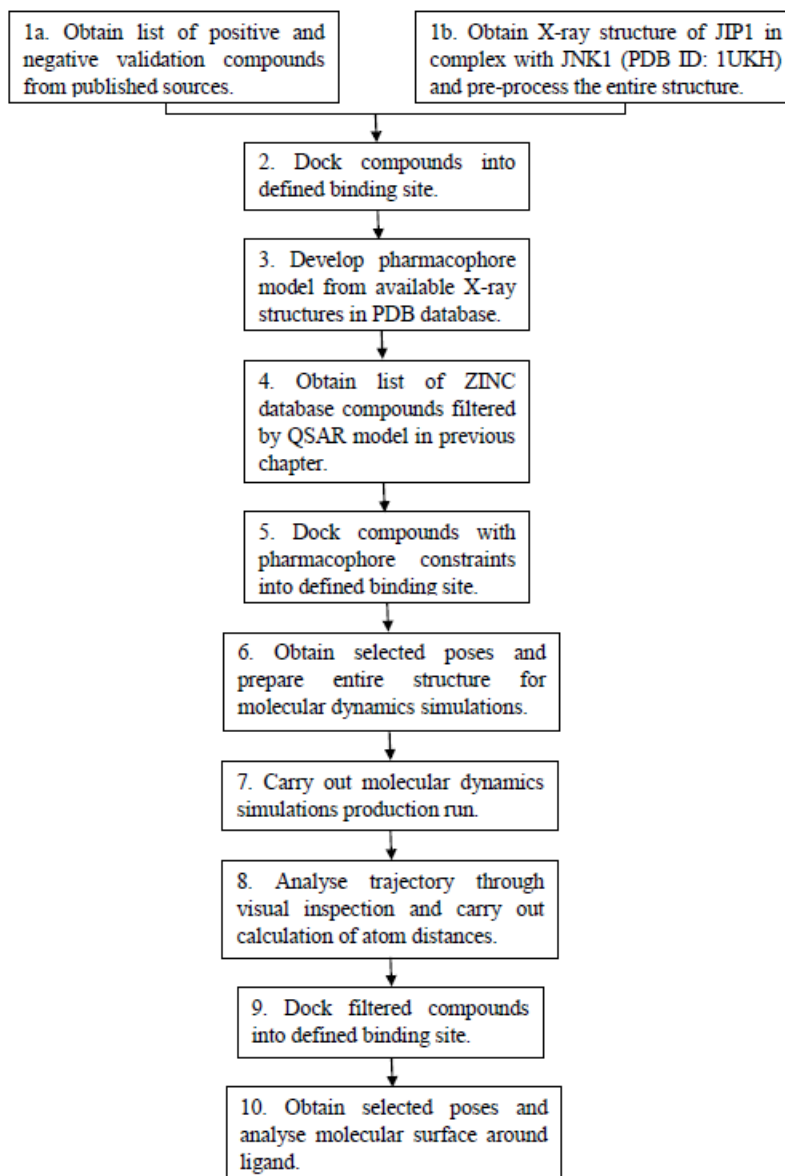
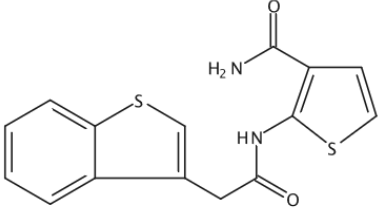
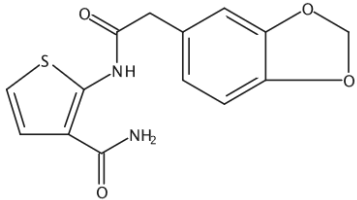
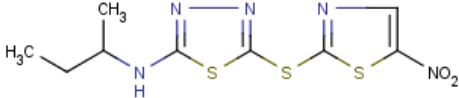
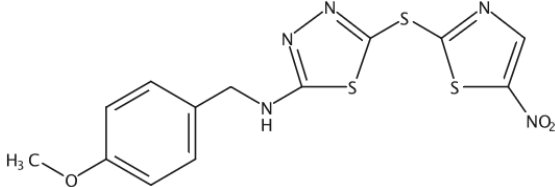
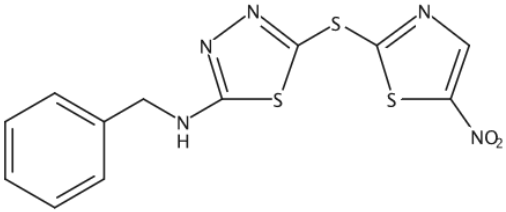
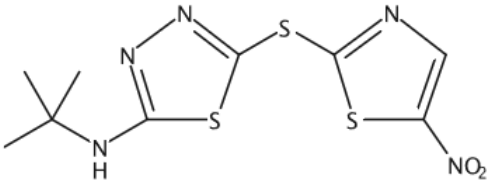
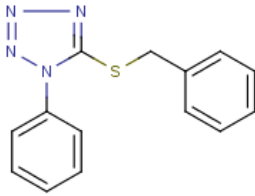
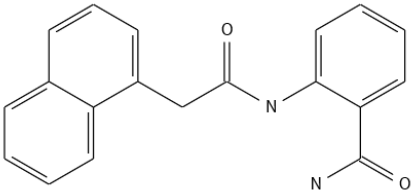
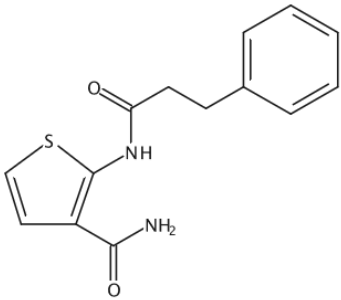


Figure 3.4: Flowchart of steps taken in this virtual screening workflow.

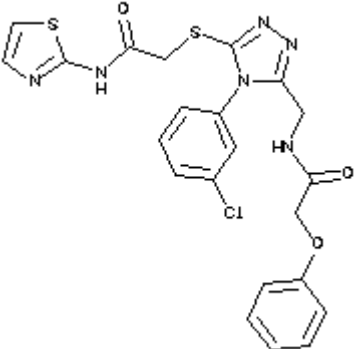
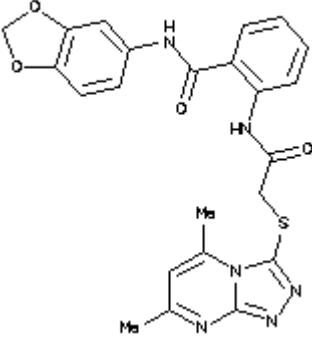
Table 3.3: Chemical structures of validation compounds drawn using Marvin Sketch[168] or obtained from SciFinder[169].

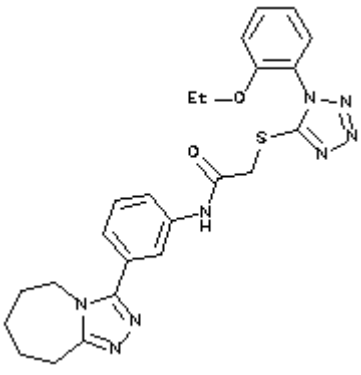
Compound name	Structure	MM/GBVI of selected pose	Published IC <sub>50</sub> (μM)
25		-4.65	1.3
19		-4.29	1.8
BI-98A10		-3.78	3.0
BI-90H8		-4.75	9.1

BI-90E2		-4.46	> 25
BI-90E7		-4.73	> 25
BI-83B3		-4.24	> 100
3		-3.64	> 100
10		-4.28	> 100

Molecular dynamics simulations were carried out on compounds from the ZINC database. Three compounds (ZINC11681157, ZINC03405497 and ZINC09369056) were retained for further analysis of interaction profiles after having a docking score of -5 or less (Table 3.4).

Table 3.4: Chemical structures and docking scores for molecular dynamics trajectories of 3 ns duration.

Compound name	Structure	MM/GBVI of selected pose
ZINC09369056	 <p>The structure of ZINC09369056 is a complex molecule featuring a central 1,2,4-triazole ring. This ring is substituted with a thiazole ring at the 1-position, a 4-chlorophenyl ring at the 2-position, and a propyl chain at the 4-position. The propyl chain is further substituted with a benzamide group at the end, which is connected to a phenyl ring via an ether linkage.</p>	-5.2
ZINC03405497	 <p>The structure of ZINC03405497 consists of a 1,2,4-triazole ring substituted with two methyl groups at the 5 and 6 positions. It is also substituted with a propyl chain at the 4-position. The propyl chain is further substituted with a benzamide group at the end, which is connected to a benzofuran ring system via an amide linkage.</p>	-5.1

ZINC11681157		-5.1
--------------	--	------

### 3.3.3 Clustering

The population size of the top 3 clusters in each validation compound and three ZINC database compounds are listed in Table 3.5. A representative pose from Cluster 1 is obtained for further analysis.

Table 3.5: Cluster size and total population of each compound. Numbers in parenthesis refer to the percentage of population in their respective clusters to the total population available in the trajectory.

Compound name	Cluster 1	Cluster 2	Cluster 3	Total population
25	63 (63.0)	25 (25.0)	12 (12.0)	100
19	75 (75.0)	11 (11.0)	8 (8.0)	100
BI-98A10	55 (55.0)	26 (26.0)	17 (17.0)	100
BI-90H8	47 (47.0)	31 (31.0)	9 (9.0)	100
BI-90E2	60 (100.0)	0 (0.0)	0 (0.0)	60
BI-90E7	13 (21.7)	13 (21.7)	8 (13.3)	60
BI-83B3	11 (18.3)	10 (16.7)	7 (11.7)	60

3	80 (80.0)	11 (11.0)	9 (9.0)	100
10	27 (45.0)	17 (28.3)	14 (23.3)	60
ZINC11681157	170 (56.7)	118 (39.3)	12 (4.0)	300
ZINC03405497	299 (99.7)	1 (0.3)	0 (0.0)	300
ZINC09369056	209 (69.7)	73 (24.3)	6 (2.0)	300

### 3.3.4 Analysis of interaction profiles

The hydrogen bond interactions between Thr 159 of JIP1 and Arg 127 of JNK1 were chosen to be included in the pharmacophore model and for further analysis of interactions. This is based on isothermal calorimetry experiments performed by Heo et al to measure dissociation constant ( $K_d$ ) values of JIP1 to bind to wild-type JNK1 with Arg 127 and to mutant JNK1 with Ala 127 where the  $K_d$  values were  $0.42 \pm 0.13 \mu\text{M}$  and  $6.4 \pm 2.2 \mu\text{M}$  respectively[57].

Among each of the listed compounds, they show a relatively stable trajectory by maintaining their distance and interaction with Arg 127 throughout the entire simulation run.

Analysis of the redocked and selected poses show that for the validation compounds, compounds 19(Figure 3.5) and 25 may potentially form pi-cationic interactions with Arg 127 through its thiophene carboxamide group and the distance between the geometric center of its thiophene group and Arg 127 is



maintained within 6 Å throughout most of the trajectory(Figures 3.6 and 3.7). Compound 19 potentially forms electrostatic interactions with Arg 127(Figure 3.8), polar non-bonded contacts with Arg 127 and Ser 161(Figure 3.9) and hydrophobic non-bonded contacts with Ala 113, Val 118, Leu 123, Leu 131 and Val 159(Figure 3.10). Compounds BI-98A10 and BI-90H8 may potentially form pi-cationic interactions with Arg 127 through its thiadiazole group(Figures 3.11 to 3.14). Both compounds may potentially form polar non-bonded contacts with Arg 127(Figures 3.15 and 3.16), and form hydrophobic non-bonded contacts with Val 118, Leu 123, Leu 131 and Val 159(Figures 3.17 and 3.18). For the decoy compounds BI-83B3, 3 and 10, when the distance between its aromatic thiophene group or 6-member ring is measured, the maximum distance is longer and the proportion of the trajectory where the distance is within 6 Å is smaller compared to compounds 19 and 25(Figures 3.19 to 3.21 and Table 3.6)[170,171].

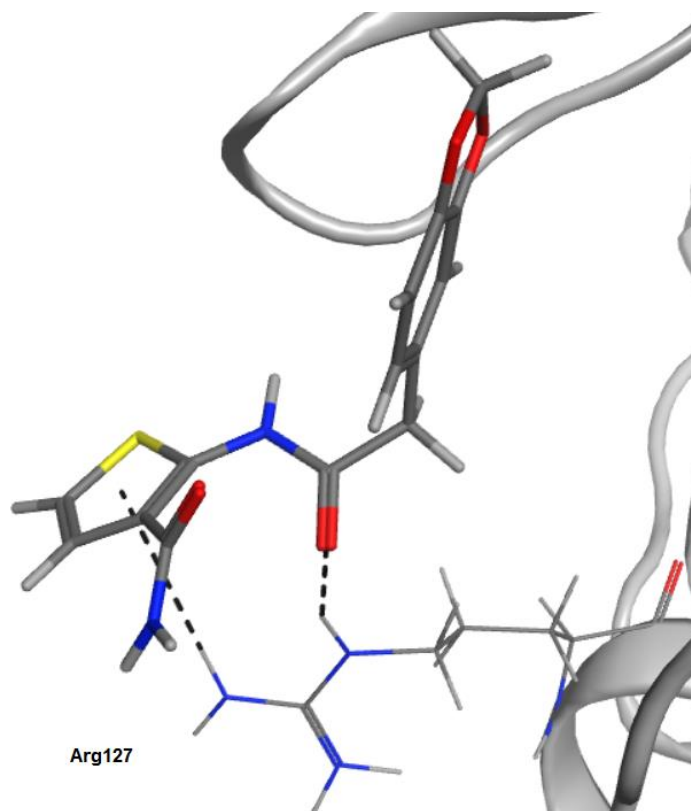


Figure 3.5: Pose of Compound 19. Potential interactions are marked by dotted lines.

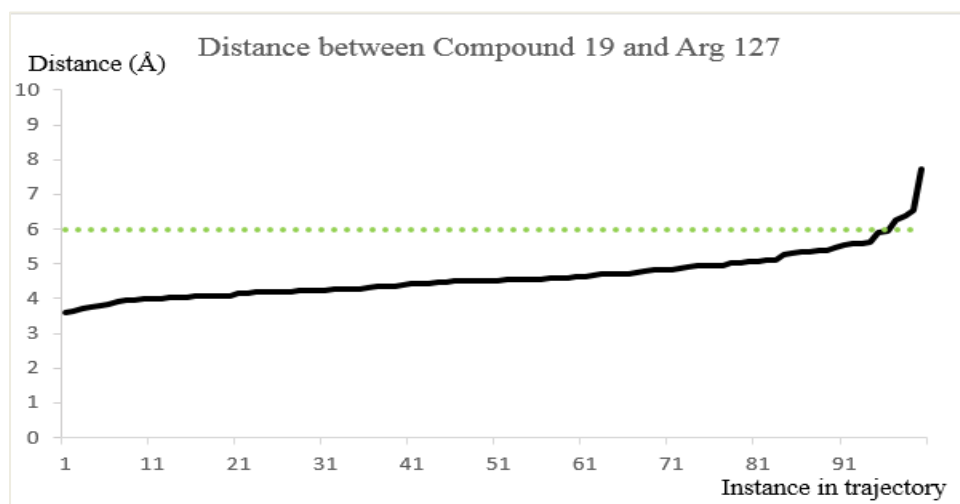


Figure 3.6: Graph of distance between geometric center of aromatic ring in Compound 19 and Arg 127 against instance in trajectory. A 6 angstrom distance cutoff is shown as a dotted line.

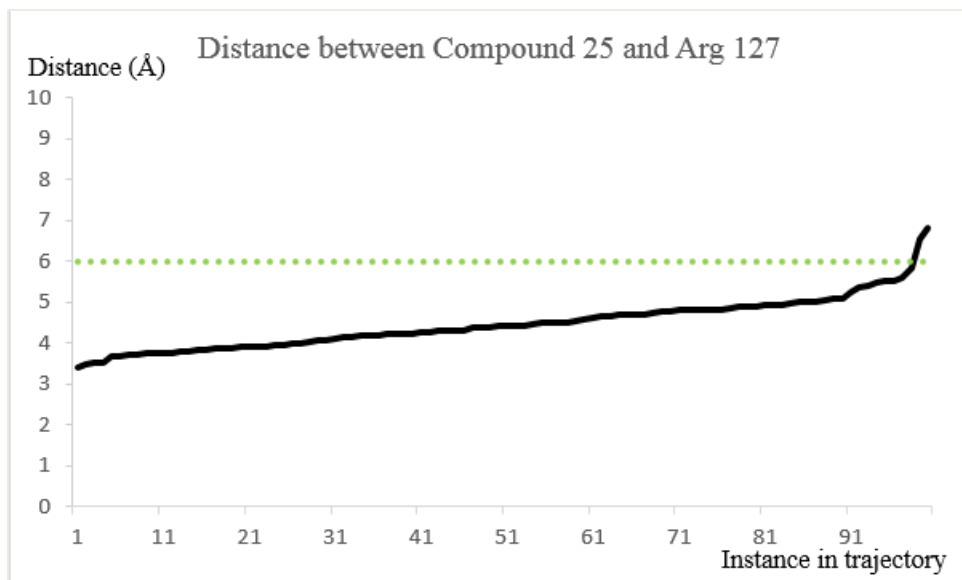


Figure 3.7: Graph of distance between geometric center of aromatic ring in Compound 25 and Arg 127 against instance in trajectory. A 6 angstrom distance cutoff is shown as a dotted line.

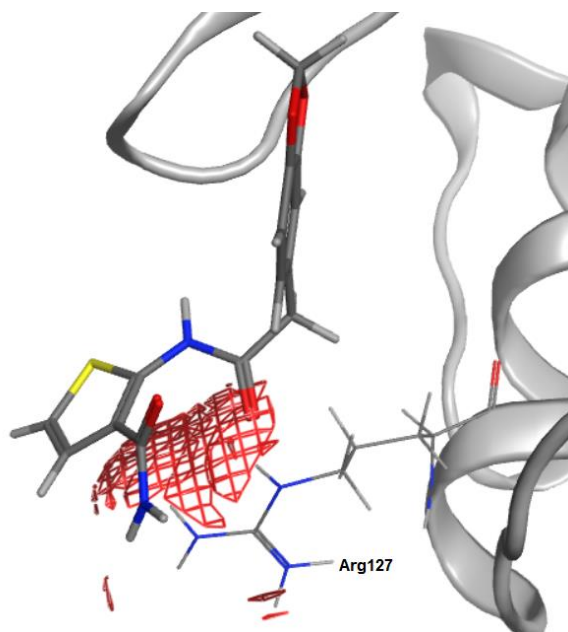


Figure 3.8: Pose of Compound 19 over ribbon diagram of JNK1. Charged residues within 4.0 Å distance are labelled. Potential electrostatic interactions within 4.0 Å distance from ligand with a preference for negatively charged atoms are shown as red spheres.

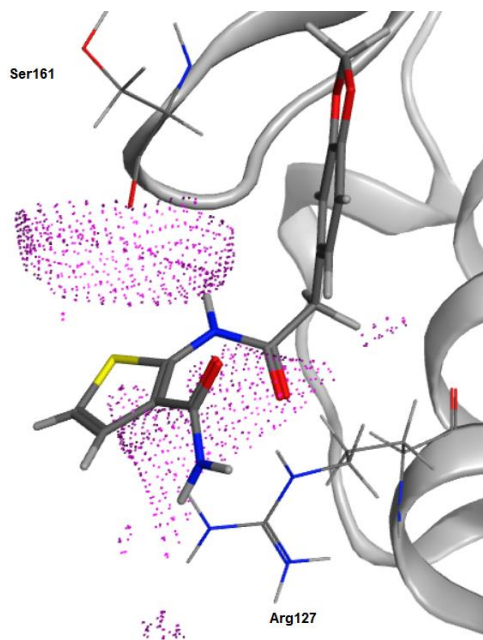


Figure 3.9: Pose of Compound 19 over ribbon diagram of JNK1. Polar residues within 4.0 Å distance are labelled. Polar contact surfaces within 4.0 Å distance from ligand are shown as dotted purple spheres.

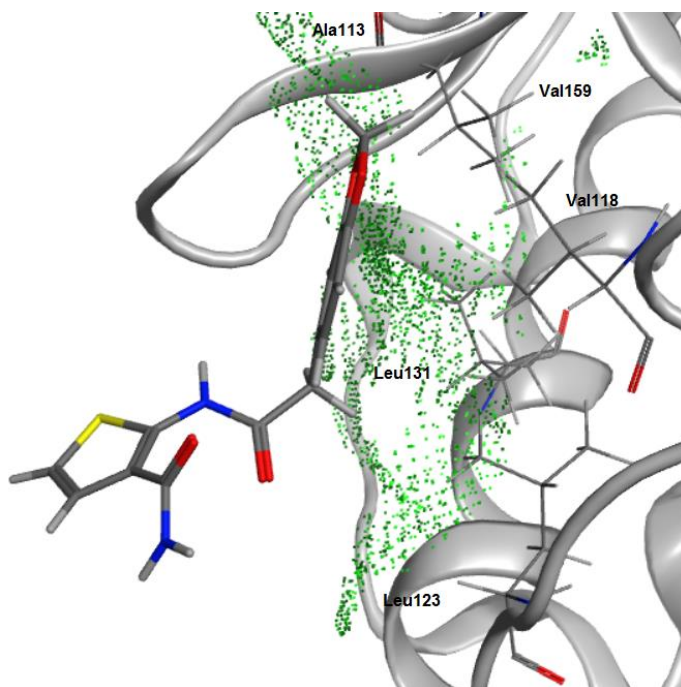


Figure 3.10: Pose of Compound 19 over ribbon diagram of JNK1. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.

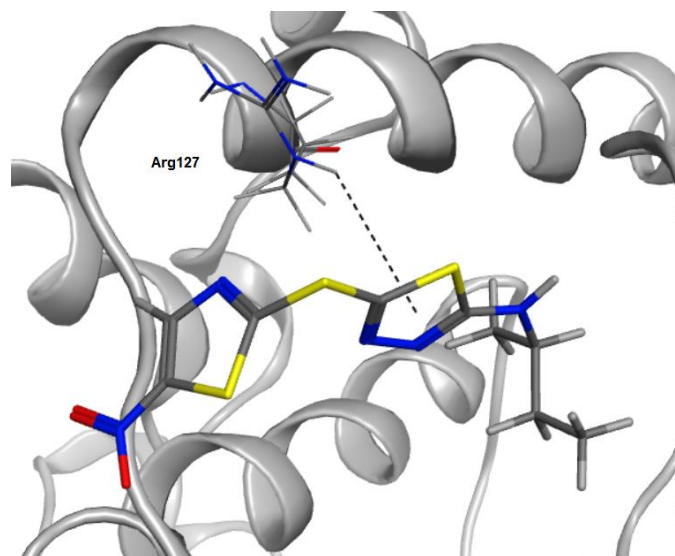


Figure 3.11: Pose of Compound BI-98A10. Potential interactions are marked by dotted lines.

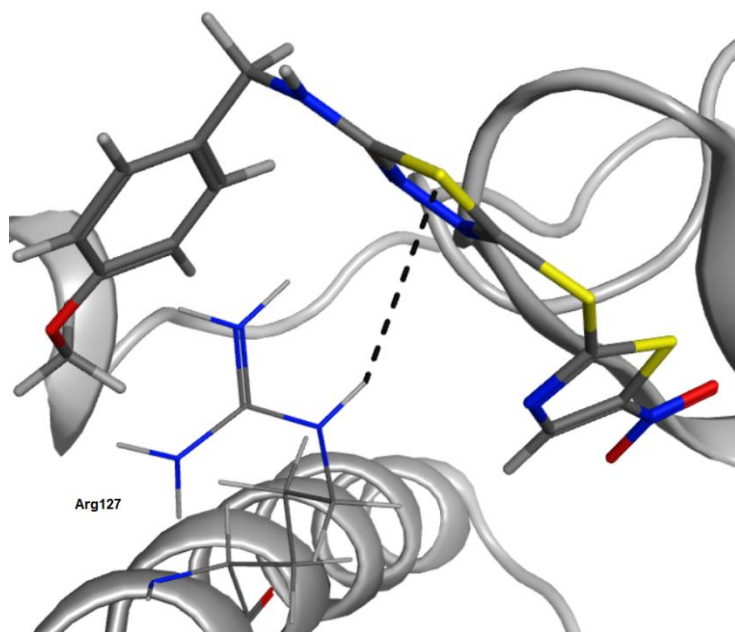


Figure 3.12: Pose of Compound BI-90H8. Potential interactions are marked by dotted lines.

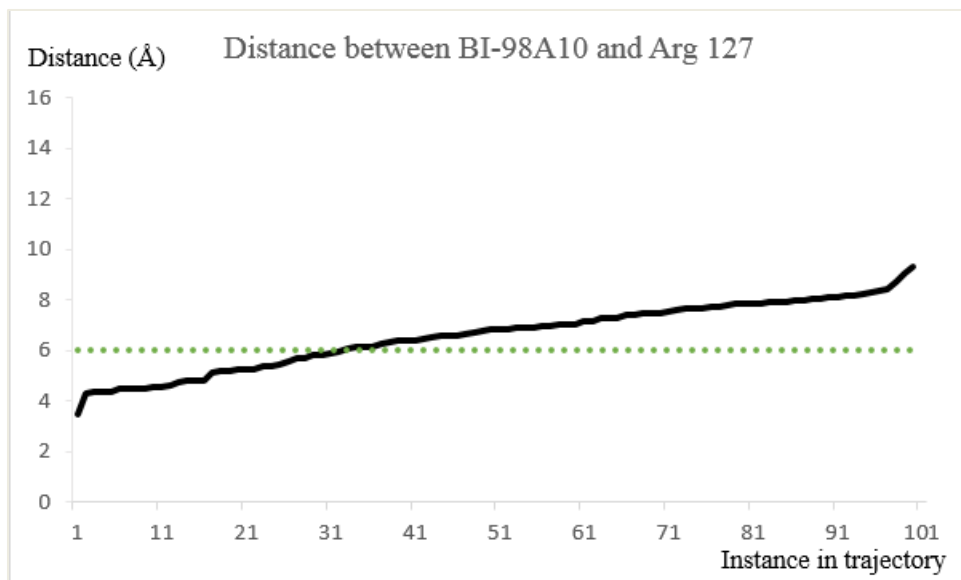


Figure 3.13: Graph of distance between geometric center of aromatic ring in Compound BI-98A10 and Arg 127 against instance in trajectory. A 6 angstrom distance cutoff is shown as a dotted line.

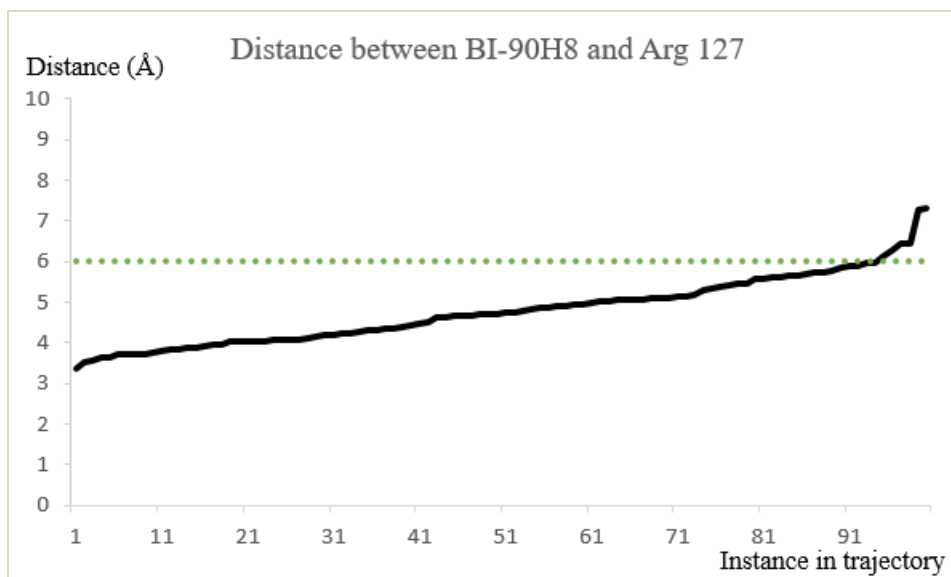


Figure 3.14: Graph of distance between geometric center of aromatic ring in Compound BI-90H8 and Arg 127 against instance in trajectory. A 6 angstrom distance cutoff is shown as a dotted line.

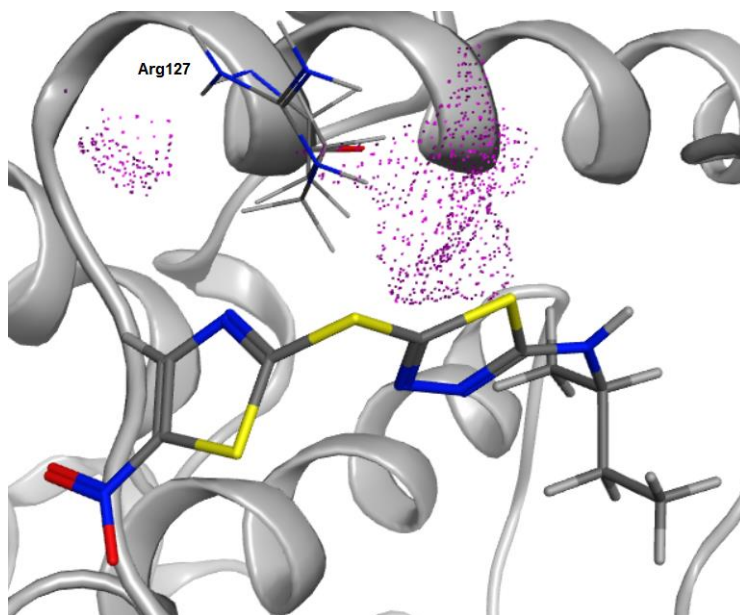


Figure 3.15: Pose of Compound BI-98A10 over ribbon diagram of JNK1. Polar residues within 4.0 Å distance are labelled. Polar contact surfaces within 4.0 Å distance from ligand are shown as dotted purple spheres.

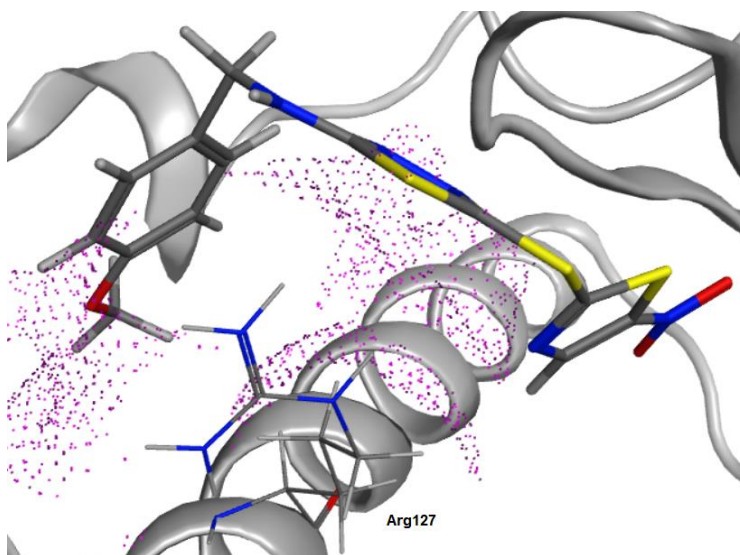


Figure 3.16: Pose of Compound BI-90H8 over ribbon diagram of JNK1. Polar residues within 4.0 Å distance are labelled. Polar contact surfaces within 4.0 Å distance from ligand are shown as dotted purple spheres.

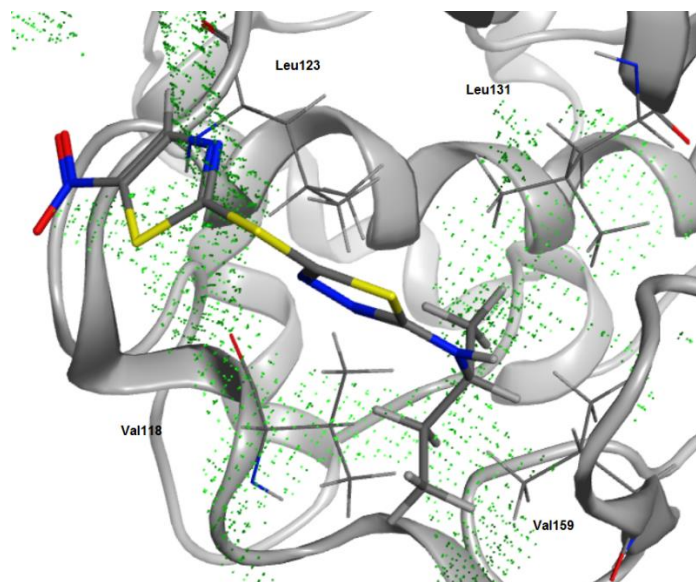


Figure 3.17: Pose of Compound BI-98A10 over ribbon diagram of JNK1. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.

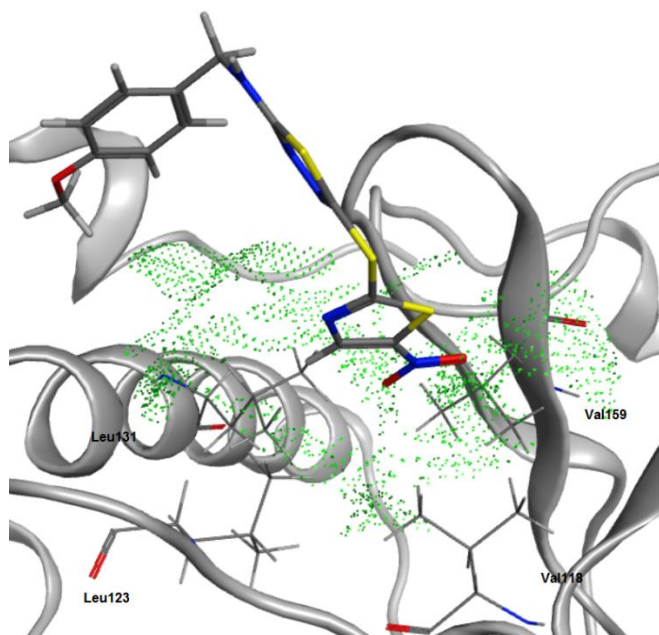


Figure 3.18: Pose of Compound BI-90H8 over ribbon diagram of JNK1. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.



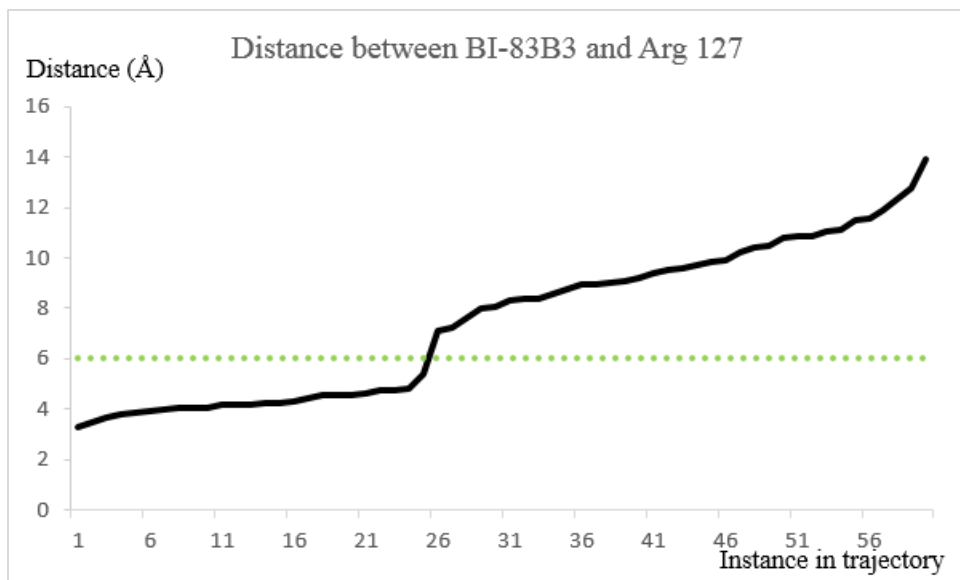


Figure 3.19: Graph of distance between geometric center of aromatic ring in Compound BI-83B3 and Arg 127 against instance in trajectory. A 6 angstrom distance cutoff is shown as a dotted line.

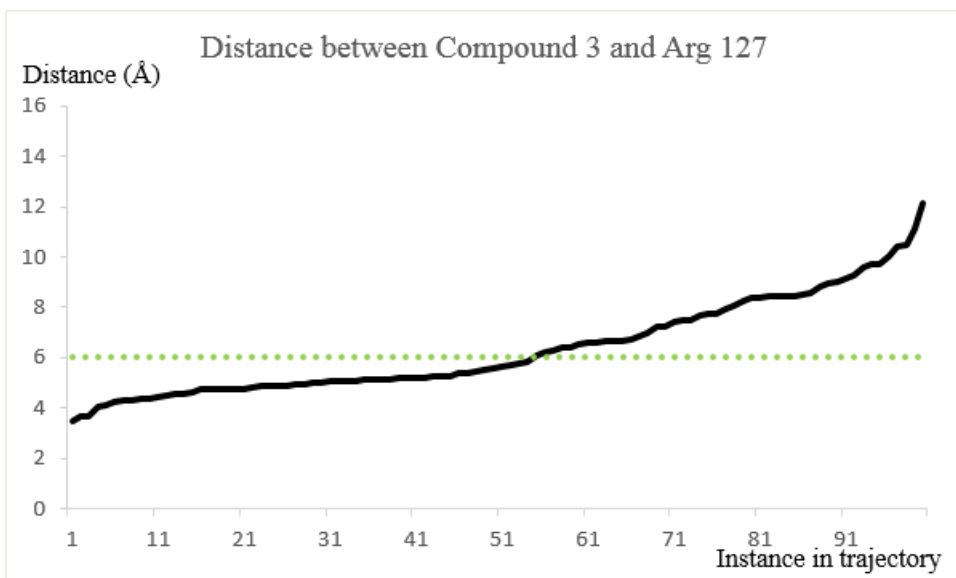


Figure 3.20: Graph of distance between geometric center of aromatic ring in Compound 3 and Arg 127 against instance in trajectory. A 6 angstrom distance cutoff is shown as a dotted line.

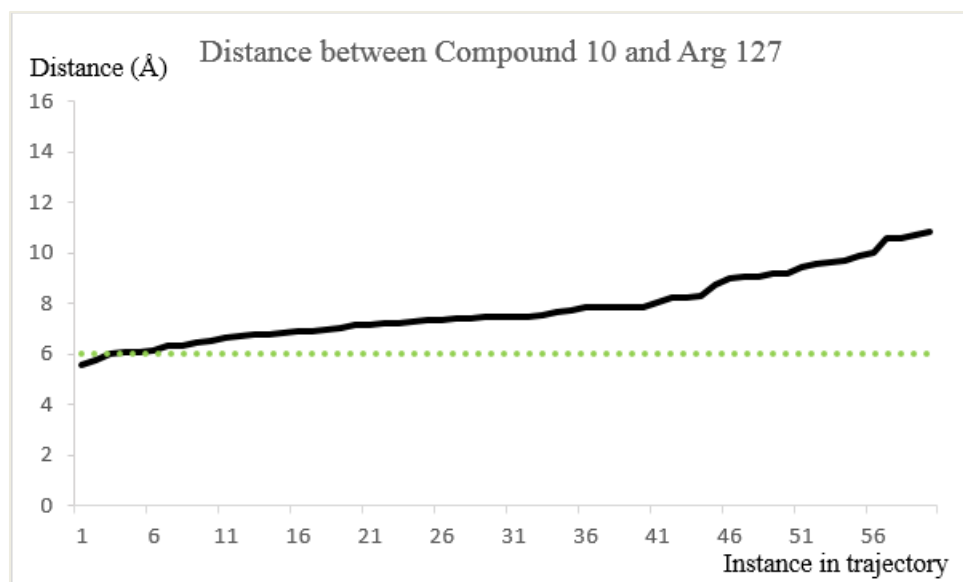


Figure 3.21: Graph of distance between geometric center of aromatic ring in Compound 10 and Arg 127 against instance in trajectory. A 6 angstrom distance cutoff is shown as a dotted line.

Table 3.6: Distance between geometric center of aromatic ring in compound and Arg 127.

Compound Name	Median distance (Å)	Standard deviation	Shortest distance (Å)	Longest distance (Å)	Percentage of frames within 6.0 Å
19	4.5	0.67	3.6	7.7	96
25	4.4	0.62	3.4	6.8	98
BI-98A10	6.8	1.3	3.4	9.3	32
BI-90H8	4.7	0.83	3.4	7.3	94
BI-83B3	8.2	3.0	3.3	14	42
3	5.6	1.9	3.5	12	54
10	7.5	1.3	5.6	11	3.3

BI-90H8 may potentially form hydrogen bond interactions with Arg 127 through its thiazole group(Figures 3.22 and 3.23). The decoy compounds BI-90E2

and BI-90E7 have a longer maximum distance and the proportion of the trajectory where the distance is within 3.5 Å is smaller compared to BI-90H8(Figures 3.24 and 3.25 and Table 3.7)[172].

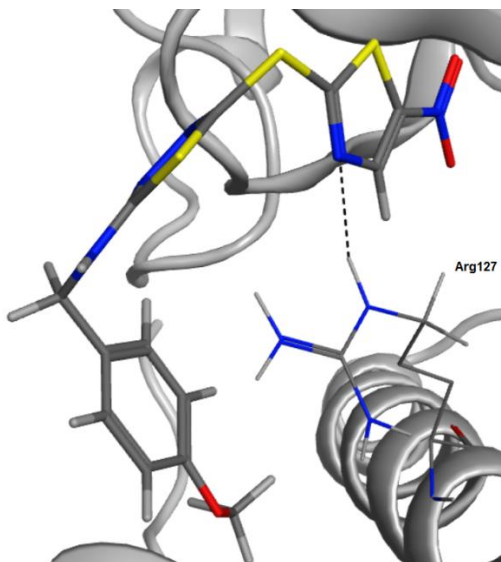


Figure 3.22: Pose of Compound BI-90H8. Potential interactions are marked by dotted lines.

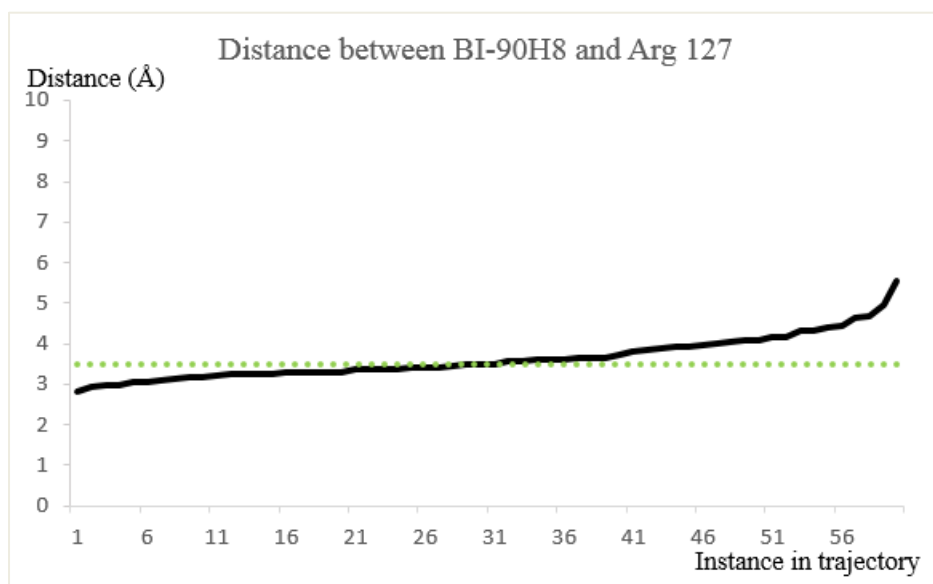


Figure 3.23: Graph of distance between large atom in Compound BI-90H8 and Arg 127 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

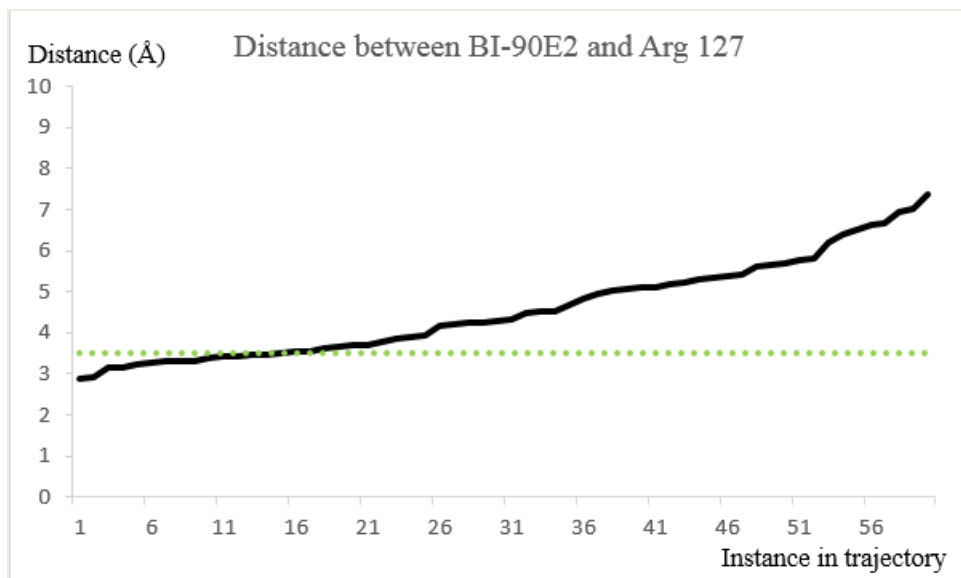


Figure 3.24: Graph of distance between large atom in Compound BI-90E2 and Arg 127 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

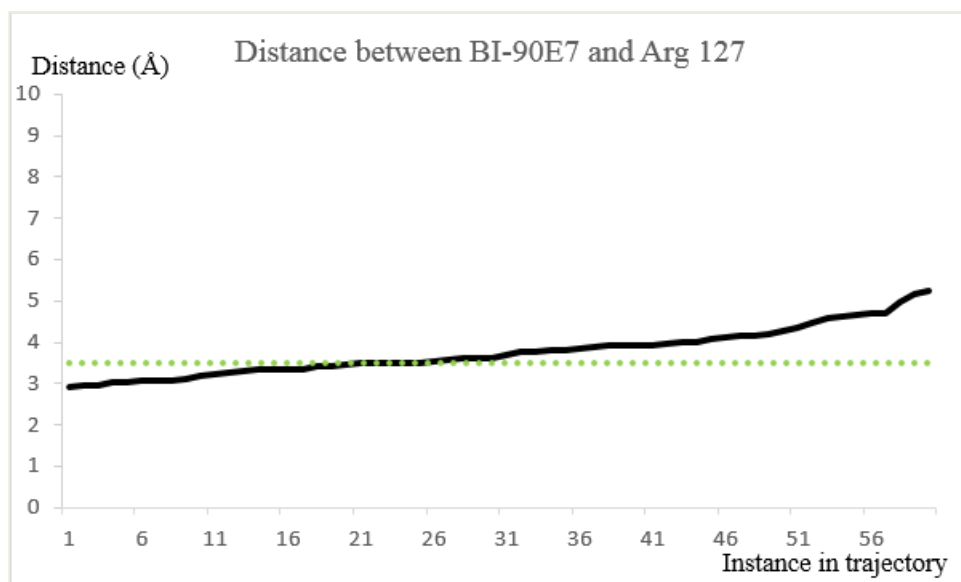


Figure 3.25: Graph of distance between large atom in Compound BI-90E7 and Arg 127 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

For the ZINC database compounds, they show ZINC09369056 changing conformation to point its thiazole sulfur towards the carbonyl after 1.6ns in the

simulation. This shows that the molecular dynamics is able to reduce the energy state of the ligand and gives valuable insights into the possible binding orientations of each ligand. ZINC03405497 is prioritized for *in vitro* testing after showing a very stable trajectory with its amide group remaining within hydrogen bond distance throughout the entire trajectory(Figure 3.26, Figure 3.27 and Table 3.7). ZINC03405497 form potential electrostatic interactions and polar non-bonded contacts with Arg 127 (Figure 3.28) and hydrophobic non-bonded contacts with Val 118, Leu 123, Leu 131, Val 159 and Trp 324(Figure 3.29).

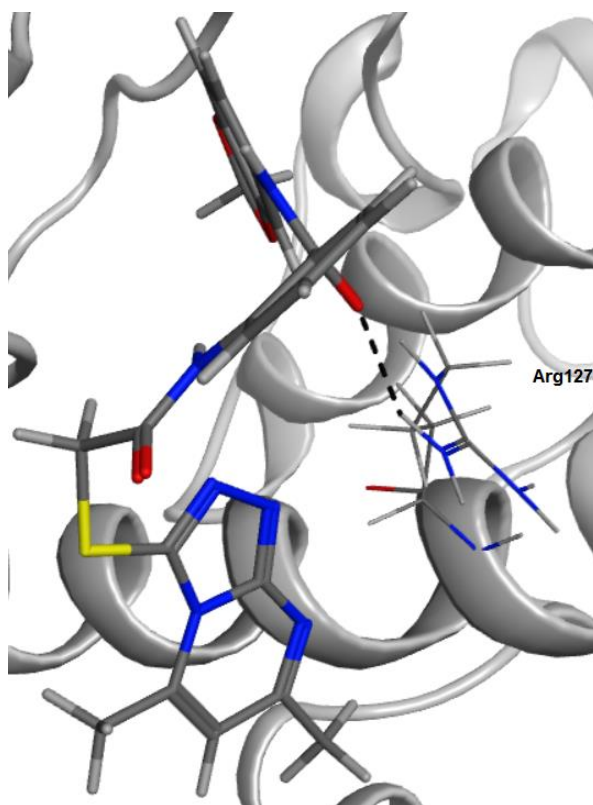


Figure 3.26: Pose of ZINC03405497. Potential interactions are marked by dotted lines.

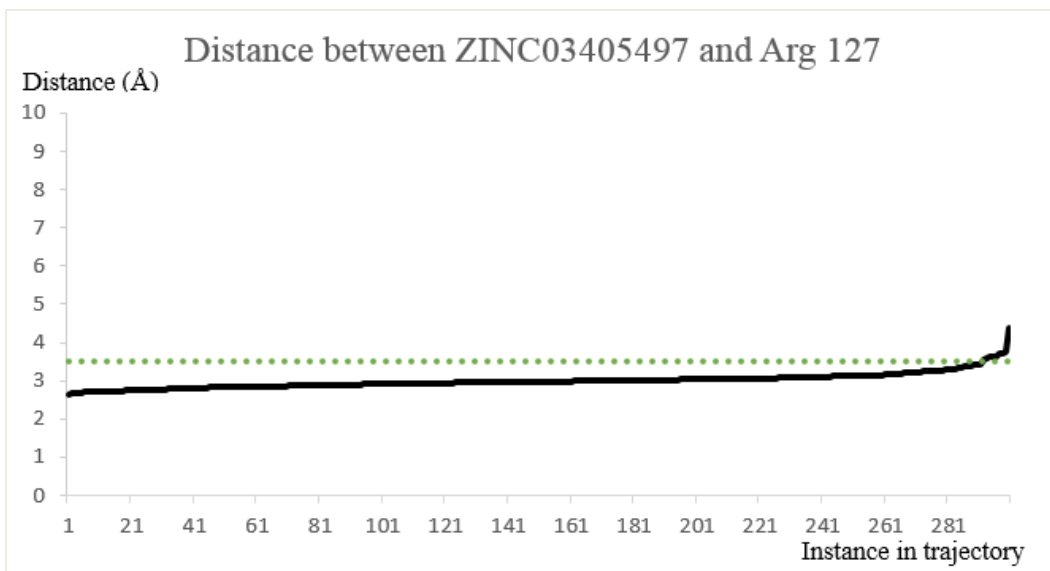


Figure 3.27: Graph of distance between large atom in ZINC03405497 and Arg 127 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

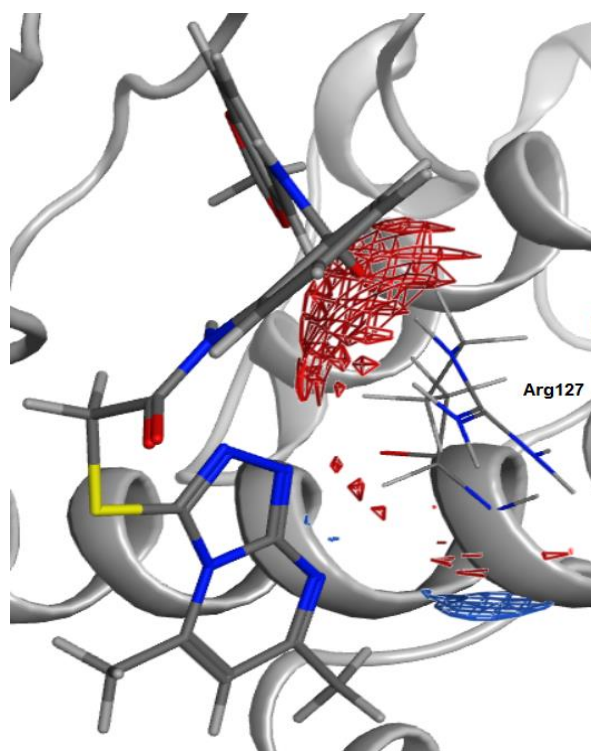


Figure 3.28: Pose of ZINC03405497 over ribbon diagram of JNK1. Charged residues within 4.0 Å distance are labelled. Potential electrostatic interactions within 4.0 Å distance from ligand with a preference for negatively charged atoms are shown as red spheres and positively charged atoms as blue spheres.

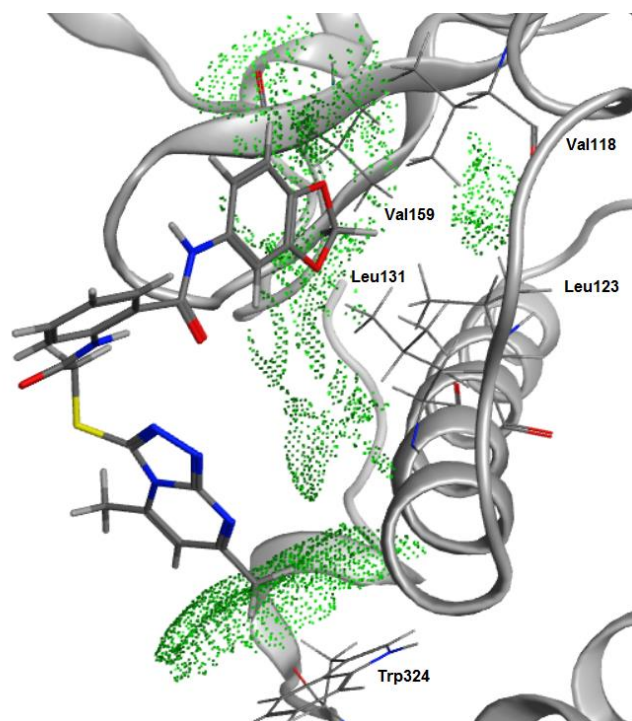


Figure 3.29: Pose of ZINC03405497 over ribbon diagram of JNK1. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.

Table 3.7: Distance between large atoms in Arg 127 of JNK1 and compound.

Compound name	Median distance (Å)	Standard deviation	Shortest distance (Å)	Longest distance (Å)	Percentage of frames within 3.5 Å
BI-90H8	3.6	0.53	2.8	5.6	52
BI-90E2	4.3	1.2	2.9	7.4	23
BI-90E7	3.8	0.57	2.9	5.2	37
ZINC11681157	5.5	0.83	3.7	8.0	0.0
ZINC03405497	3.0	0.20	2.6	4.4	97
ZINC09369056	7.5	1.6	3.0	12.7	0.67

From the analysis of interactions in the molecular dynamics trajectories, the results are consistent with the reported small molecule inhibitors where thiazoles, thiadiazoles and thiophene carboxamides were investigated[82-84], as well as the QSAR study where JNK1-JIP1 inhibitors have hydrogen bond donor and acceptor properties and do not appear to form interactions through six-membered rings.

### **3.4 Conclusion**

In doing a virtual screening of ZINC database compounds to identify potent compounds that could potentially inhibit JNK1 at the JIP1 binding site, one ZINC database compound was identified that could be prioritized for *in vitro* testing.

The entire virtual screening process is less accurate in ranking and filtering compounds among the validation compounds in the published JNK1-JIP1 inhibitors dataset. These challenges remain difficult to address due to the large size of the JIP binding pocket where the exact location and binding pose for small molecule inhibitors are not known. Future work could include testing the identified compounds for biological activity and to refine the model by choosing a different part of the JIP1 binding site for docking, molecular dynamics simulations and analysis of interactions.



## Chapter 4

### Development of structure-based model for DFG binding site inhibitors

#### 4.1 Summary

The compound BIRB796 (Doramapimod) has been found to bind to the DFG binding site of p38 $\alpha$  and JNK2. Virtual screening of a subset of ZINC database compounds with a Tanimoto coefficient of up to 0.7 based on ChemAxon fingerprints[168] was carried out by docking and analysis of interaction profiles. A consensus model that analyzed docking scores using three docking software was used to filter and prioritize compounds for *in vitro* testing. 9 ZINC database compounds and the positive control compound BIRB796 were selected for *in vitro* testing using the KINOMEScan test from DiscoverX to determine K<sub>d</sub> values of the tested compounds on JNK1, JNK2 and p38 $\alpha$ .

#### 4.2 Methods

##### 4.2.1 Ranking of DFG inhibitors based on scoring function

The X-ray crystal structures of JNK2 with the DFG-in and DFG-out conformation were available in the RCSB Protein Databank website (PDB structures 3E7O and 3NPC respectively)[56,65,173]. The structures were

processed by adding hydrogen atoms, assigning tautomer and ionization states to the structure and capping terminal ends of the entire protein structure, removing water molecules in the structure and carrying out energy minimization of the structure using the Assisted Model Building with Energy Refinement 99(AMBER99)[174] or Chemistry at HARvard Macromolecular Mechanics 27(CHARMM27) forcefield[154,175].

A purchasable subset of compounds of up to 0.7 Diversity Index based on ChemAxon fingerprints that were compared to the compound within the most densely populated cluster in the ZINC chemical database[168,176] were downloaded and docking of the compounds was carried out on the DFG binding site of JNK2.

The DFG binding site of JNK2 was defined by covering the following amino acid residues within the search space of the docking procedure. These amino acid residues consist of 12 hydrophobic residues (Ile 32, Val 40, Ala 53, Leu 76, Leu 77, Ile 86, Met 108, Leu 110, Met 111, Leu 142, Leu 168 and Phe 170), 1 polar residue (Gln 37) and 6 polar and charged residues (Lys 55, Arg 69, Arg 72, Glu 73, Glu 109 and Asp 169).

#### 4.2.2 Analysis of interaction profiles of identified DFG site inhibitors

After refining the list of compounds based on the purchasable compounds criteria to enable convenient procurement of the compound from commercial vendors, diversity index metric and the docking score in the structure-based study, the list of compounds were further refined for further analysis by identifying docked poses which fulfill at least one of the following criteria:

- 1) The compound fits into the binding pocket that is defined to be the binding pocket that the compound BIRB796 occupies in the X-ray crystal structure of JNK2.
- 2) The compound forms a H-bond interaction with Glu 73 or Asp 168

MGLTools, Ligplot and Discovery Studio Visualizer were used to visually inspect the interactions that were formed. The binding poses and interaction map profiles were eventually used to select a list of promising compounds for purchase through commercial vendors and to determine K<sub>d</sub> values of the compounds through *in vitro* testing on the KINOMEScan assay platform by DiscoverX[177].

#### 4.2.3 Preparation of compounds for *in vitro* testing

The solubility profiles of selected compounds were calculated using MOE 2011.10 which calculates the LogS values based on the sum of all the atoms and

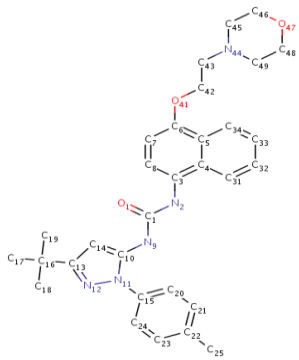
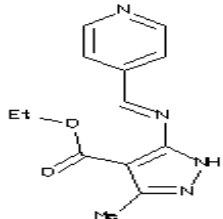
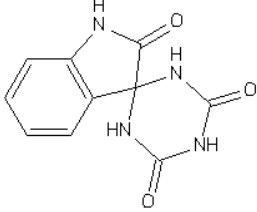
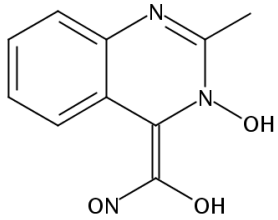
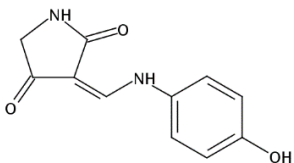
chemical bonds present in the molecule[178]. Compounds with LogS values above 4.0 or below -4.0 were not selected for purchase. The chemical structures and Chemical Accession Service (CAS) numbers of selected compounds were used to obtain permission from the relevant authority to import into the country[169]. The compounds were prepared in 100  $\mu$ L of dimethyl sulfoxide (DMSO) solution at 10mM concentration under collaboration with Life Sciences Institute and Laboratory of Liver Cancer and Drug-induced Liver Diseases Research at National University of Singapore.

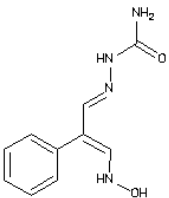
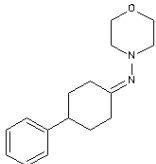
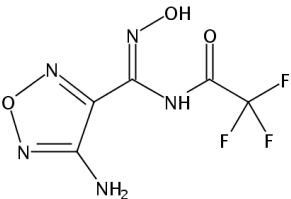
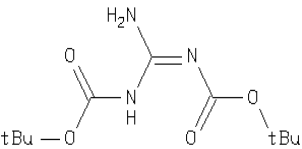
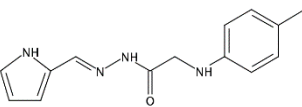
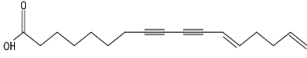
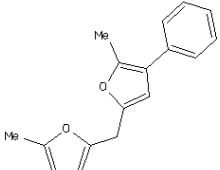
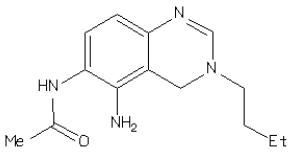
### **4.3 Results and discussion**

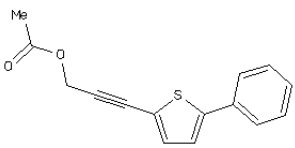
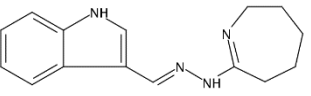
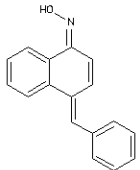
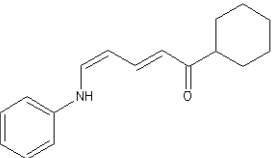
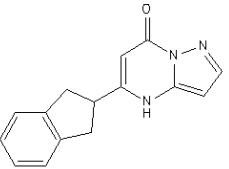
#### **4.3.1 Consensus scoring functions using different docking software**

The docking score and Tanimoto coefficients for each compound compared to the positive validation compound BIRB796 are listed in Table 4.1. Compounds with a higher docking score in SYBYL and compounds with a more negative docking score in MOE or AutoDock Vina indicate a higher chance of binding and hence higher potency of the compound.

Table 4.1: Results of docking scores on various ZINC compounds using different docking software.

ZINC ID	Chemical structure	SYBYL CScore	MOE AMBER99 Score	MOE CHARMM27 Score	AutoDock Vina Score	Tanimoto Coefficient
BIRB796		16	-9.6	-14	-	-
ZINC00088624		4.6	-6.8	-7.7	-6.3	0.25
ZINC00210429		5.1	-6.0	-7.1	-8.2	0.25
ZINC00344359		5.5	-6.0	-6.9	-8.4	0.16
ZINC02026664		3.8	-5.9	-6.1	-8.0	0.16

ZINC04266305		7.1	-6.4	-7.5	-7.9	0.11
ZINC05019440		5.7	-7.1	-7.1	-9.3	0.22
ZINC06320732		2.6	-6.1	-6.5	-7.3	0.09
ZINC06692943		6.4	-6.4	-7.7	-7.4	0.08
ZINC13118854		7.3	-6.9	-7.8	-8.8	0.16
ZINC31166436		8.3	-7.7	-6.8	-7.6	0.06
ZINC05417635		7.0	-7.1	-8.3	-9.4	0.12
ZINC01718148		5.7	-7.6	-6.6	-7.0	0.24

ZINC28541016		6.4	-6.3	-7.9	-7.5	0.12
ZINC03178796		7.4	-7.4	-7.1	-9.2	0.22
ZINC08659433		5.1	-6.7	-8.0	-9.8	0.13
ZINC05799999		7.1	-7.1	-8.0	-8.9	0.12
ZINC39955766		5.6	-6.5	-7.2	-9.8	0.3

### 4.3.2 KINOMEScan assay

The KINOMEScan assay platform by DiscoverRx was used to determine the  $K_d$  values of selected ZINC database compounds on JNK1, JNK2 and p38 $\alpha$ . The positive validation control BIRB796 showed significant binding and  $K_d$  values at nanomolar concentrations on JNK2 and p38 $\alpha$ , while all the ZINC database compounds did not show any significant binding at concentrations of up to 40 $\mu$ M. The percentage inhibition profile for ZINC13118854 was inconclusive as one experimental arm showed no change in signal strength and the second

experimental arm showed near full reduction in signal strength at 40  $\mu\text{M}$  concentration (Table 4.2).

Table 4.2: KINOMEScan assay results.

Name of compound	JNK1	JNK2	p38 $\alpha$
BIRB796 / Doramapimod	Kd = 15 $\mu\text{M}$	Kd = 1.8 nM	Kd = 4.7 nM
ZINC00088624	Kd > 40 $\mu\text{M}$ % inhibition = 2 %	Kd > 40 $\mu\text{M}$ % inhibition = 19 %	Kd > 40 $\mu\text{M}$ % inhibition = 1 %
ZINC13118854	Kd > 40 $\mu\text{M}$ No inhibition	Kd > 40 $\mu\text{M}$ % inhibition = 7 %	Kd > 40 $\mu\text{M}$ Inconclusive
ZINC05019440	Kd > 40 $\mu\text{M}$ No inhibition	Kd > 40 $\mu\text{M}$ No inhibition	Kd > 40 $\mu\text{M}$ % inhibition = 3 %
ZINC06692943	Kd > 40 $\mu\text{M}$ No inhibition	Kd > 40 $\mu\text{M}$ % inhibition = 3 %	Kd > 40 $\mu\text{M}$ % inhibition = 10%
ZINC04266305	Kd > 40 $\mu\text{M}$ % inhibition = 20 %	Kd > 40 $\mu\text{M}$ % inhibition = 8 %	Kd > 40 $\mu\text{M}$ % inhibition = 9 %
ZINC02026664	Kd > 40 $\mu\text{M}$ No inhibition	Kd > 40 $\mu\text{M}$ % inhibition = 12 %	Kd > 40 $\mu\text{M}$ % inhibition = 3 %
ZINC00210429	Kd > 40 $\mu\text{M}$ No inhibition	Kd > 40 $\mu\text{M}$ % inhibition = 2 %	Kd > 40 $\mu\text{M}$ % inhibition = 1 %
ZINC00344359	Kd > 40 $\mu\text{M}$ No inhibition	Kd > 40 $\mu\text{M}$ % inhibition = 2 %	Kd > 40 $\mu\text{M}$ No inhibition
ZINC06320732	Kd > 40 $\mu\text{M}$ No inhibition	Kd > 40 $\mu\text{M}$ % inhibition = 2 %	Kd > 40 $\mu\text{M}$ % inhibition = 2 %



This shows that random docking even with consensus scoring and analysis of interaction fingerprint profile is insufficient to accurately rank and classify compounds with a very high chance for false positives and a very low hit rate. Further work that includes docking with pharmacophore constraints and molecular dynamics simulations will be required to refine the computational model.

#### **4.4 Conclusion**

In doing a virtual screening of ZINC database compounds to identify potent compounds that could potentially inhibit JNK2 at the DFG binding site, 9 compounds were found to show no significant binding using the KINOMEScan assay platform. Future work could include testing a larger number of identified compounds for biological activity.

## **Chapter 5**

### **Structure-based study of human Pregnane X Receptor (PXR) activators**

#### **5.1 Summary**

The compound rifampicin has been found to bind to human PXR and an X-ray crystal structure of rifampicin in human PXR is available in the Protein Databank. A structure-based study was carried out where rifampicin, 5 non-nucleoside reverse transcriptase inhibitors (NNRTIs) and 1 chemical compound pregnenolone carbonitrile were docked in the ligand-binding domain of human PXR with molecular dynamics simulations and docking to filter and select poses for further study.

#### **5.2 Structures of human PXR**

The structures of human PXR consist of a ligand-binding domain, a co-activator domain and a complex binding interface between PXR and RXR. The ligand-binding domain was chosen as the initial site for study in this project as there are 9 X-ray crystal structures that are available in the Protein Databank database[75,86,109-113], as well as several site-directed mutagenesis studies in the literature which help to elucidate the location and mechanism of binding for various strong activators within the ligand-binding domain[85,86,112].

Attempting a docking study and molecular dynamics study of various ligands in this pocket has several challenges. Due to the very large and highly flexible nature of this binding pocket[75,111,112,114], this binding domain has a very high capacity to change conformation and accept a wide variety of ligands, with a reported ligand promiscuity index of 1.006 for human PXR by using molecular quantum number (MQN) distances and ChEMBL activity datasets[23,179]. In addition, there is also currently a large knowledge gap in this area where the mechanism of binding and the X-ray crystal structure or NMR structure of PXR is not known with regard to partial agonists and antagonists that bind to this pocket. Hence, the study of agonists or activators of PXR will be carried out in this project.

The interactions that may play a role in activating human PXR occur at the following residues: Ser-247, Gln-285, Phe-288, Trp-299, Tyr-306, His-407 and Arg-410 as mutagenesis studies have shown changes in human PXR activity on hyperforin[112], SR12813, rifampicin[86] or bisphenol A[85].

The human PXR ligand binding domain consist of 15 hydrophobic residues (Val-211, Leu-240, Met-243, Phe-251, Phe-281, Phe-288, Trp-299, Leu-308, Met-323, Leu-324, Leu-411, Ile-414, Phe-420, Met-425, and Phe-429), four polar residues (Ser-247, Cys-284, Gln-285, and Tyr-306), and four charged residues (Glu-321, His-327, His-407, and Arg-410)[180].

The co-activator domain functions by accepting co-activators in PXR and RXR and increasing the length, duration and stability of binding and activation by ligands, as well as by decreasing the possible number of poses that the ligands might adopt on binding at the ligand-binding site[111]. Due to the difficulty in designing and studying protein-protein docking and interactions, the study of possible compounds for co-activators will not be carried out in this project.

### **5.3 Materials and methods**

#### **5.3.1 Biological activity**

Biological activity data was obtained from collaboration with Faculty of Pharmaceutical Sciences in University of British Columbia at Vancouver. Firefly/Renilla luciferase assay was used to obtain CYP3A4 gene expression and PXR activation. Rifampicin was used as the positive control, pregnenolone carbonitrile (PCN) was used as the negative control, dimethyl sulfoxide (DMSO) was used as the vehicle control, and five non-nucleoside reverse transcriptase inhibitors (NNRTIs) delavirdine, etravirine, rilpivirine, nevirapine and efavirenz were tested for biological activity[181].

### 5.3.2 Initial docking of compounds

The X-ray crystal structure of human PXR as a monomer with rifampicin as the co-crystal structure (PDB structure: 1SKX) was used in this docking study. The structures were processed by adding hydrogen atoms, assigning protons and partial charges to the structure and capping terminal ends of the entire protein structure[152], removing water molecules in the structure and carrying out energy minimization of the structure using the Assisted Model Building with Energy Refinement 99(AMBER99)[174] or Chemistry at HARvard Macromolecular Mechanics 27 (CHARMM27) forcefield[154,175]. The docking function in MOE was used to dock each molecule to human PXR. The induced-fit docking setting was used, due to the highly flexible region of the ligand binding site[75,111,112,114]. This setting uses a London dG scoring function, a force field refinement scheme to refine the docked molecules and to tether the positions of the side chains of the 23 residues surrounding the active site to allow a certain degree of movement of each side chain. The molecular mechanics generalized Born/volume integral (MM/GBVI) binding free energy is calculated as a final pose rescoring function to refine the docking pose. One hundred possible poses were determined for each molecule[156-158].

The top-ranked pose with the best MM/GBVI binding free energy was selected for analysis of interaction with human PXR if it fulfilled at least one of the following criteria[74]:

- 1) the compound fits into the Trp-299 site;
- 2) a H-bond acceptor of the compound interacts with either the His-407 or Gln-285 side chain;
- 3) the compound forms an interaction with Ser-247, Phe-288 or Trp-299.

### **5.3.3 Analysis of interactions**

A similar workflow to Figure 3.4 was carried out where molecular dynamics simulations were carried out on the top-ranked docked pose from the initial docking study. Compounds which show an unstable trajectory or had a docking score that did not meet a cut-off criteria were not included in the redocking experiments. Redocking experiments using MOE 2013.10 were carried out for the structure of rifampicin from the PubChem database and for filtered compounds. Poses which have hydrogen bond or pi interactions that are consistent with molecular dynamics simulations are selected for further analysis of potential electrostatic interactions and non-bonded contact surfaces.

## 5.4 Results and discussion

### 5.4.1 Initial docking results

To validate the docking process, the structure of rifampicin taken from PDB database was redocked into the same ligand binding domain and the best binding pose was copied and superimposed over the original binding position in PDB structure 1SKX and had a RMSD of 1.02 Å. This shows that the docking process was able to dock rifampicin as a positive control accurately onto the ligand-binding domain.

Table 5.1 shows the MM/GBVI binding free energy for the top-ranked pose of each compound and the various controls. Compared with rifampicin (positive control), the non-nucleoside reverse transcriptase inhibitors (NNRTIs) had weaker binding free energy. Among the five individual NNRTIs, the rank order was delavirdine > etravirine > rilpivirine > nevirapine > efavirenz. This shows that the docking process is able to rank the positive compounds rifampicin, etravirine and rilpivirine accurately. The docking process is also able to rank the negative compounds nevirapine and PCN accurately. Delavirdine was ranked as a false positive, and efavirenz was ranked as a false negative.

The correlation curves for MM/GBVI binding free energy or molecular weight against observed experimental values are shown in Figures 5.1 and 5.2

with a squared correlation coefficient of 0.44 and 0.51 respectively. This shows that there is a certain trend and a higher chance to predict higher fold increase in human PXR activity with increasing molecular weight and more negative MM/GBVI binding free energy values. This could be due to the increased number of pharmacophoric features that would be present in larger molecule compounds and the inherent capacity for PXR to adopt large changes in conformation within its ligand-binding domain and hence having a very high ligand promiscuity index.

Table 5.1: List of docked scores

<b>Name of compound</b>	<b>Molecular weight (g/mol)</b>	<b>Human PXR Activity (Fold increase in Firefly/Renilla luciferase assay over vehicle control)</b>	<b>MM/GBVI Binding Free Energy (kcal/mol)</b>
Rifampicin	822.94	20	-13.1
Delavirdine	456.56	2.0	-9.1
Etravirine	435.28	17	-7.9
Rilpivirine	366.42	11.5	-7.9
Nevirapine	266.30	1.3	-6.6
Pregnenolone carbonitrile (PCN)	341.49	1.5	-6.5
Efavirenz	315.68	7.0	-6.1



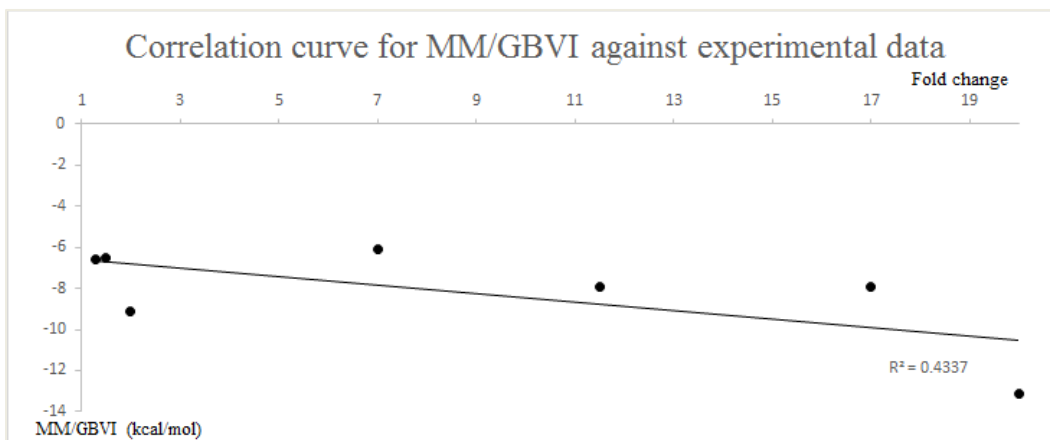


Figure 5.1: Correlation curve for MM/GBVI binding free energy (y-axis) against fold increase in human PXR activity over vehicle control (x-axis).

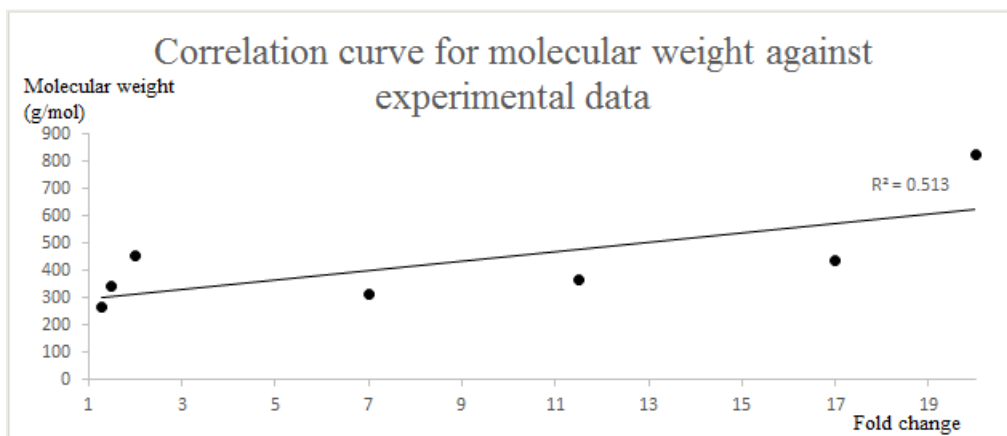


Figure 5.2: Correlation curve for molecular weight (y-axis) against fold increase in human PXR activity over vehicle control (x-axis).

#### 5.4.2 Analysis of interactions

Table 5.2 lists the amino acid residues that are in contact and may form interactions with the selected ligand pose. The molecular dynamics and redocking results are presented for each compound in this section.

Table 5.2: Amino acid residues involved in the binding of compounds to human PXR

<b>Compound</b>	<b>Hydrophobic residues</b>	<b>Polar residues</b>	<b>Figures</b>
Etravirine	Leu 240, Met 243, Phe 281, Phe 288, Met 323, Leu 411, Phe 420, Met 425, Phe 429	Gln 285	5.3, 5.5
Rilpivirine	Val 211, Met 243, Phe 288, Trp 299, Met 323, Leu 324, Leu 411, Ile 414, Phe 420, Met 425, Phe 429	Gln 285	5.6, 5.8
Efavirenz	Leu 240, Met 243, Phe 288, Trp 299	His 407	5.9, 5.11

### Rifampicin

The X-ray structure of rifampicin in human PXR show potential hydrogen bond interactions that may take place between the ester group in rifampicin and Ser 247 in human PXR(Table 5.3). The redocking process is less successful in docking the larger structure of rifampicin taken from PubChem database where all the poses have a RMSD of greater than 2.0 Å.

## Etravirine

Potential hydrogen bond interactions take place between the nitrile group in etravirine and Gln 285(Figures 5.3, 5.4 and Table 5.3).

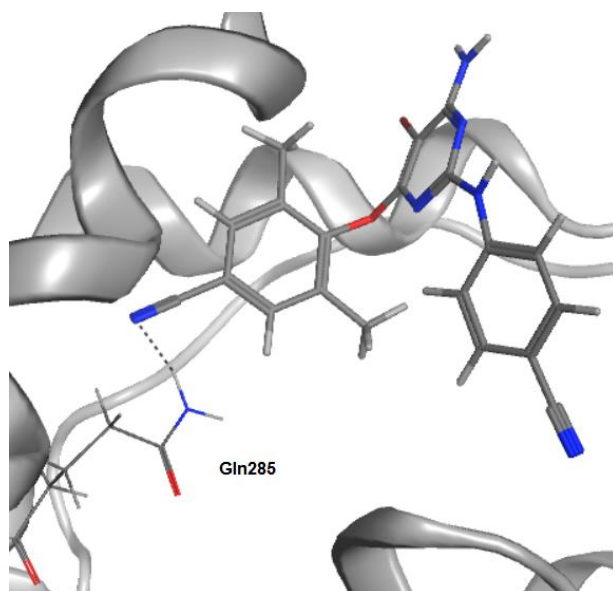


Figure 5.3: Pose of etravirine. Potential interactions are marked by dotted lines.

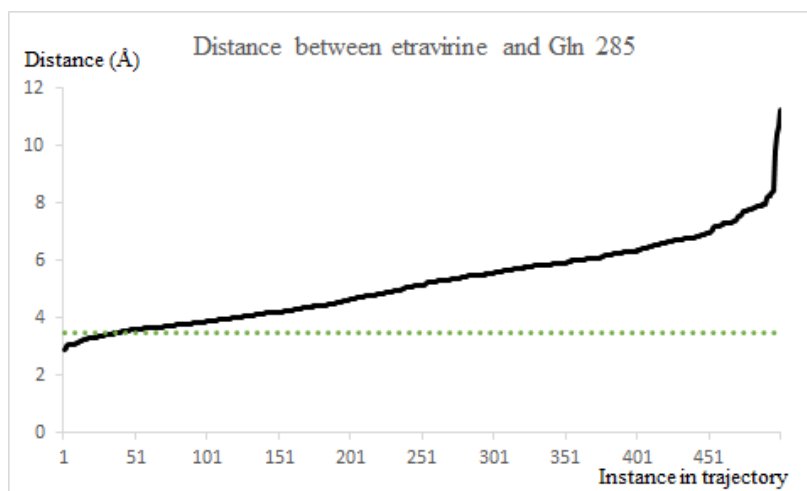


Figure 5.4: Graph of distance between large atom in etravirine and Gln 285 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

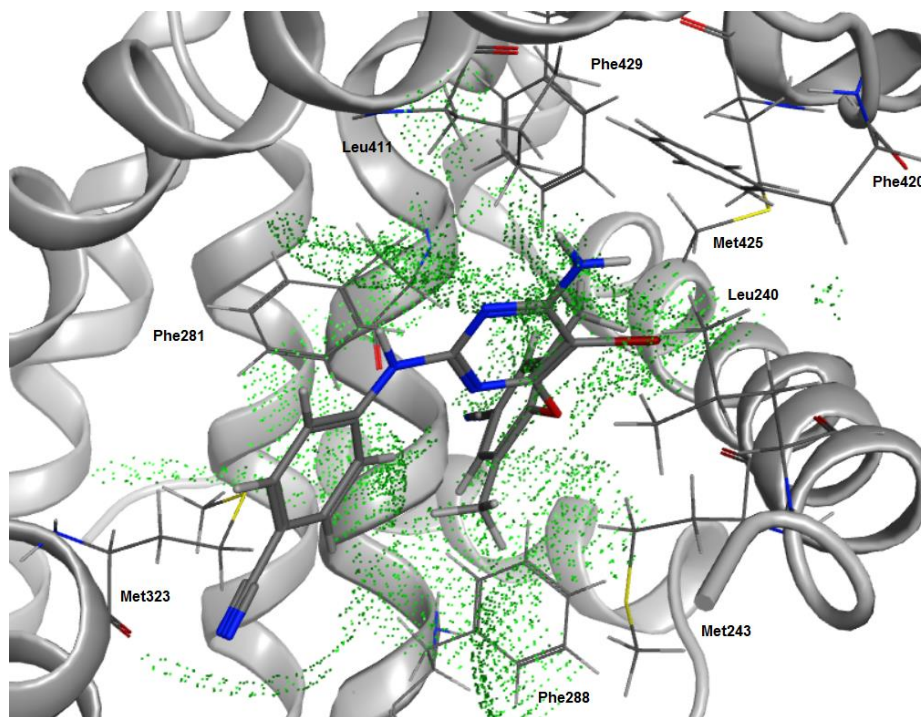


Figure 5.5 Pose of etravirine over ribbon diagram of human PXR. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.

### Rilpivirine

Potential hydrogen bond interactions take place between the nitrile group in rilpivirine and Gln 285(Figures 5.6, 5.7 and Table 5.3).

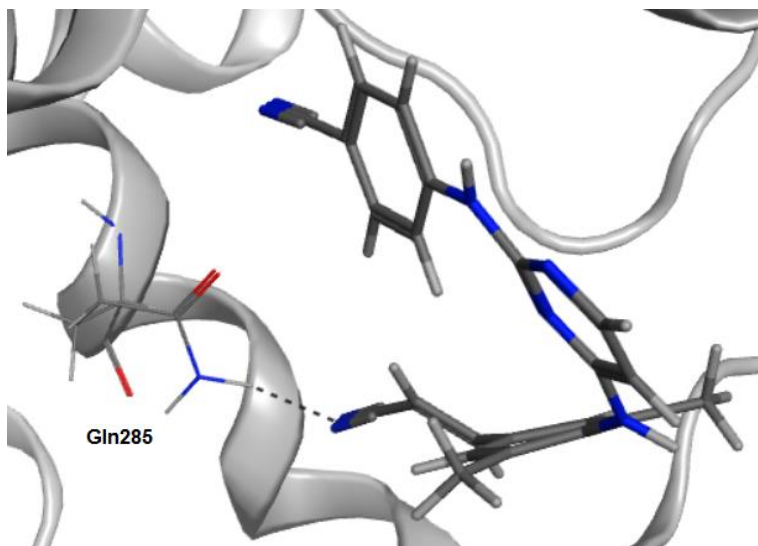


Figure 5.6: Pose of rilpivirine. Potential interactions are marked by dotted lines.

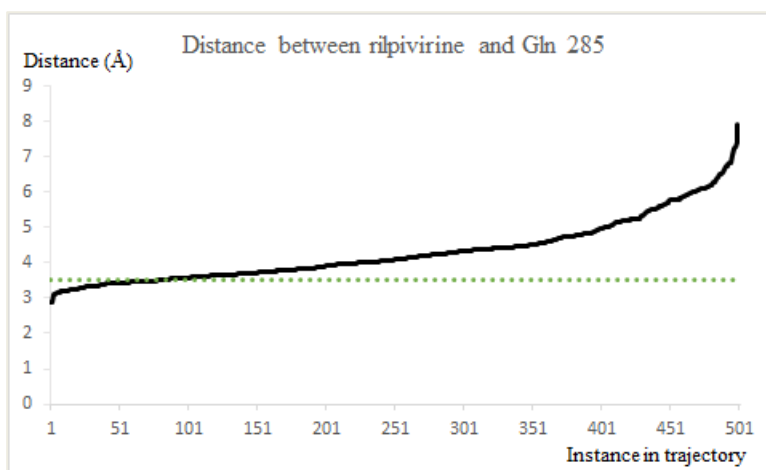


Figure 5.7: Graph of distance between large atom in rilpivirine and Gln 285 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

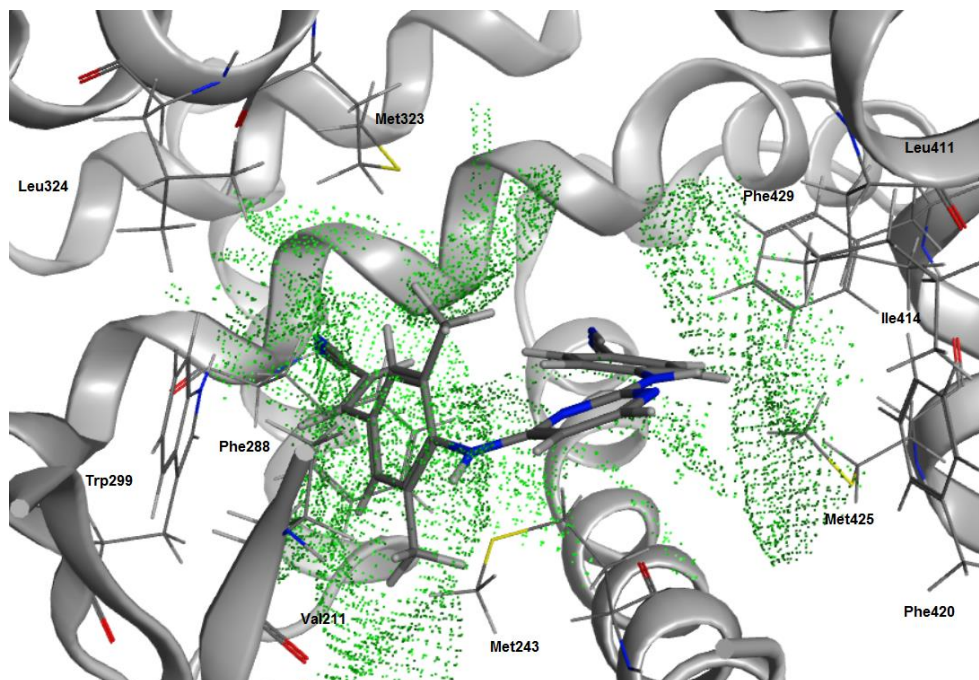


Figure 5.8: Pose of rilpivirine over ribbon diagram of human PXR. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.

### Efavirenz

Potential hydrogen bond interactions take place between the amide group in efavirenz and His 407 in human PXR (Figures 5.9, 5.10 and Table 5.3).

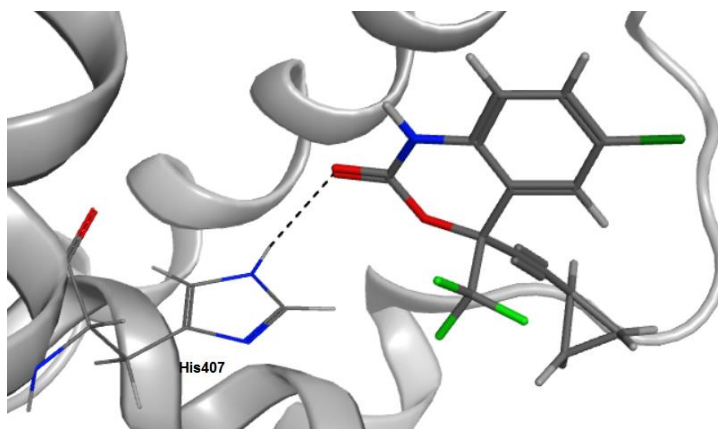


Figure 5.9: Pose of efavirenz. Potential interactions are marked by dotted lines.

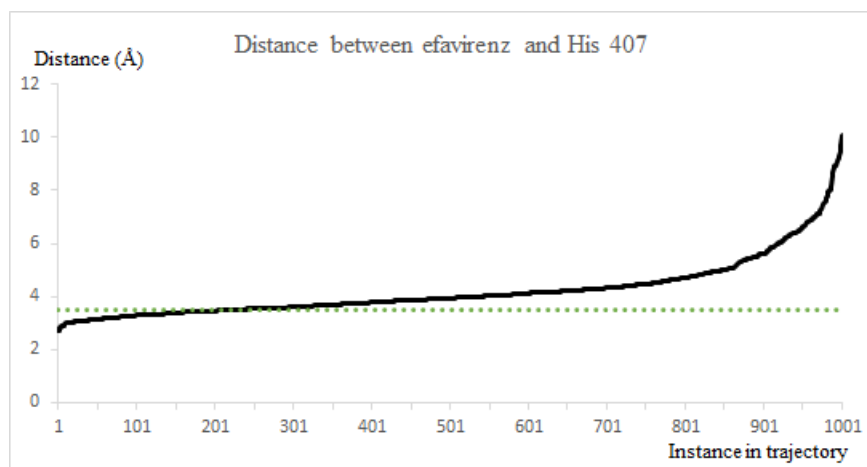


Figure 5.10: Graph of distance between large atom in efavirenz and His 407 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

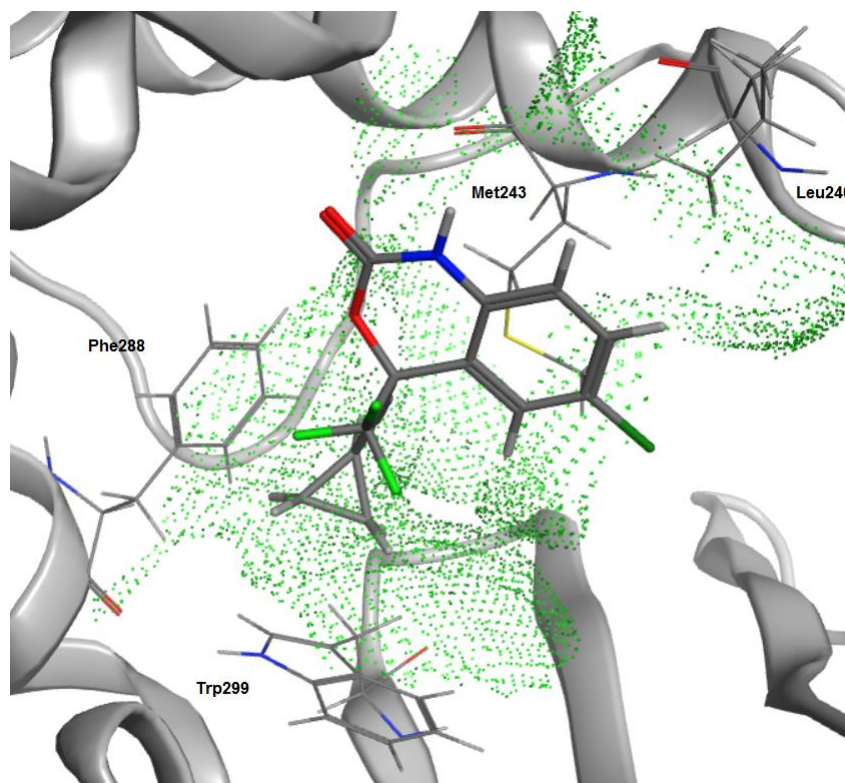


Figure 5.11: Pose of efavirenz over ribbon diagram of human PXR. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.

## Delavirdine

No consistent set of hydrogen bond interactions or pi-interactions are found among poses obtained through clustering in molecular dynamics simulations. The distance between the amide group in delavirdine and His 407 is shown in Figure 5.12 and Table 5.3.

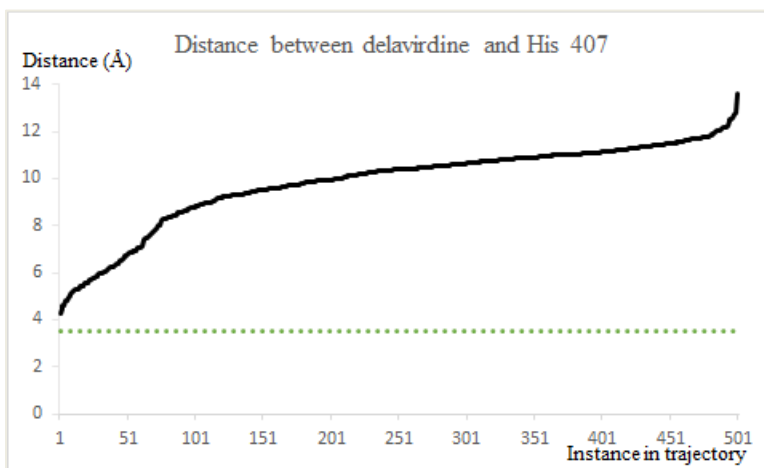


Figure 5.12: Graph of distance between large atom in efavirenz and His 407 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

## Pregnenolone carbonitrile (PCN)

No consistent hydrogen bond interactions of within 3.5 Å in length or aromatic pi interactions are found among the representative poses. The distance between the nitrile group in PCN and His 407 is shown in Figure 5.13 and Table 5.3.



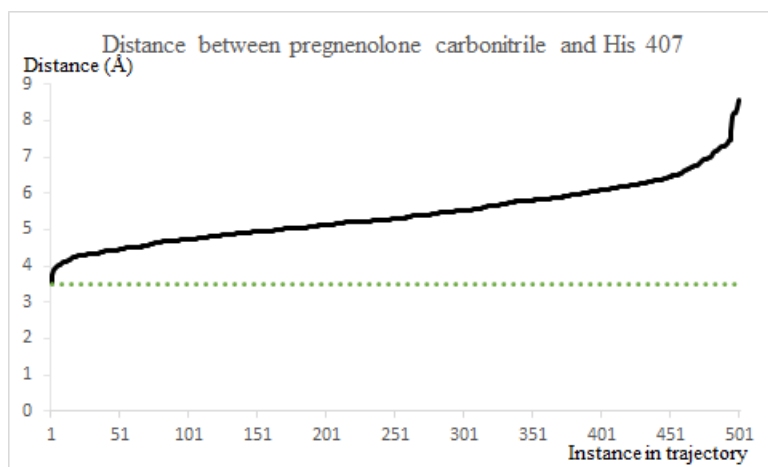


Figure 5.13: Graph of distance between large atom in PCN and His 407 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

### Nevirapine

No consistent hydrogen bond interactions of within 3.5 Å in length or aromatic pi interactions are found among the representative poses. The distance between the amide group in nevirapine and His 407 is shown in Figure 5.14 and Table 5.3.

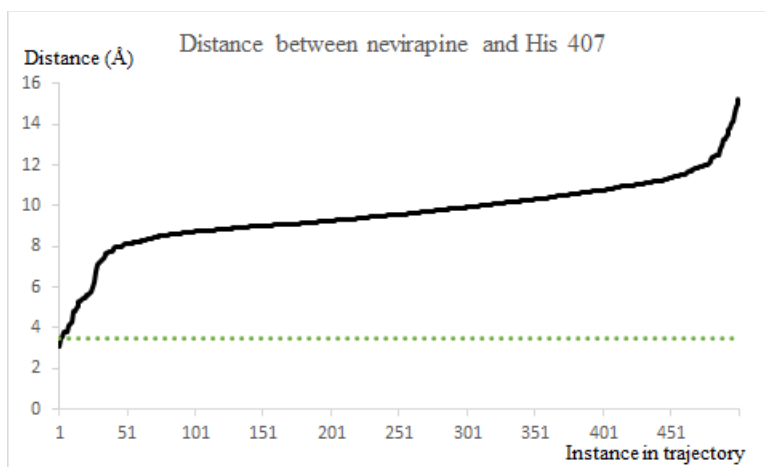


Figure 5.14: Graph of distance between large atom in nevirapine and His 407 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

Table 5.3: Distance between first set of compounds and residue where potential hydrogen bond interactions might be present.

Name of compound	Atom - Residue	Median distance (Å)	Standard deviation	Shortest distance (Å)	Longest distance (Å)	Percentage of trajectory within 3.5 Å
Rifampicin	O – Ser 247	4.6	0.76	2.7	8.1	11
Etravirine	N – Gln 285	5.1	1.4	2.9	11.3	8.0
Rilpivirine	N – Gln 285	4.1	0.89	2.9	7.9	16
Efavirenz	O – His 407	3.9	1.1	2.7	10	22
Delavirdine	O – His 407	10	1.8	4.3	14	0
Pregnenolone carbonitrile	N – His 407	5.3	0.81	3.5	8.6	0
Nevirapine	O – His 407	9.6	1.7	3.1	15	0.40

## 5.5 Conclusion

A structure-based study was carried out on 5 NNRTIs, one antibiotic and one chemical compound on the X-ray crystal structure of human PXR. The workflow was able to filter and rank human PXR activators based on a set of possible binding orientations and docked scores. Additional future work will refine the model by applying a similar method on a larger set of compounds on both human and zebrafish PXR in the next chapter.

## Chapter 6

### Structure-based study on human and zebrafish PXR activators

#### 6.1 Summary

A homology model of zebrafish PXR was created using templates of human PXR from the Protein Databank and the primary structure of zebrafish PXR in the UniProt database. A structure-based study was carried out where 10 approved drugs were docked in the ligand-binding domain of human and zebrafish PXR. Molecular dynamics simulations were carried out on the docked poses and a separate set of docking experiments were carried out to select poses for further analysis of binding surfaces and binding orientations.

#### 6.2 Introduction to zebrafish PXR

While attempts have been made to develop *in vitro* assays to profile efficiently the effects of new compounds on CYP3A4 expression levels, these efforts are made difficult by species-specific effects that have limited the use of animal tissues and cells for testing purposes. Since analysis of the effects of compounds on CYP3A4 gene expression has been largely restricted to laborious assays involving human liver tissue, an alternative *in vivo* study was carried out by Metabolic Profile Research Group in which zebrafish *Danio rerio* was used to

investigate the effects of various approved drugs on CYP3A4 gene expression and PXR activation in zebrafish due to the easier accessibility, lower cost and shorter waiting period in obtaining zebrafish for testing of biological activity[122,124-126]. Although human microsomes are available commercially, a zebrafish animal model is sufficiently small to be used in small assay well plates for high-throughput screening.

Induction of zebrafish CYP3A4 gene expression is similar to human CYP3A4 gene expression for most of the tested approved drugs where most of the compounds are able to induce an increase in CYP3A4 gene expression[126,182]. Two of the tested compounds rifampicin and pioglitazone show differences in biological activity between human and zebrafish and acetaminophen does not appear to induce increased expression of CYP3A4 in both human and zebrafish. Hence, acetaminophen will be used as the negative control compound in this study.

## **6.3 Materials and methods**

### **6.3.1 Biological activity**

Biological activity data was obtained from published studies[182] and from collaboration with Metabolic Profiling Research Group in Department of Pharmacy, National University of Singapore. CYP3A4 gene expression data was

obtained using the P450-Glo™ CYP3A4 assay kit. Rifampicin and dexamethasone were used as the positive controls, acetaminophen was used as the negative control in this study, and seven approved drugs phenytoin, carbamazepine, nafcillin, efavirenz, pioglitazone, prednisone and rufinamide were tested for biological activity[126].

### **6.3.2 Homology modeling of zebrafish PXR**

There are nine X-ray crystallography structures of human pregnane X receptor (PXR)-ligand complexes in the RCSB Protein Data Bank[75,86,109-113]. As rifampicin was chosen as the positive validation compound for this study, the structure (PDB ID: 1SKX) was chosen as that structure has rifampicin as the co-crystallized ligand[86]. For zebrafish PXR, only the primary structure is available in the UniProt database[183]. Therefore, a homology model of zebrafish PXR is created using the PDB structure 1SKX as a template.

Molecular Operating Environment (MOE) 2011.10[153] was used for creating the homology model of zebrafish PXR and for docking the individual compounds to the 1SKX structure. 1SKX was first processed by removing bound ligand and water molecules. Hydrogen atoms were then added, and ionization states were assigned using the Protonate3D function in MOE[152]. The Assisted Model Building with Energy Refinement 99 (AMBER99) forcefield was used to create the homology model, and to carry out energy minimization of the

homology model and the structure 1SKX[154]. The target sequence has an identity value of 41.4%. The Maestro software in Schrödinger 2015-3[184] is used to carry out comparison of amino acid sequence between the homology model of zebrafish PXR and human PXR(Figure 6.1) and to carry out Ramachandran plot of the PDB structures 1SKX(Figure 6.2) and the created zebrafish homology model(Figure 6.3). All non-glycine and non-binding site residues which do not fulfill the dihedral angle criteria[185] are removed from the model where in the zebrafish homology model, the removed residues are Ile 195, Lys 228, His 359 and Ser 433.

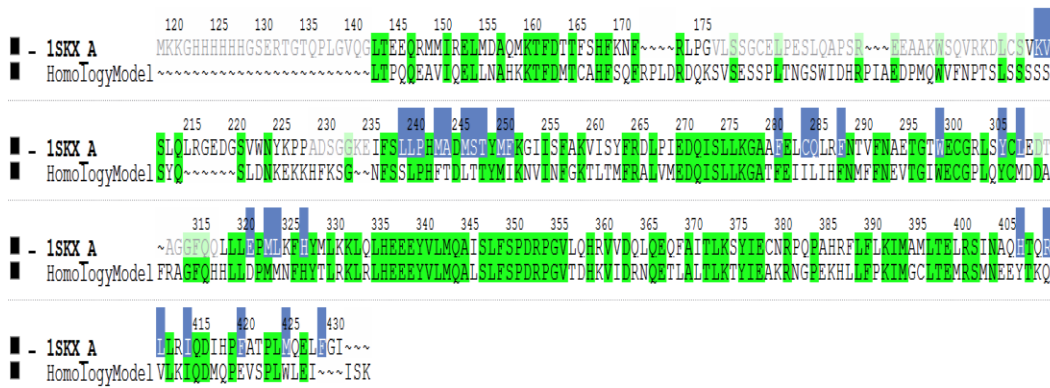


Figure 6.1: Amino acid sequence for homology model of zebrafish PXR and human PXR using PDB structure 1SKX as template. Conserved residues are highlighted in green and important residues that are reported to activate human PXR are highlighted in blue.

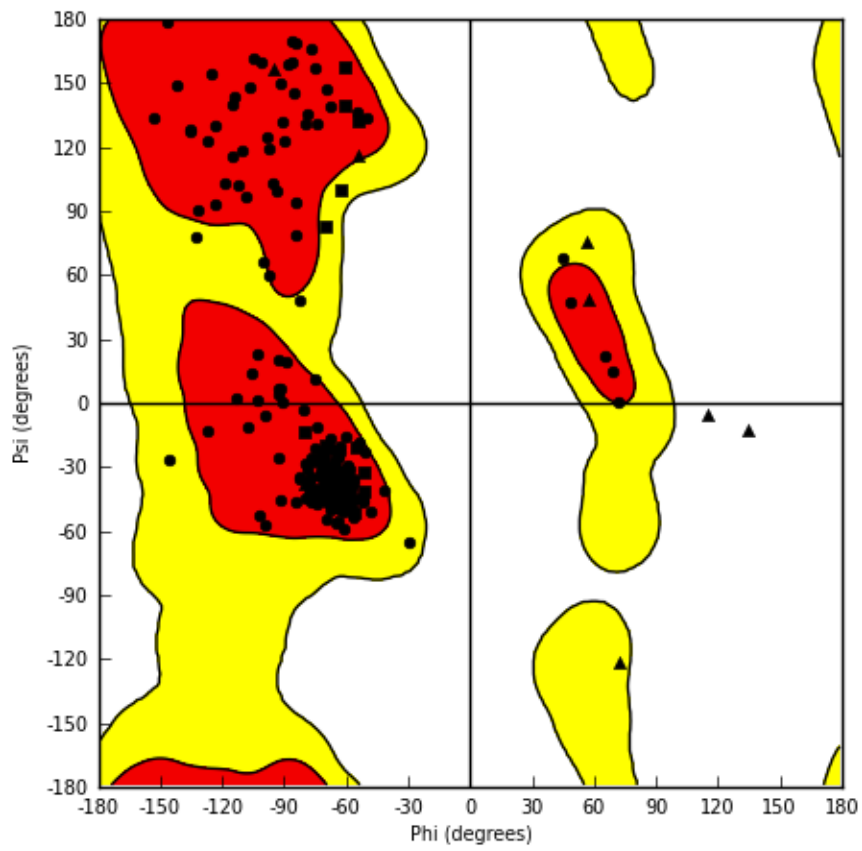


Figure 6.2: Ramachandran plot of X-ray structure of human PXR.

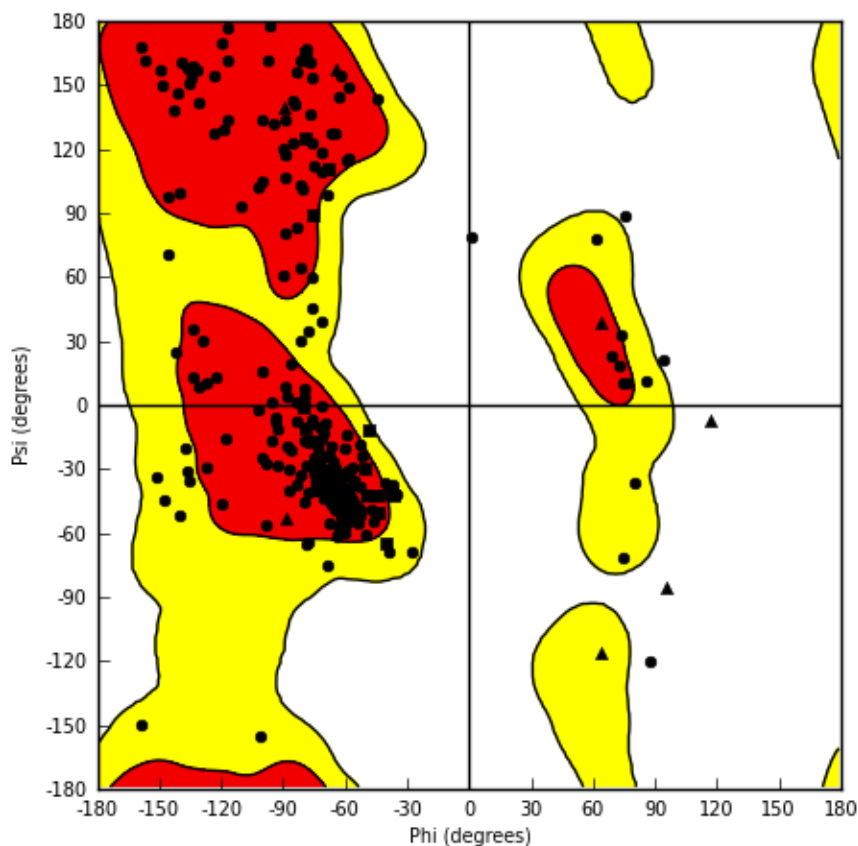


Figure 6.3: Ramachandran plot of homology model of zebrafish PXR.

### 6.3.3 Structure-based study of compounds

An initial docking process similar to section 5.3.2 was carried out in this phase of the research project as it was able to accurately dock the positive validation compound rifampicin within 2.0 Å of its X-ray structure and also rank the compound accurately as a strong human PXR activator. The molecular mechanics generalized born/volume integral (MM/GBVI) binding free energy is calculated as a final pose rescoring function to refine the docking pose. Five hundred possible poses were determined for each molecule[156-158].



The individual atom coordinates of each ligand's possible docked pose were then recorded and hierarchical clustering using the Ward method[186,187] was carried out on each ligand to determine the most likely poses that were predicted to occur. The representative pose from the most populated cluster was selected for further analysis.

The pose was superimposed on the binding pocket of zebrafish PXR to compare their structural interaction fingerprint. Energy minimization was carried out to refine the docked pose if there is a clash with any residue in the binding pocket. JMP Version 10[188] was used to perform the hierarchical clustering to choose representative poses from the most populated cluster for each docked compound.

Molecular dynamics simulation, docking and analysis of binding site surface was carried out similar to the workflow in Figure 3.4.

## **6.4 Results and Discussion**

### **6.4.1 Structures of human and zebrafish PXR**

Table 6.1 below lists the amino acid residues which surround the ligand-binding domain of human and zebrafish PXR.

Table 6.1: List of amino acid residues that form the ligand-binding domain in human PXR and zebrafish PXR. Important residues that are reported to activate human PXR are highlighted in bold.

<b>Residue in human PXR</b>	<b>Corresponding Residue in zebrafish PXR</b>
Val 211	Thr
Leu 240	Leu
Met 243	Phe
<b>Ser 247</b>	Thr
Phe 251	Ile
Phe 281	Phe
Cys 284	Ile
<b>Gln 285</b>	Leu
<b>Phe 288</b>	Phe
<b>Trp 299</b>	Trp
<b>Tyr 306</b>	Tyr
Leu 308	Met
Glu 321	Asp
Met 323	Met
Leu 324	Met
His 327	His
<b>His 407</b>	Tyr
<b>Arg 410</b>	Gln
Leu 411	Val
Ile 414	Ile

Phe 420	Glu
Met 425	Trp
Phe 429	Ile

#### 6.4.2 Initial docking results

Due to the large number of poses that were predicted by docking for each compound, an analysis of the interaction profile was carried out based on the top-scoring docked pose, as well as through carrying out hierarchical clustering using the Ward method[186,187] to filter and refine the number of docked poses for further analysis. To validate the docking and hierarchical clustering process, the structure of rifampicin obtained from PDB database was redocked into the same ligand binding domain and the top-scoring pose and the binding pose from the most populated cluster had a root mean squared distance (RMSD) of 1.0 Å with the original binding position in PDB structure 1SKX. This shows that the docking process was able to dock rifampicin as a positive control accurately onto the ligand-binding domain. Table 6.2 lists the docked score of the representative pose from the most populated cluster.

Table 6.2: List of docked scores

Name of compound	Human PXR		Zebrafish PXR	
	Activity	MM/GBVI Binding Free Energy (kcal/mol)	Activity	MM/GBVI Binding Free Energy (kcal/mol)
Rifampicin	Strong inducer	-13.1	Weak/non-inducer	-9.2
Pioglitazone	Weak/non-inducer	-6.9	Strong inducer	-7.3
Nafcillin	Moderate inducer	-6.7	Moderate inducer	-6.7
Prednisone	Weak/non-inducer	-6.7	Weak/non-inducer	-6.6
Dexamethasone	Strong inducer	-6.4	Strong inducer	-6.4
Efavirenz	Moderate inducer	-6.1	Moderate inducer	-5.5
Rufinamide	Weak/non-inducer	-5.4	Weak/non-inducer	-5.5
Phenytoin	Strong inducer	-5.3	Strong inducer	-5.4

Carbamazepine	Strong inducer	-5.1	Strong inducer	-4.7
Acetaminophen	Weak/non-inducer	-4.6	Weak/non-inducer	-4.6

The docking and clustering process was able to rank and classify most of the compounds as possible activators. However, it ranked certain weak or non-activators as strong activators and vice versa. Additional further study through molecular dynamics simulations and analysis of structural interactions would be required.

### 6.4.3 Analysis of interactions

Table 6.2 lists the amino acid residues that are in contact and may form interactions with the selected ligand pose. The molecular dynamics and redocking results are presented for each compound in this section.

Table 6.3: Amino acid residues involved in the binding of compounds to human PXR.

Compound	Hydrophobic residues	Polar residues	Figures
Phenytoin	Val 211, Met 243, Phe 281, Phe 288, Trp 299, Met 323	Gln 285	6.6, 6.8

Carbamazepine	Met 243, Phe 281, Phe 288, Trp 299, Met 323, Leu 411, Met 425, Phe 429	Ser 247	6.12, 6.14
Dexamethasone	Val 211, Leu 240, Met 243, Phe 251, Phe 281, Phe 288, Trp 299, Met 323, Leu 411	Gln 285	6.18 , 6.20
Nafcillin	Val 211, Leu 240, Met 243, Trp 299, Leu 308, Met 323, Leu 324, Leu 411, Ile 414, Phe 420	Gln 285	6.24, 6.26
Prednisone	Val 211, Leu 240, Met 243, Phe 281, Trp 299, Met 323, Leu 324, Leu 411, Ile 414	Gln 285	6.37, 6.39

Table 6.4: Amino acid residues involved in the binding of compounds to zebrafish PXR.

<b>Compound</b>	<b>Hydrophobic residues</b>	<b>Figures</b>
Phenytoin	Phe 243, Phe 288, Trp 299	6.9, 6.11
Carbamazepine	Thr 211, Phe 243, Phe 288, Trp 299, Met 323, Met 324	6.15, 6.17
Dexamethasone	Thr 211, Leu 240, Phe 243, Phe 281, Phe 288	6.21, 6.23
Nafcillin	Thr 211, Let 240, Phe 243, Met 323, Met 324	6.27, 6.29
Efavirenz	Thr 211, Leu 240, Phe 243, Met 323, Met 324	6.30, 6.32

Pioglitazone	Thr 211, Leu 240, Phe 243, Thr 247, Trp 299, Val 411, Ile 414, Trp 425	6.33, 6.35
Prednisone	Thr 211, Leu 240, Phe 243, Trp 299, Met 323, Met 324	6.40, 6.42

### Rifampicin

The X-ray structure of rifampicin in human PXR show potential hydrogen bond interactions that may take place between the ester group in rifampicin and Ser 247 in human PXR(Table 5.3). Molecular dynamics simulations show rifampicin in human PXR have a larger percentage of instances within hydrogen bond distance to Ser 247 in the trajectory(Figure 6.4) compared to rifampicin in zebrafish PXR with the equivalent residue Thr 247(Figure 6.5). The redocking process is less successful in docking the larger structure of rifampicin taken from PubChem database in both human and zebrafish PXR. In human PXR, poses which meet a cut-off filter of having a docking score of -5 or less are obtained, however all the poses have a RMSD of greater than 2.0 Å. In zebrafish PXR, all poses do not meet the cut-off filter and have a docking score of greater than -5.

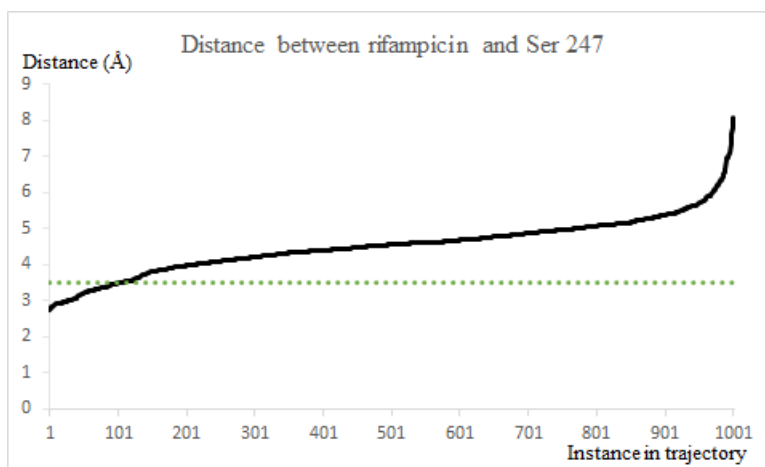


Figure 6.4: Graph of distance between large atom in rifampicin and Ser 247 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

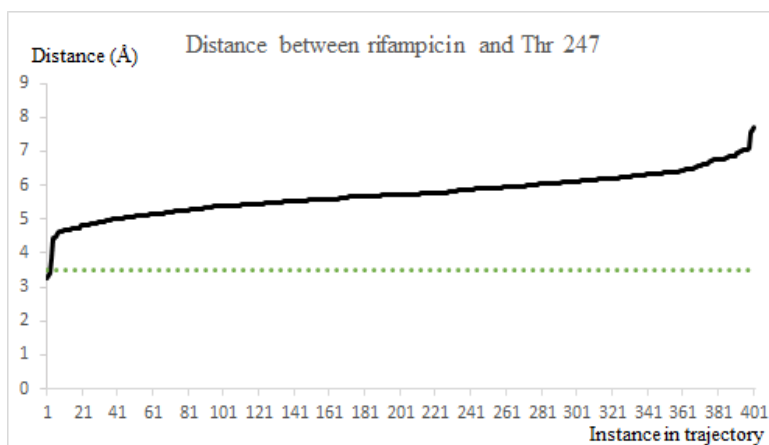


Figure 6.5: Graph of distance between large atom in rifampicin and Thr 247 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

### Phenytoin

Potential hydrogen bond interactions occur between the amide group in phenytoin and Gln 285 in human PXR (Figures 6.6 and 6.7). In zebrafish PXR, potential pi interactions occur between the aromatic rings in phenytoin and the



conserved residue Phe 288(Figures 6.9 and 6.10) . Hence, phenytoin was predicted to activate both human and zebrafish PXR.

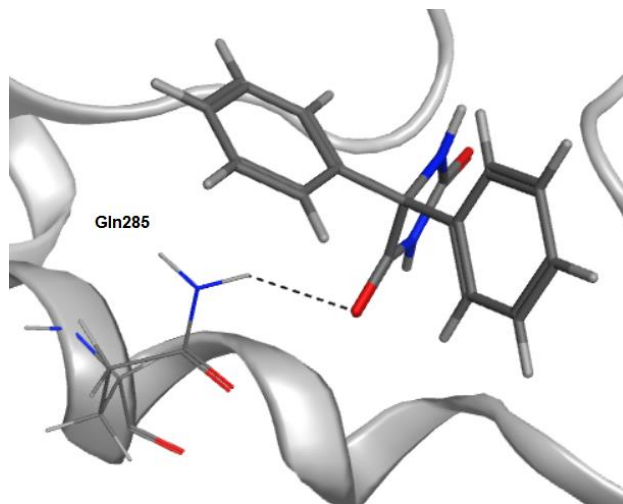


Figure 6.6: Pose of phenytoin in human PXR. Potential interactions are marked by dotted lines.

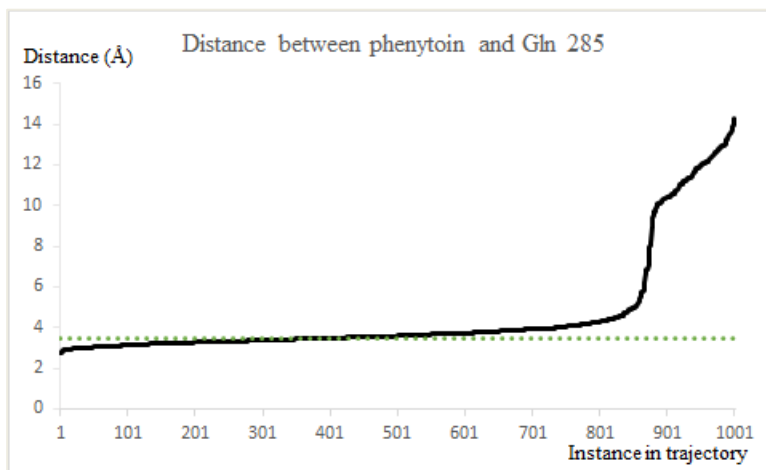


Figure 6.7: Graph of distance between large atom in phenytoin and Gln 285 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

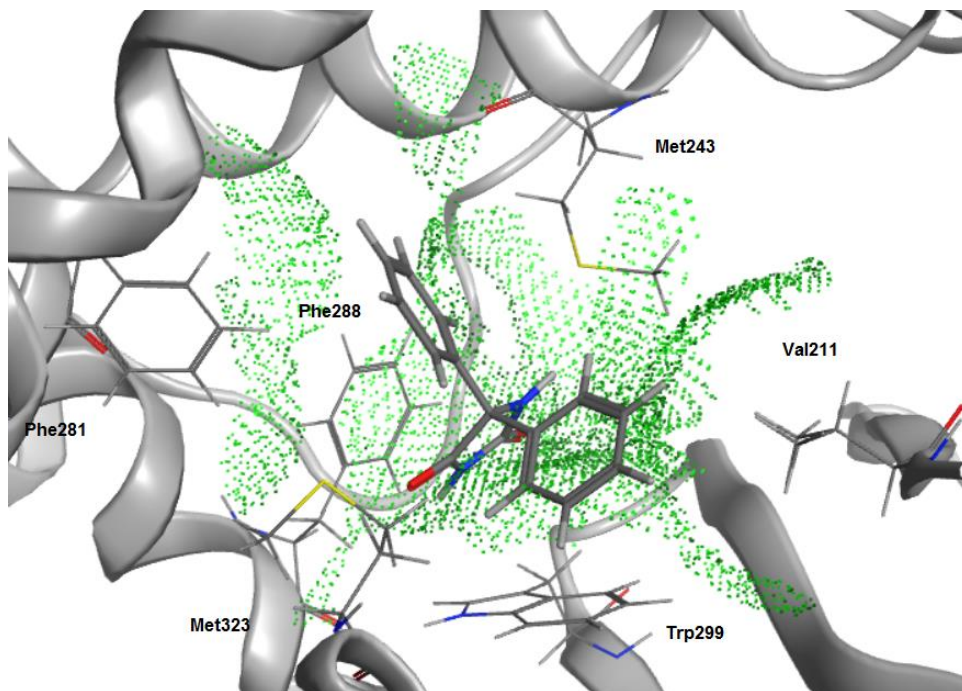


Figure 6.8: Pose of phenytoin over ribbon diagram of human PXR. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.

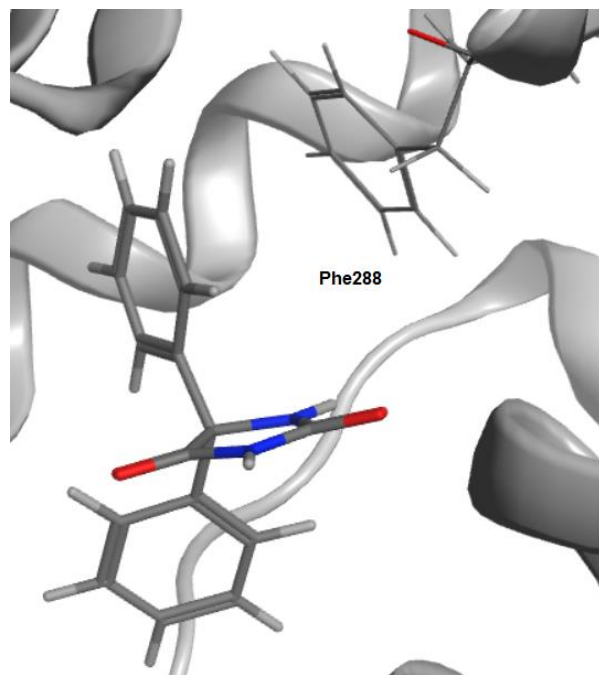


Figure 6.9: Pose of phenytoin in zebrafish PXR.

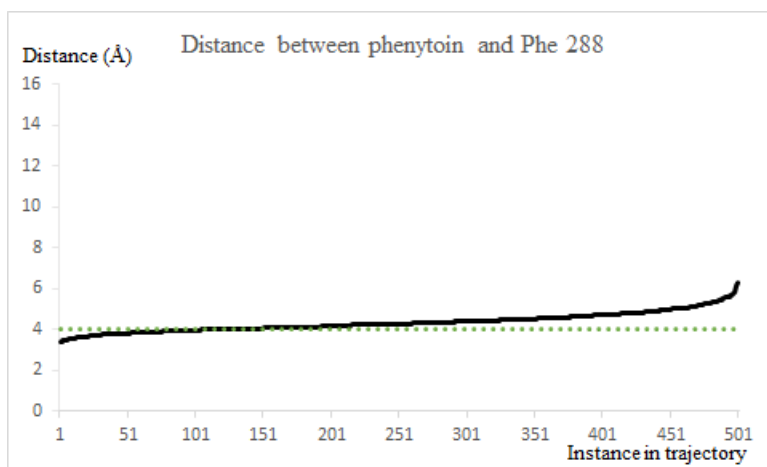


Figure 6.10: Graph of distance between geometric center of aromatic ring in phenytoin and Phe 288 against instance in trajectory. A 4.0 angstrom distance cutoff is shown as a dotted line.

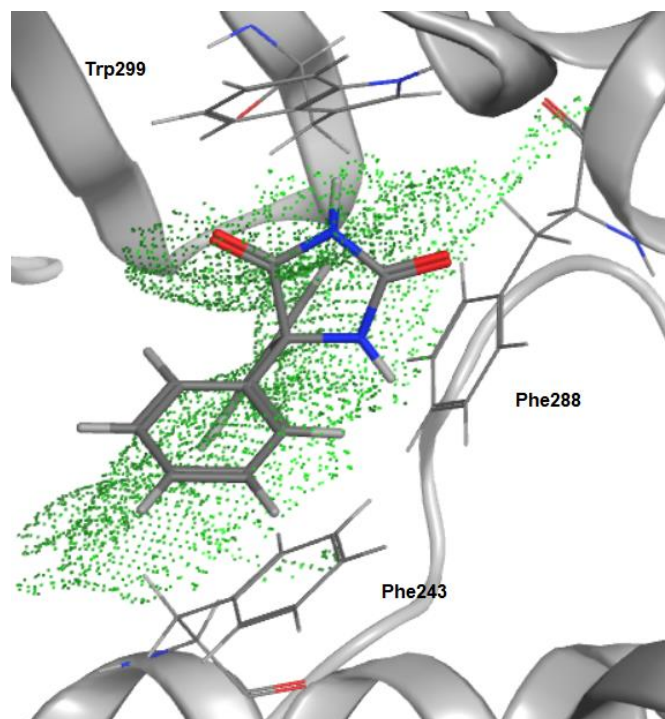


Figure 6.11: Pose of phenytoin over ribbon diagram of zebrafish PXR. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.

## Carbamazepine

Potential hydrogen bond interactions occur between the amide group in carbamazepine and Ser 247 in human PXR(Figures 6.12 and 6.13) . In zebrafish PXR, potential pi interactions occur between the aromatic rings in carbamazepine and the conserved residue Phe 288(Figures 6.15 and 6.16). Hence, carbamazepine was predicted to activate both human and zebrafish PXR.

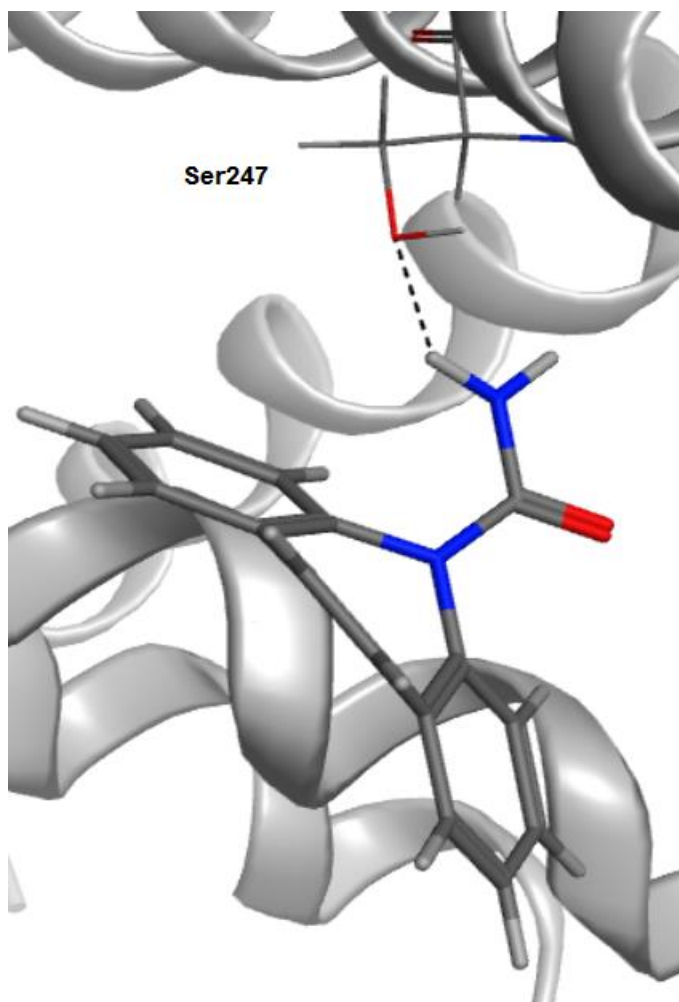


Figure 6.12: Pose of carbamazepine in human PXR. Potential interactions are marked by dotted lines.

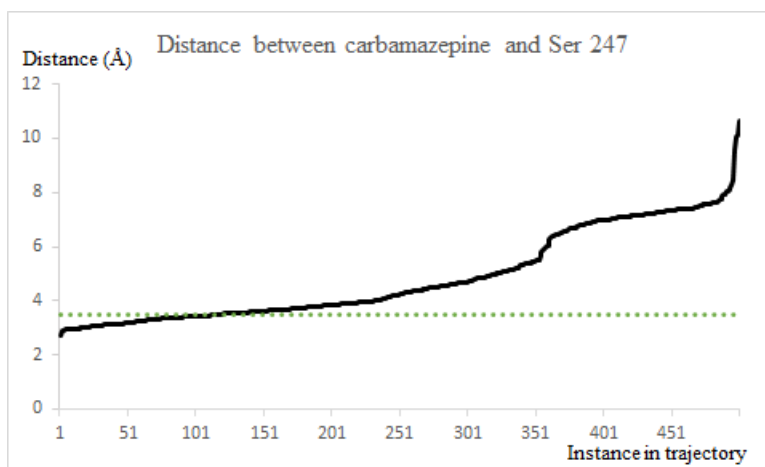


Figure 6.13: Graph of distance between large atom in carbamazepine and Ser 247 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

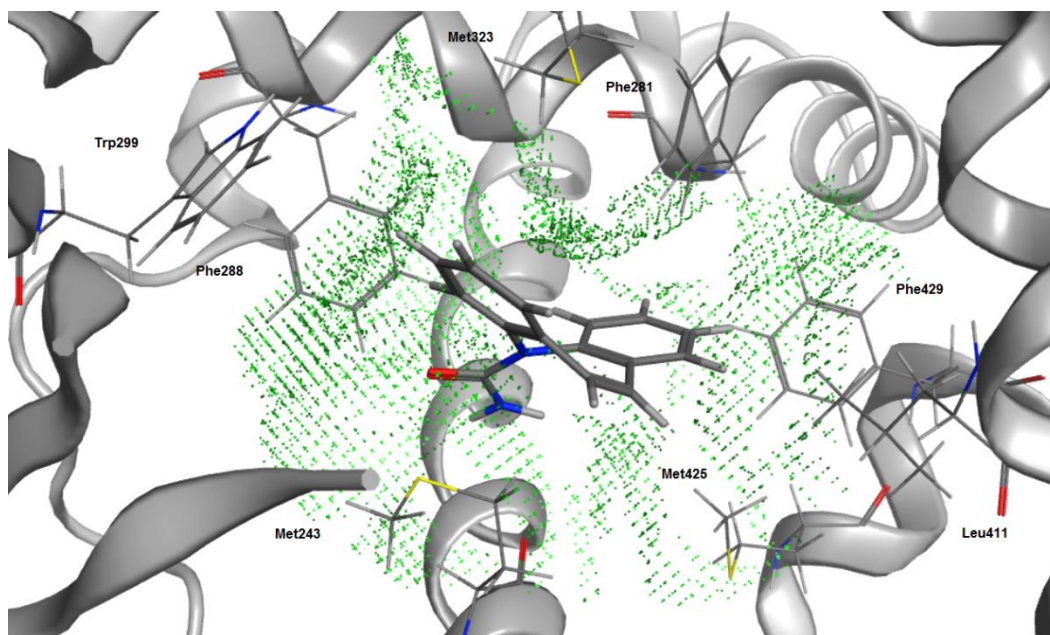


Figure 6.14: Pose of carbamazepine over ribbon diagram of human PXR. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.

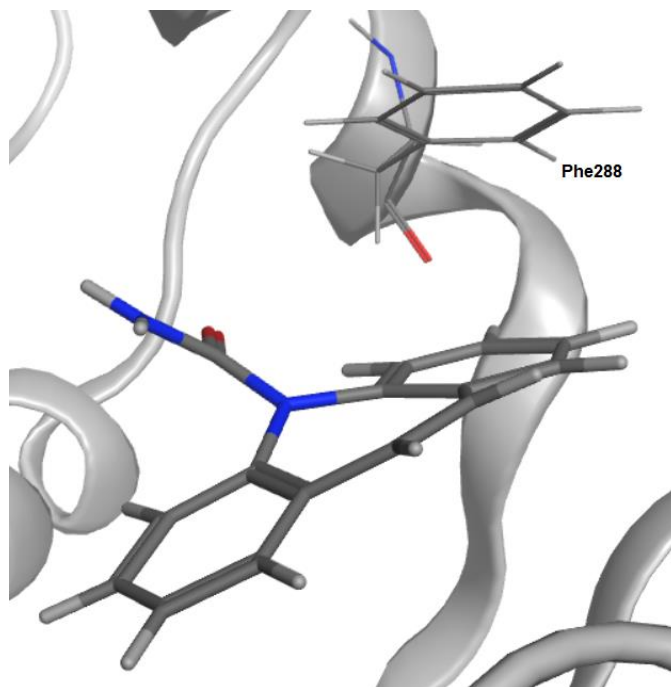


Figure 6.15: Pose of carbamazepine in zebrafish PXR.

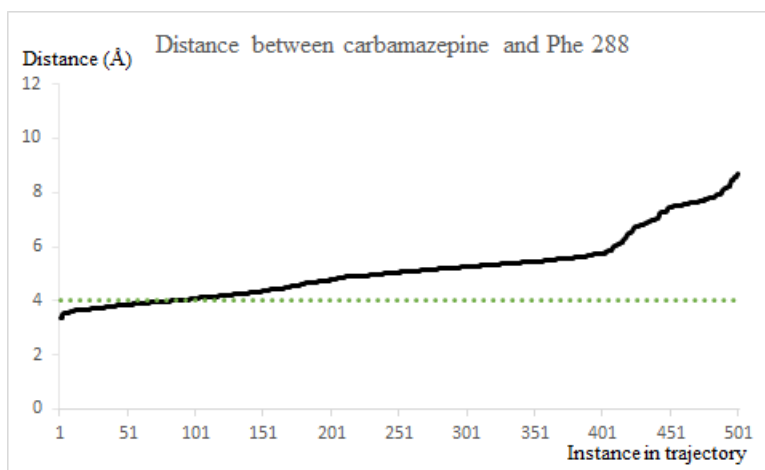


Figure 6.16: Graph of distance between large atom in carbamazepine and Phe 288 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

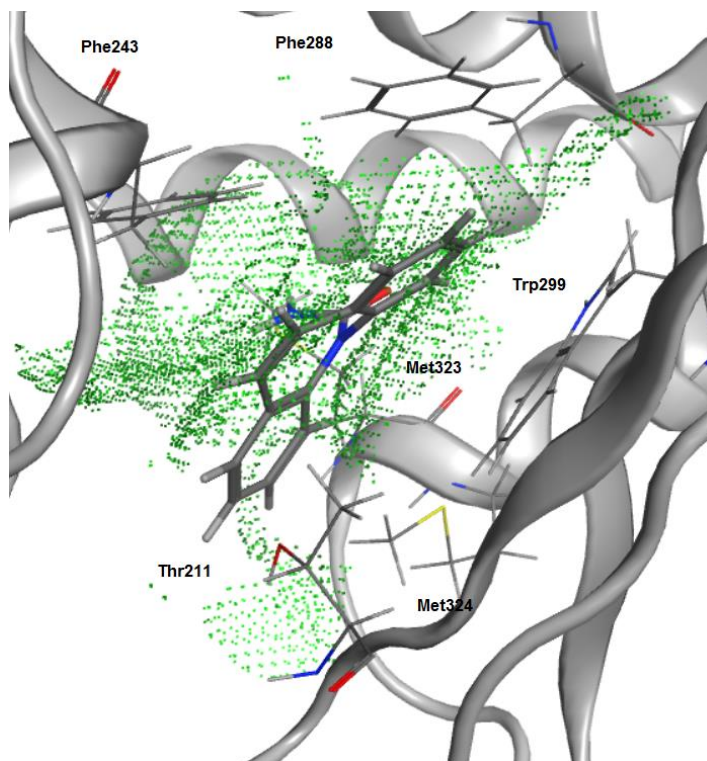


Figure 6.17: Pose of carbamazepine over ribbon diagram of zebrafish PXR. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.

### Dexamethasone

Potential hydrogen bonds occur between the hydroxyl groups in dexamethasone and Gln 285 in human PXR(Figures 6.18 and 6.19). In zebrafish PXR, potential hydrogen bonds occur between the hydroxyl groups in dexamethasone and the conserved residue Met 323(Figures 6.21 and 6.22).

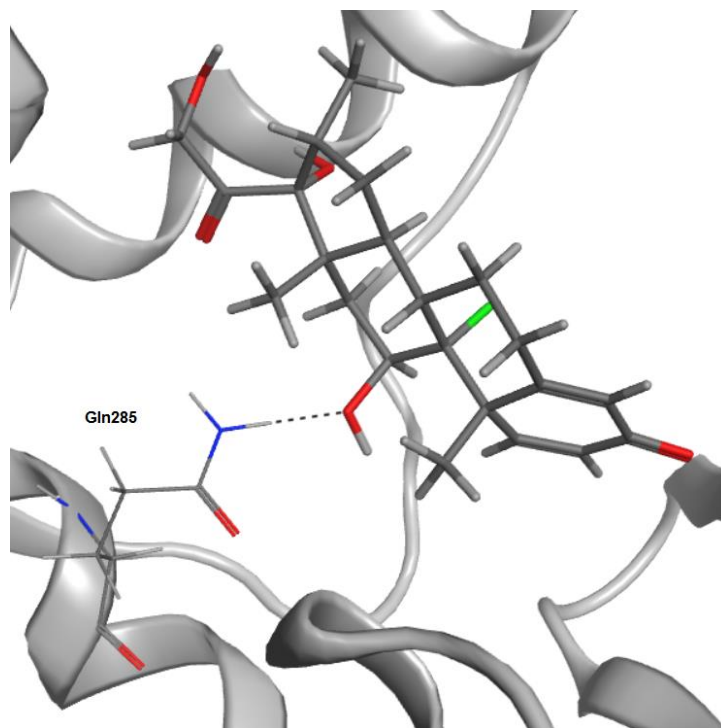


Figure 6.18: Pose of dexamethasone in human PXR. Potential interactions are marked by dotted lines.

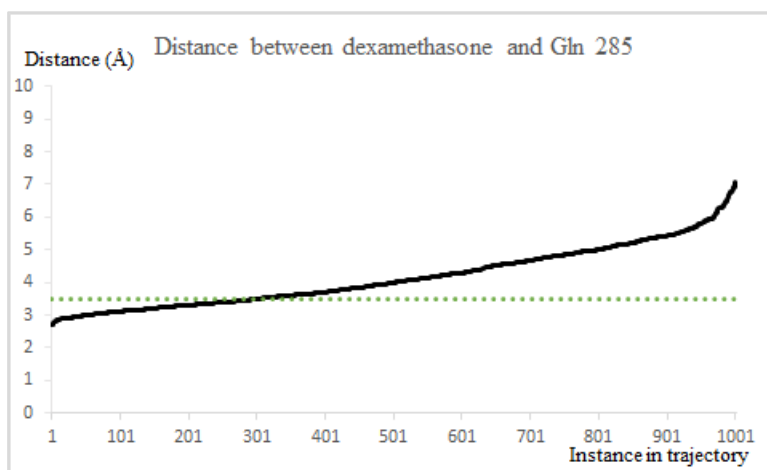


Figure 6.19: Graph of distance between large atom in dexamethasone and Gln 285 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.



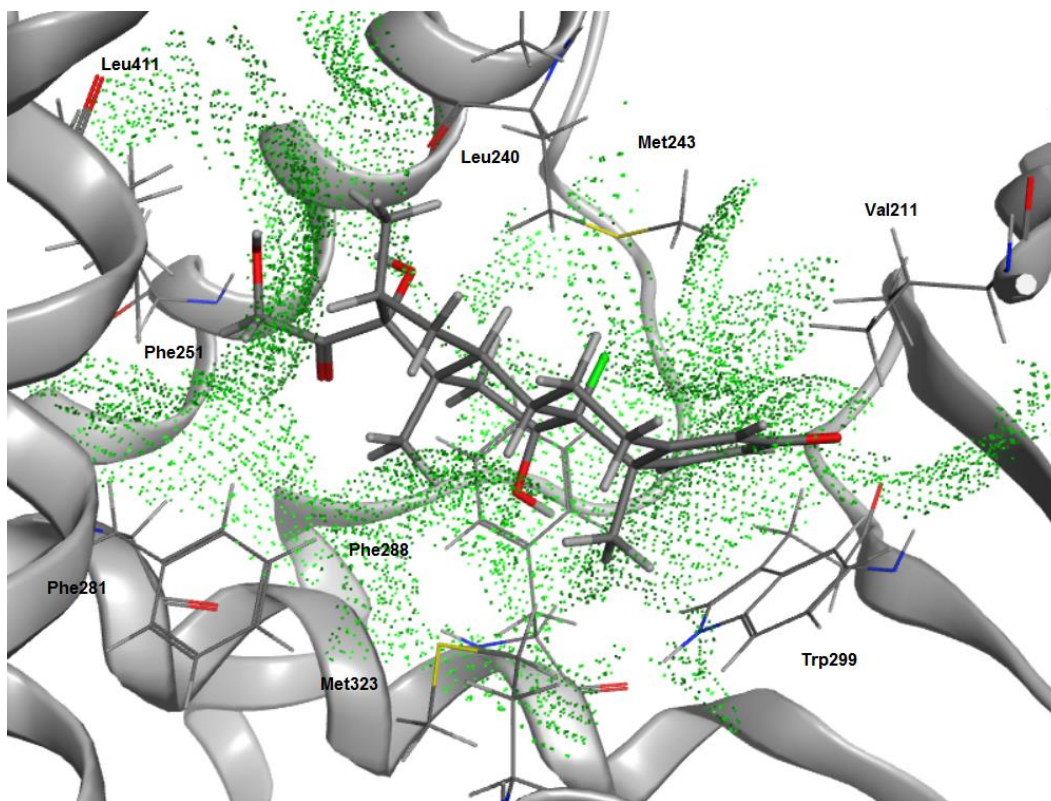


Figure 6.20: Pose of dexamethasone over ribbon diagram of human PXR. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.

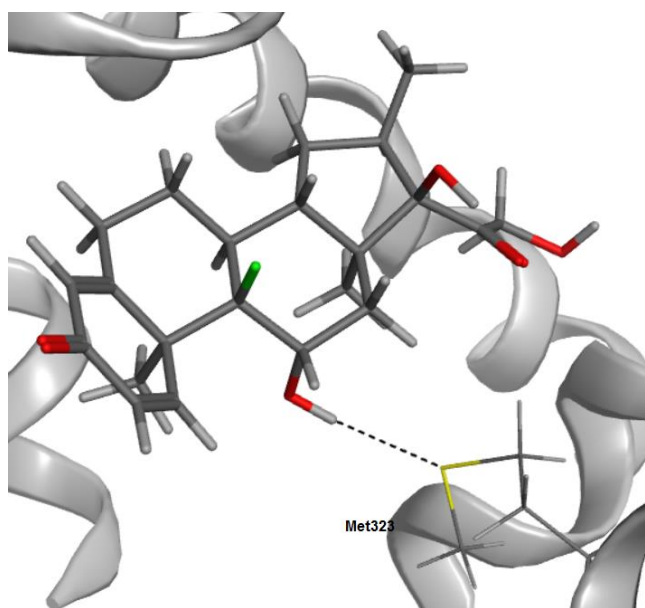


Figure 6.21: Pose of dexamethasone in zebrafish PXR. Potential interactions are marked by dotted lines.

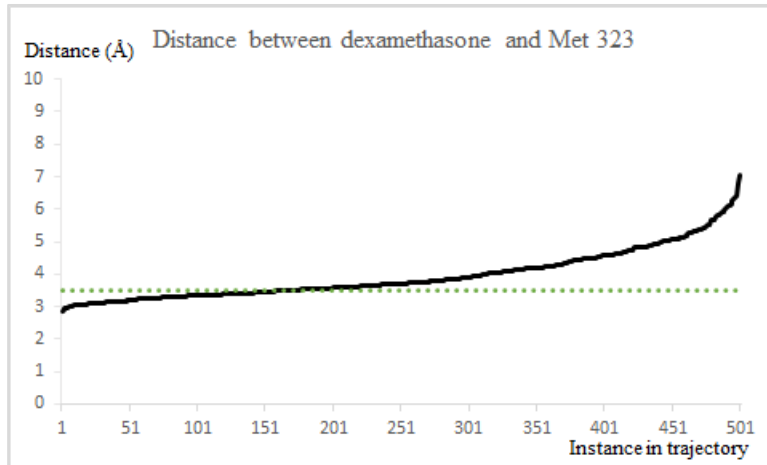


Figure 6.22: Graph of distance between large atom in dexamethasone and Met 323 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

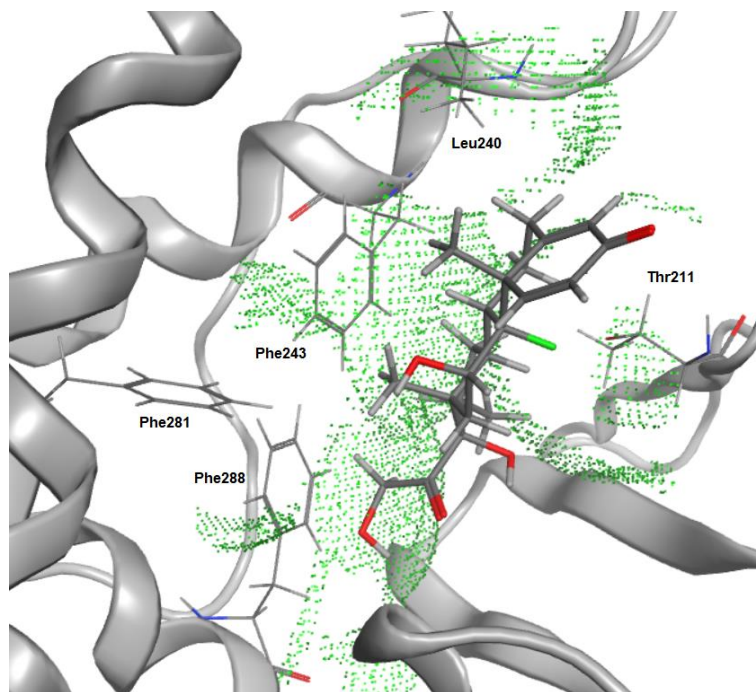


Figure 6.23: Pose of dexamethasone over ribbon diagram of zebrafish PXR. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.

## Nafcillin

Potential hydrogen bonds occur between the amide group in nafcillin and Gln 285 in human PXR(Figures 6.24 and 6.25). In zebrafish, potential hydrogen bonds occur between the amide group and Met 323(Figures 6.27 and 6.28).

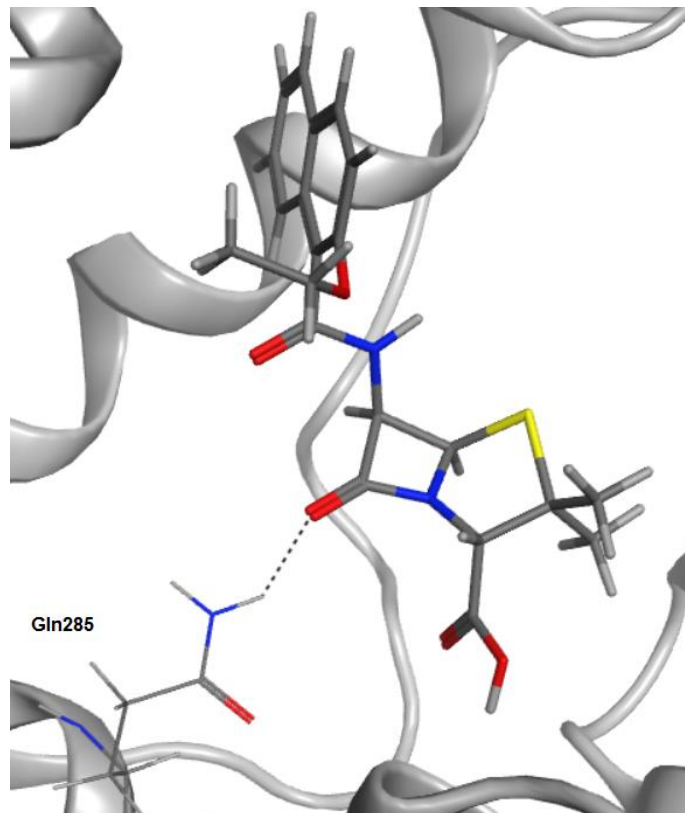


Figure 6.24: Pose of nafcillin in human PXR. Potential interactions are marked by dotted lines.

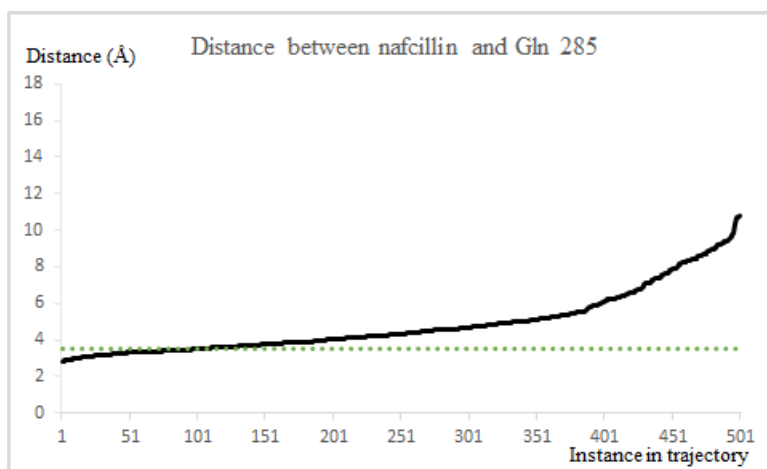


Figure 6.25: Graph of distance between large atom in nafcillin and Gln 285 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

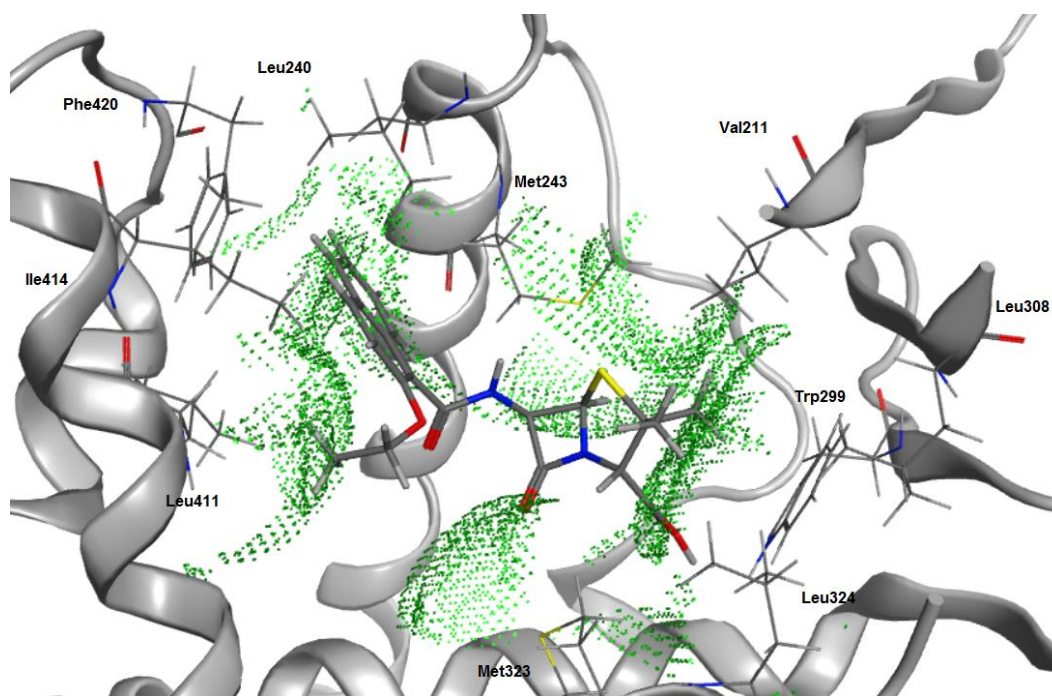


Figure 6.26: Pose of nafcillin over ribbon diagram of human PXR. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.

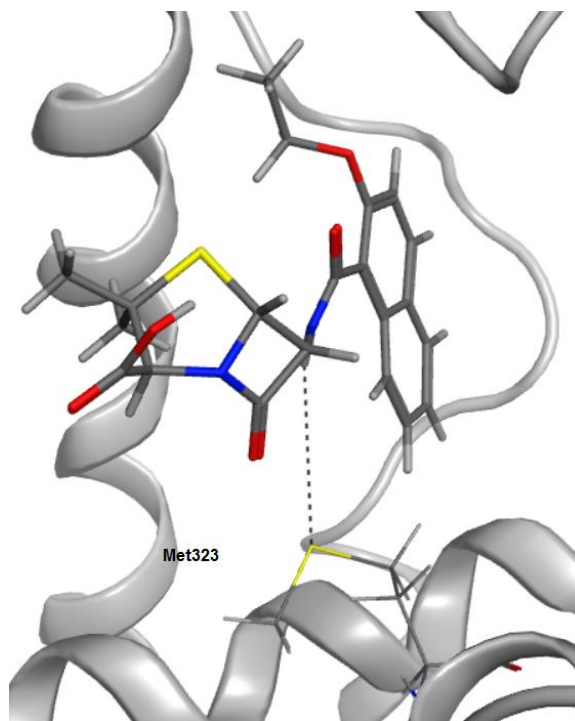


Figure 6.27: Pose of nafcillin in zebrafish PXR. Potential interactions are marked by dotted lines.

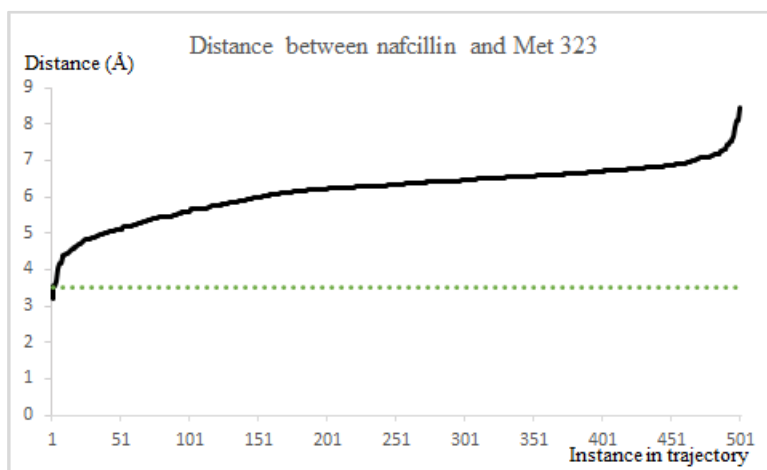


Figure 6.28: Graph of distance between large atom in nafcillin and Met 323 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

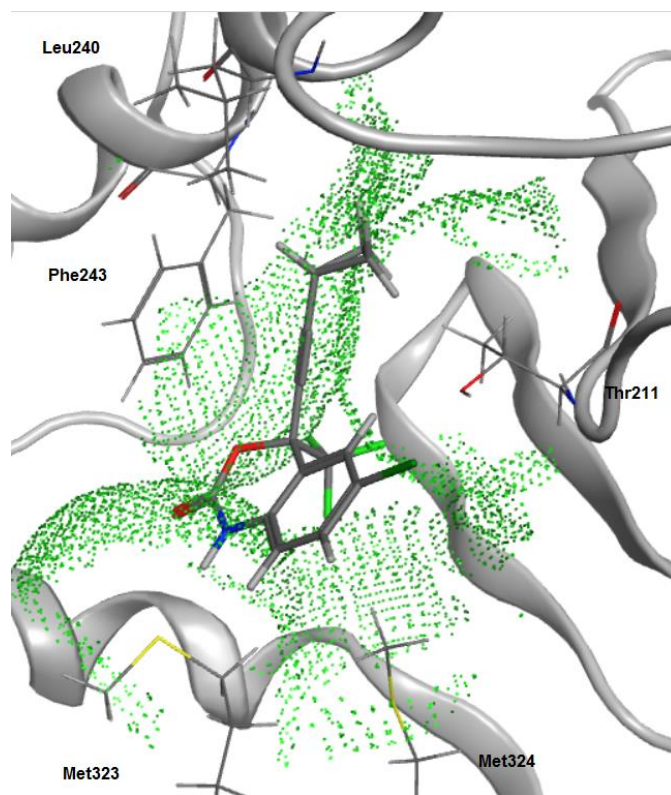


Figure 6.29: Pose of nafcillin over ribbon diagram of zebrafish PXR. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.

### Efavirenz

In zebrafish PXR, potential hydrogen bonds occur between the amide group in efavirenz and the conserved residue Met 323(Figure 6.30 and 6.31).

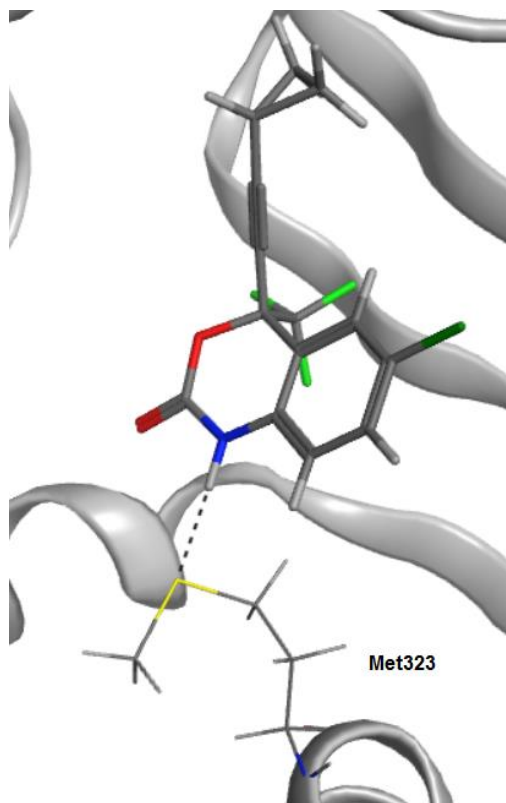


Figure 6.30: Pose of efavirenz in zebrafish PXR. Potential interactions are marked by dotted lines.

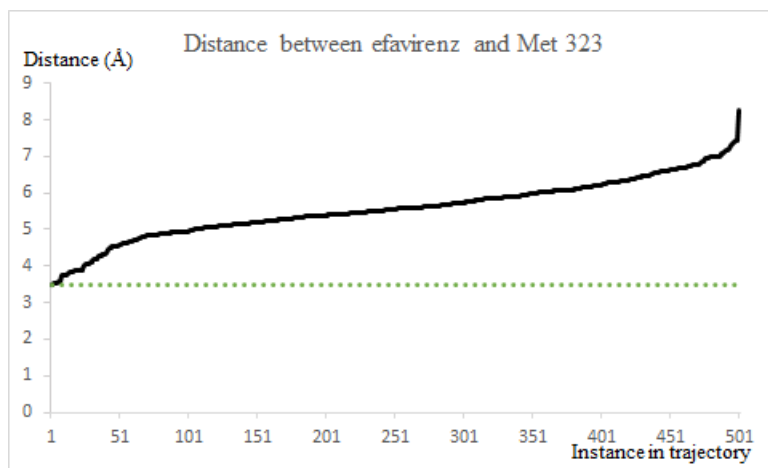


Figure 6.31: Graph of distance between large atom in efavirenz and Met 323 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

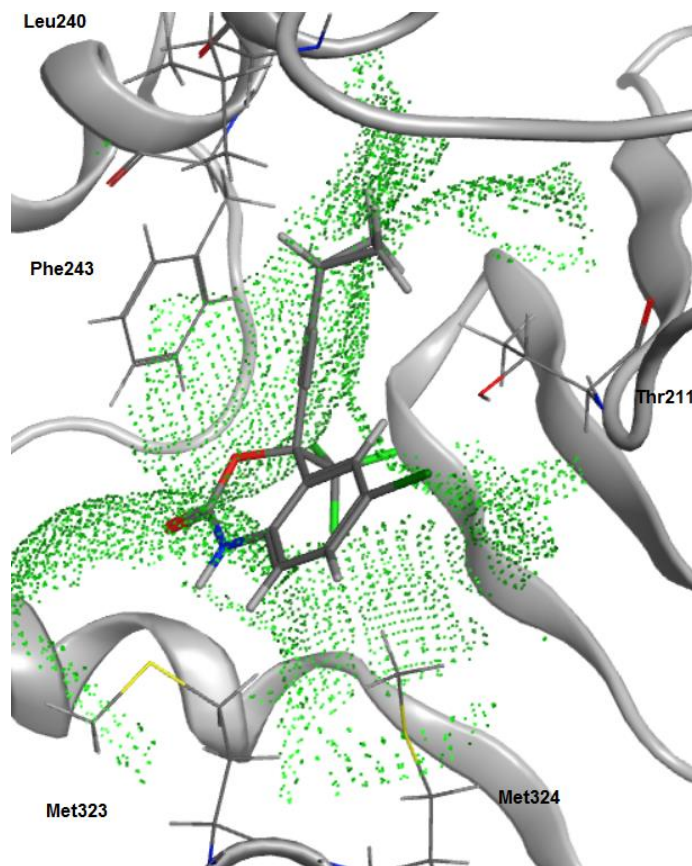


Figure 6.32: Pose of efavirenz over ribbon diagram of zebrafish PXR. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.

### Pioglitazone

Potential pi interactions occur between the thiazolidinedione group in pioglitazone and the conserved residue Trp 299 in zebrafish PXR(Figures 6.33 and 6.34). In human PXR, this distance is longer during molecular dynamics simulations(Figure 6.36). Hence, pioglitazone is predicted to activate zebrafish PXR more than human PXR.



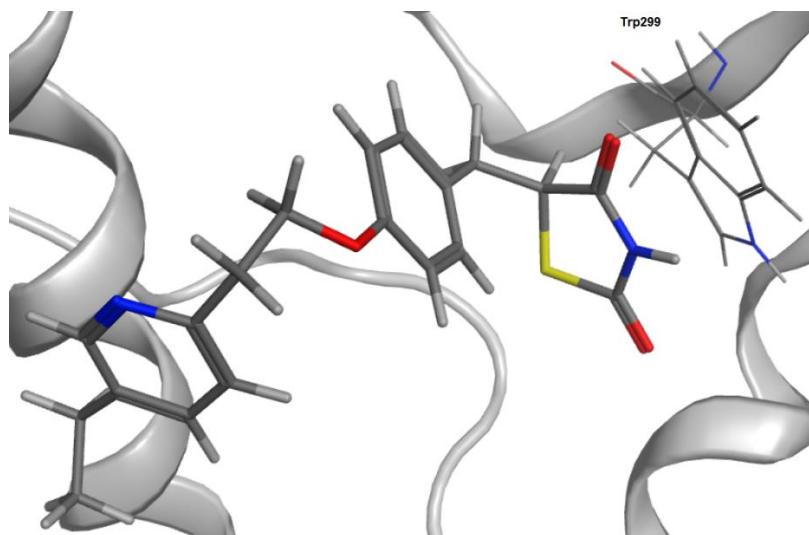


Figure 6.33: Pose of pioglitazone in zebrafish PXR.

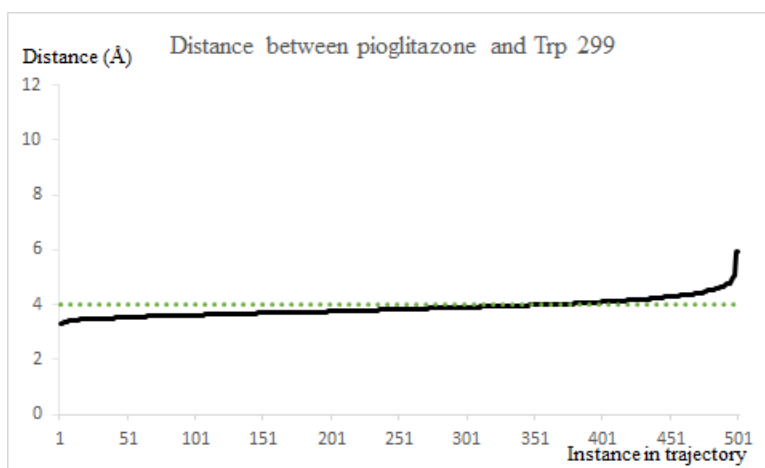


Figure 6.34: Graph of distance between geometric center of aromatic ring in pioglitazone and Trp 299 in zebrafish PXR against instance in trajectory. A 4.0 angstrom distance cutoff is shown as a dotted line.

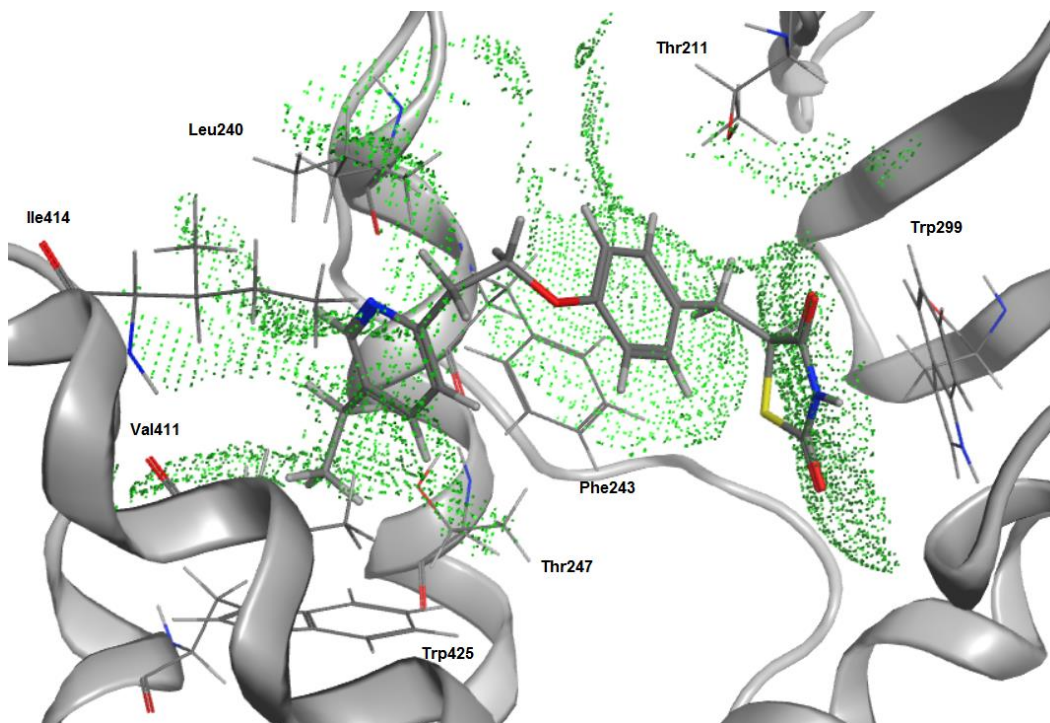


Figure 6.35: Pose of pioglitazone over ribbon diagram of zebrafish PXR. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.

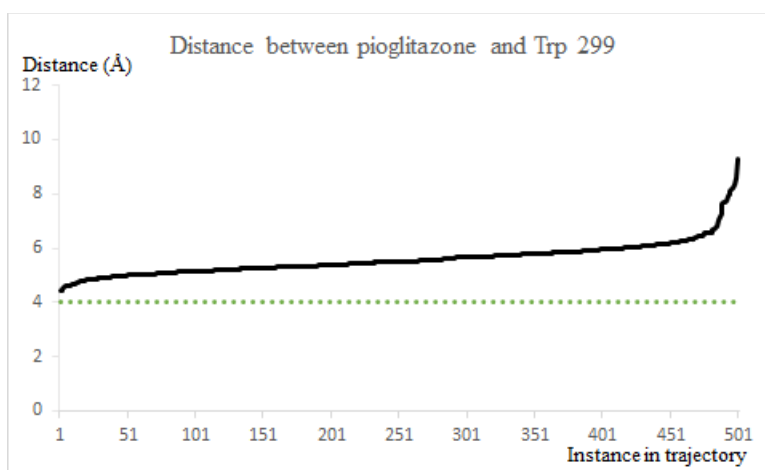


Figure 6.36: Graph of distance between geometric center of aromatic ring in pioglitazone and Trp 299 in human PXR against instance in trajectory. A 4.0 angstrom distance cutoff is shown as a dotted line.

## Prednisone

Potential hydrogen bonds occur between the carbonyl group in prednisone and Gln 285 in human PXR(Figures 6.37 and 6.38). In zebrafish PXR, potential hydrogen bonds occur between the hydroxyl group in prednisone and the conserved residue Met 323(Figures 6.40 and 6.41).

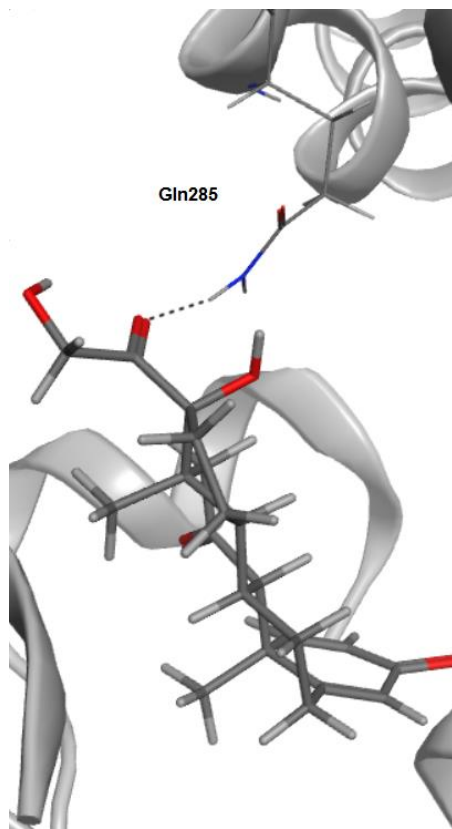


Figure 6.37: Pose of prednisone in human PXR. Potential interactions are marked by dotted lines.

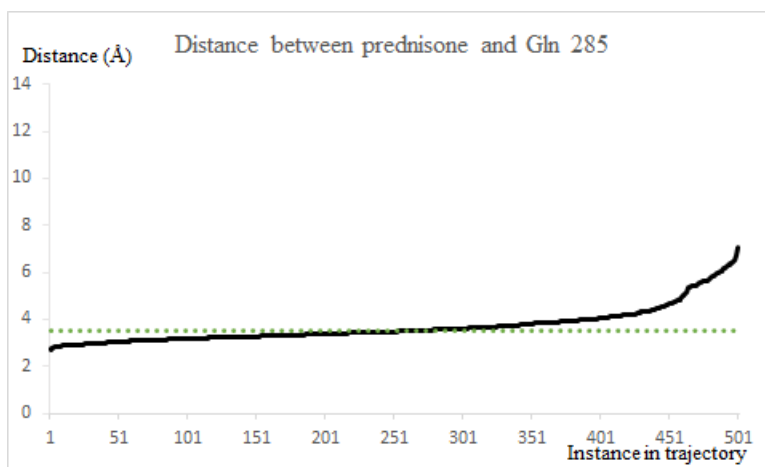


Figure 6.38: Graph of distance between large atom in prednisone and Gln 285 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

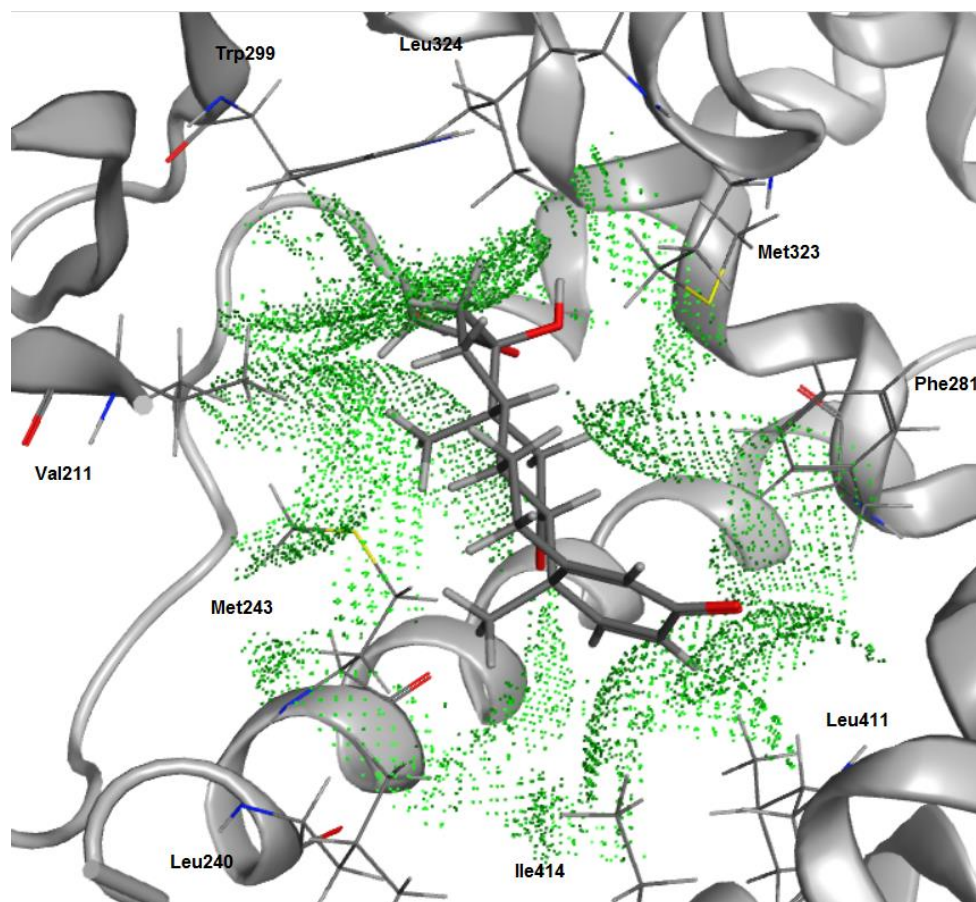


Figure 6.39: Pose of prednisone over ribbon diagram of human PXR. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.

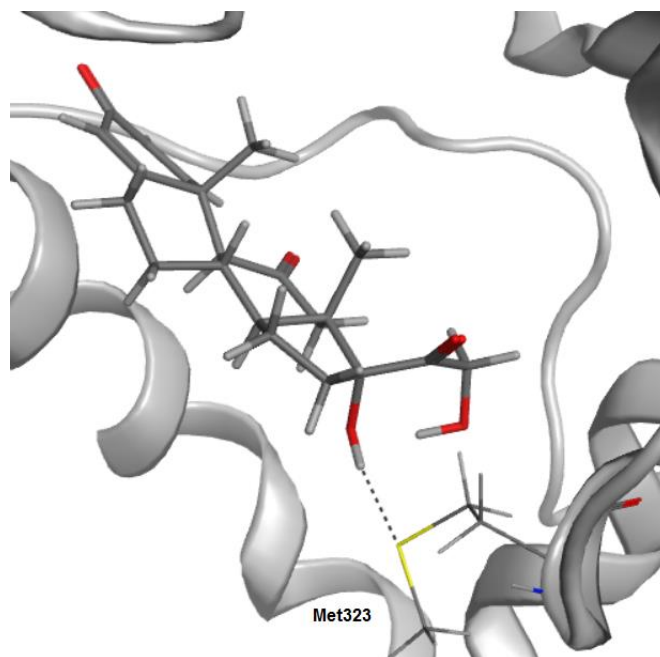


Figure 6.40: Pose of prednisone in zebrafish PXR. Potential interactions are marked by dotted lines.

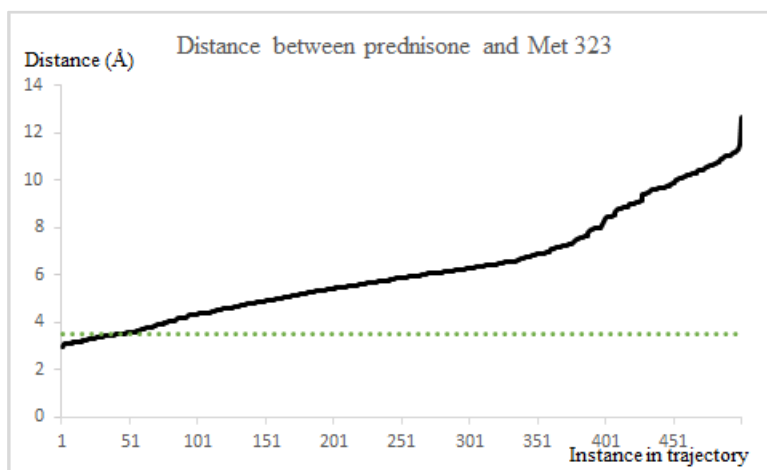


Figure 6.41: Graph of distance between large atom in prednisone and Met 323 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

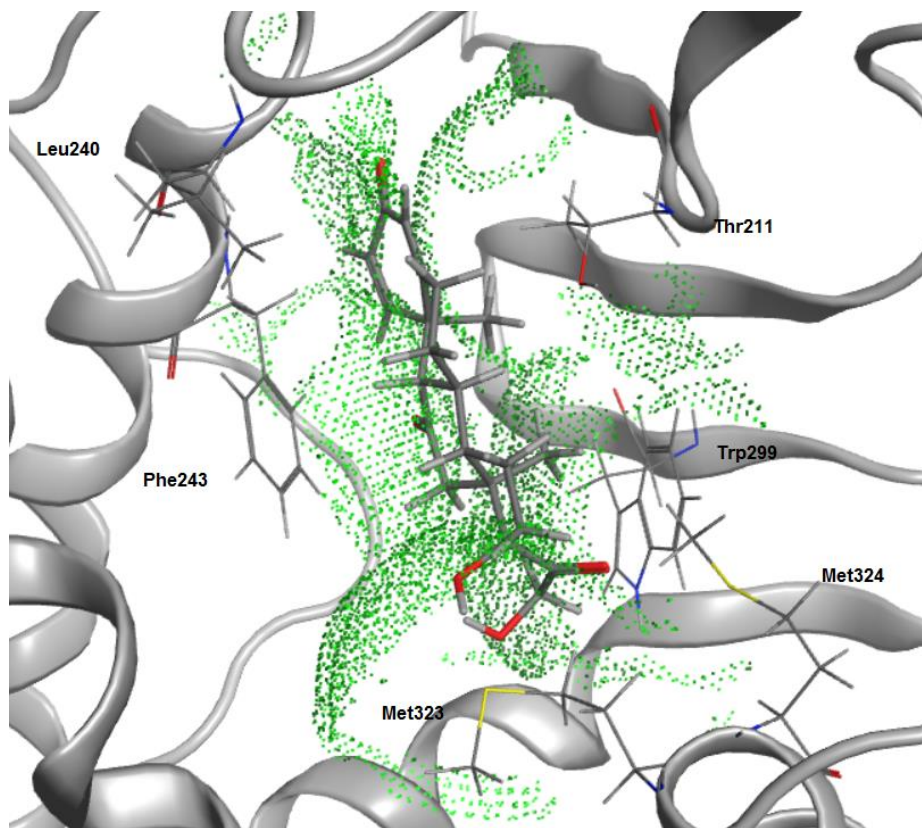


Figure 6.42: Pose of prednisone over ribbon diagram of zebrafish PXR. Hydrophobic residues within 4.0 Å distance are labelled. Hydrophobic contact surfaces within 4.0 Å distance from ligand are shown as dotted green spheres.

### Rufinamide

No consistent set of hydrogen bond or pi interactions are observed in human PXR(Figures 6.43 to 6.46) and in zebrafish PXR(Figures 6.47 to 6.49).

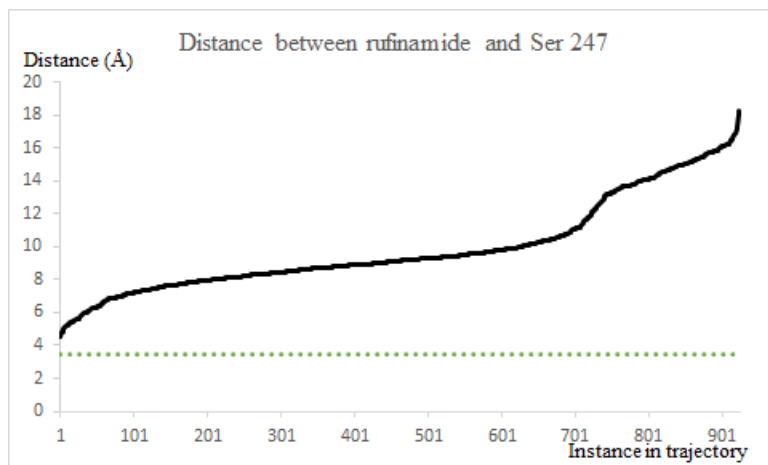


Figure 6.43: Graph of distance between large atom in rufinamide and Ser 247 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

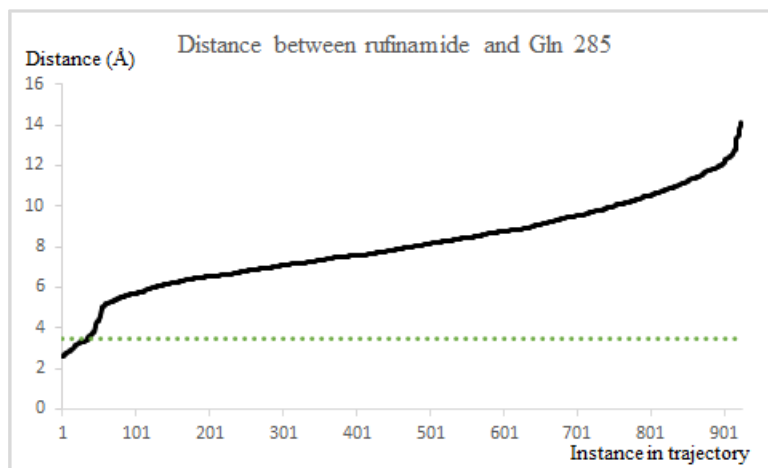


Figure 6.44: Graph of distance between large atom in rufinamide and Gln 285 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

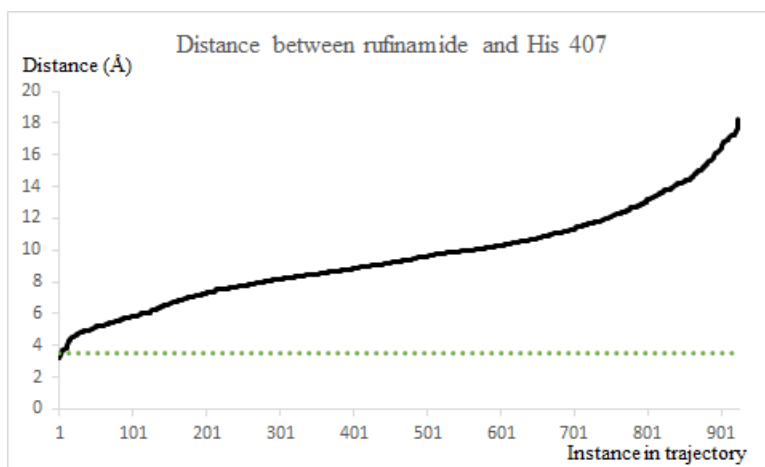


Figure 6.45: Graph of distance between large atom in rufinamide and His 407 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

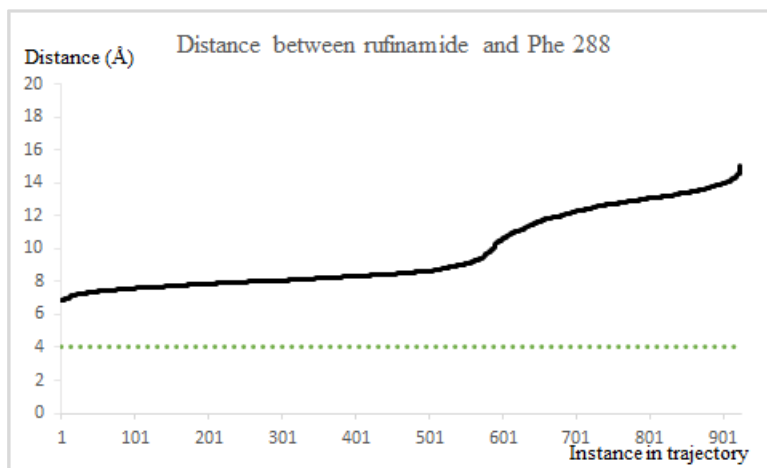


Figure 6.46: Graph of distance between geometric center of aromatic ring in rufinamide and Phe 288 in human PXR against instance in trajectory. A 4.0 angstrom distance cutoff is shown as a dotted line.



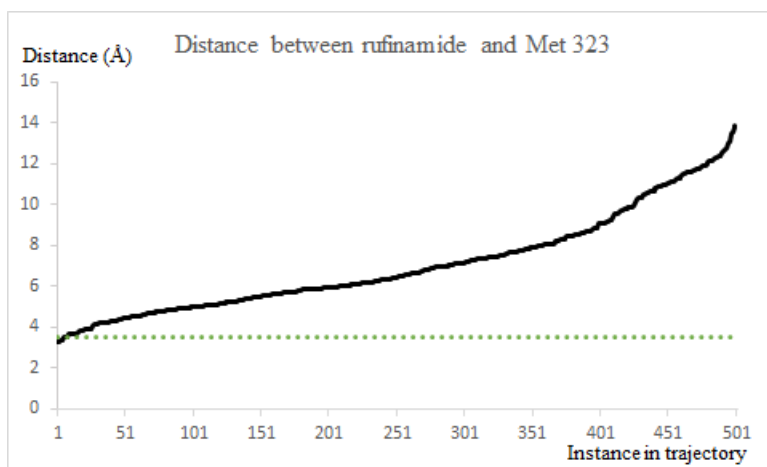


Figure 6.47: Graph of distance between large atom in rufinamide and Met 323 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

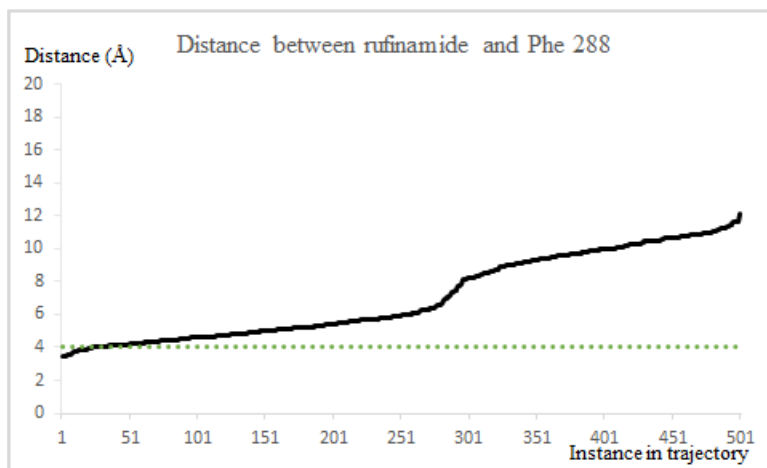


Figure 6.48: Graph of distance between geometric center of aromatic ring in rufinamide and Phe 288 in zebrafish PXR against instance in trajectory. A 4.0 angstrom distance cutoff is shown as a dotted line.

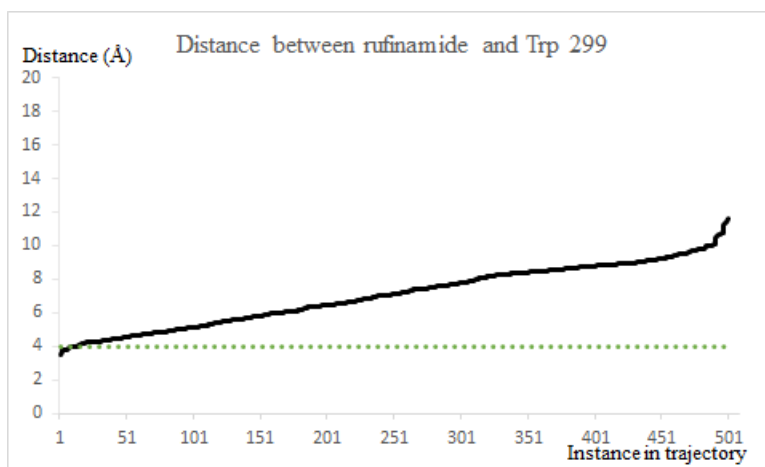


Figure 6.49: Graph of distance between geometric center of aromatic ring in rufinamide and Trp 299 in zebrafish PXR against instance in trajectory. A 4.0 angstrom distance cutoff is shown as a dotted line.

### Acetaminophen

No consistent set of hydrogen bond or pi interactions are observed in human PXR(Figures 6.50 to 6.53) and in zebrafish PXR(Figures 6.54 and 6.55).

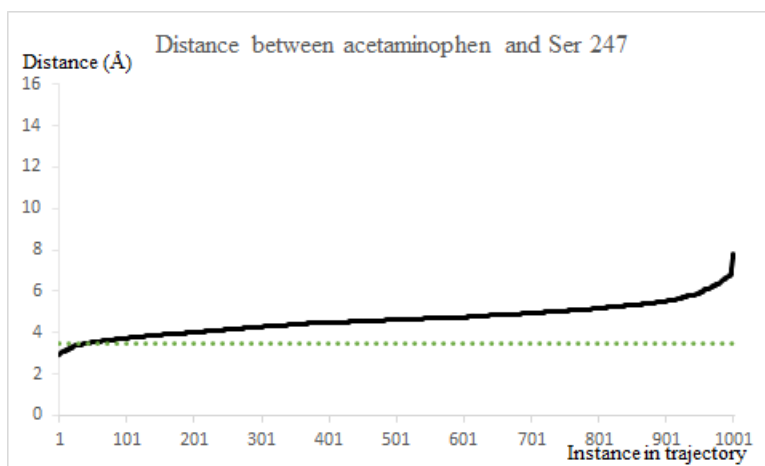


Figure 6.50: Graph of distance between large atom in acetaminophen and Ser 247 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

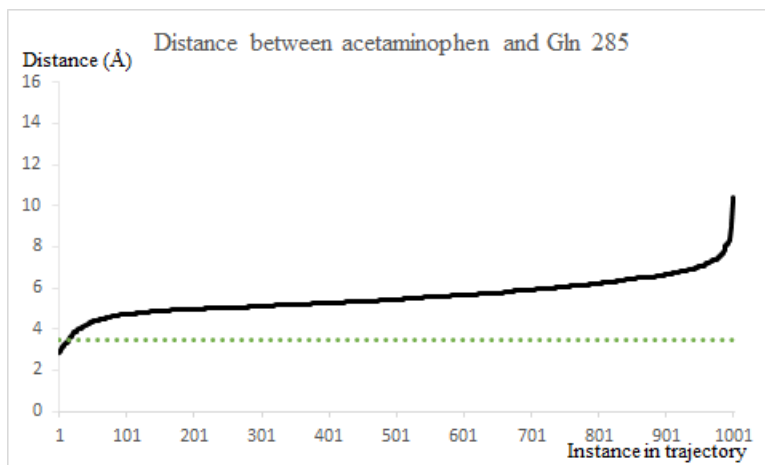


Figure 6.51: Graph of distance between large atom in acetaminophen and Gln 285 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

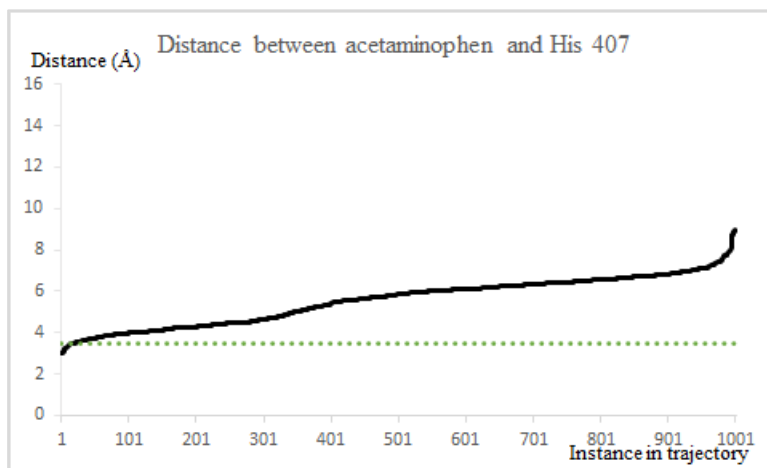


Figure 6.52: Graph of distance between large atom in acetaminophen and His 407 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

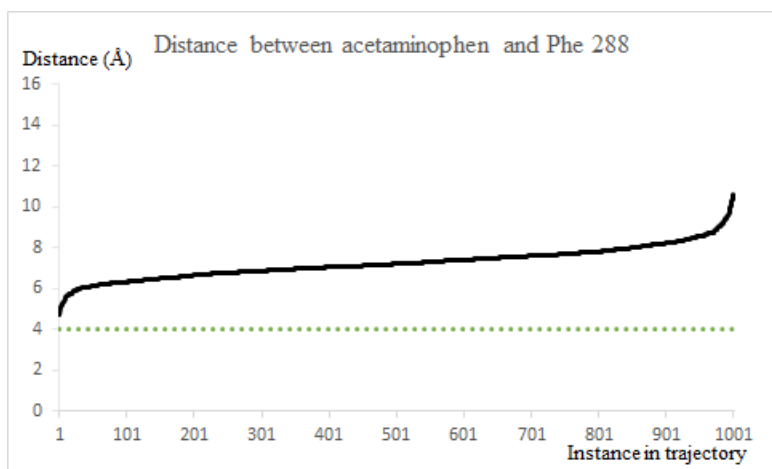


Figure 6.53: Graph of distance between geometric center of aromatic ring in acetaminophen and Phe 288 in human PXR against instance in trajectory. A 4.0 angstrom distance cutoff is shown as a dotted line.

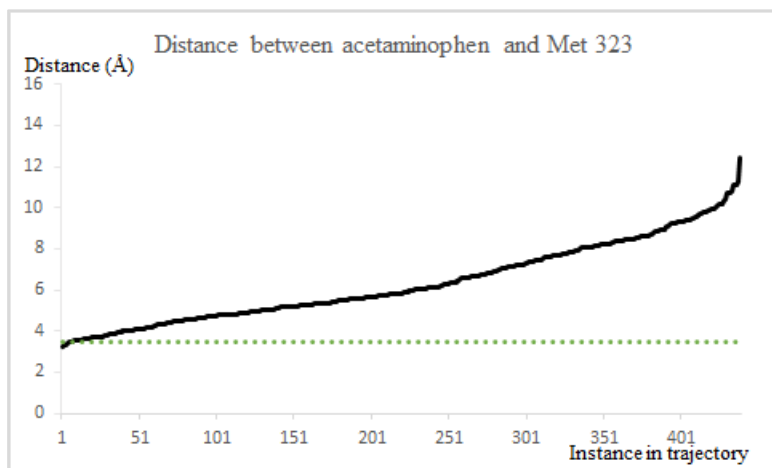


Figure 6.54: Graph of distance between large atom in acetaminophen and Met 323 against instance in trajectory. A 3.5 angstrom distance cutoff is shown as a dotted line.

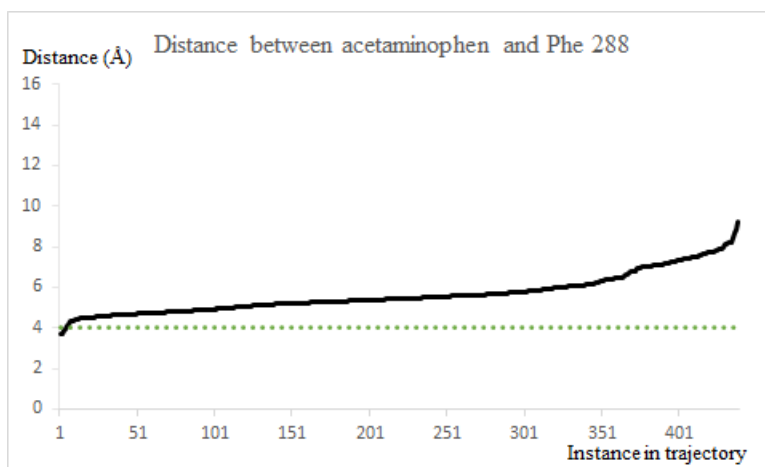


Figure 6.55: Graph of distance between geometric center of aromatic ring in acetaminophen and Phe 288 in zebrafish PXR against instance in trajectory. A 4.0 angstrom distance cutoff is shown as a dotted line.

Table 6.5: Distance between second set of compounds and residue where potential hydrogen bond interactions might be present.

Name of compound	Atom - Residue	Median distance (Å)	Standard deviation	Shortest distance (Å)	Longest distance (Å)	Percentage of trajectory within 3.5 Å
Phenytoin (Human PXR)	O – Gln 285	3.6	2.7	2.8	14	43
Carbamazepine (Human PXR)	N – Ser 247	4.2	1.7	2.9	11	23
Dexamethasone (Human PXR)	O – Gln 285	4.0	0.92	2.7	7.1	30
Dexamethasone (Zebrafish PXR)	O – Met 323	3.7	0.76	2.9	7.0	34
Nafcillin (Human PXR)	O – Gln 285	4.3	1.7	2.8	11	20
Nafcillin (Zebrafish PXR)	N – Met 323	6.3	0.70	3.2	8.4	0.60
Efavirenz (Zebrafish PXR)	N – Met 323	5.6	0.80	3.5	8.3	0.40

Prednisone (Human PXR)	O – Gln 285	3.5	0.77	2.7	7.1	52
Prednisone (Zebrafish PXR)	O – Met 323	5.9	2.2	3.0	13	8.4
Rufinamide (Human PXR)	N – Ser 247	9.2	2.8	4.5	18	0.0
Rufinamide (Human PXR)	O – Gln 285	7.9	2.1	2.6	14	3.7
Rufinamide (Human PXR)	O – His 407	9.3	3.0	3.2	18	0.33
Rufinamide (Zebrafish PXR)	O – Met 323	6.4	2.4	3.3	14	1.2
Acetaminophen (Human PXR)	O – Ser 247	4.6	0.72	2.9	7.8	4.4
Acetaminophen (Human PXR)	O – Gln 285	5.4	0.87	2.8	10	1.5
Acetaminophen (Human PXR)	O – His 407	5.8	1.1	3.0	9.0	1.7
Acetaminophen (Zebrafish PXR)	O – Met 323	5.8	1.9	3.3	12	1.6

Table 6.6: Distance between geometric center of aromatic rings in second set of compounds and residue where potential pi interactions might be present.

Name of compound	Median distance (Å)	Standard deviation	Shortest distance (Å)	Longest distance (Å)	Percentage of trajectory within 4.0 Å
Phenytoin (Zebrafish PXR)	4.3	0.47	3.4	6.3	23

Carbamazepine (Zebrafish PXR)	5.0	1.2	3.4	8.7	17
Pioglitazone (Human PXR)	5.5	0.61	4.4	9.3	0
Pioglitazone (Zebrafish PXR)	3.8	0.33	3.3	5.9	72
Rufinamide (Human PXR)	8.5	2.3	6.8	11	0
Rufinamide (Zebrafish PXR)	5.9	2.5	3.4	12	5.0
Acetaminophen (Human PXR)	7.2	0.76	4.7	11	0
Acetaminophen (Zebrafish PXR)	5.4	0.96	3.7	9.2	0.91

#### 6.4.4 Molecular dynamics as a filter tool

An initial docking with pharmacophore constraints was carried out. This process was able to filter out poses for molecular dynamics simulations, however this method to filter and select poses was less accurate in showing a pose that is consistent with molecular dynamics simulations. The selected poses were also less likely to show a stable ligand conformer. Hence, molecular dynamics was carried out and clustering of each trajectory was carried out to select representative poses for further analysis. Initially, attempts were made to correlate biological activity with molecular dynamics trajectory data by measuring ligand RMSD, all-atoms RMSD, distance between large atoms and distance between geometric center of aromatic rings. Obtaining the distance between large atoms where hydrogen bonds could potentially form and distance between geometric

center of aromatic rings where pi interactions could potentially form was able to obtain a more accurate correlation with the obtained biological activity in human and zebrafish PXR. The selection of poses through clustering of a molecular dynamics trajectory was less accurate in obtaining stable low-energy conformers of the ligand. Attempts were made to redock the ligand on the same binding pocket, select poses that are consistent with the molecular dynamics trajectory data and fulfill a docking score cut-off criteria between reported PXR activators and non-activators and energy minimization of the ligand using a forcefield parameterized for small organic molecules was carried out to select and obtain the best lowest energy conformers for further analysis of electrostatic, polar and hydrophobic surfaces surrounding the ligand. This method of combining molecular dynamics simulations with earlier docking methods that include pharmacophore modelling, as well as docking for pose prediction with ligand energy minimization and analysis of binding site surfaces appear to provide more insight on the interactions between the ligand and binding site, and could potentially be used as a starting point for future work in studying PXR activity using compounds with similar structures to the studied compounds and approved drugs.



## 6.5 Conclusion

From available X-ray crystal structures of human PXR, a homology model of zebrafish PXR was developed using the human PXR crystal structure as a template.

A workflow that is similar to the workflow being carried out in chapter 3 appears to show promising results that could be used as starting points to study further human and zebrafish PXR activity as well as to improve on the prediction of CYP3A4 activity using computational models. The data are consistent with available X-ray structures and site-directed mutagenesis studies on human PXR in the literature where three non-conserved polar residues(Ser 247, Gln 285 and His 407) and three conserved hydrophobic residues(Phe 288, Trp 299 and Met 323) could have an effect on human and zebrafish PXR activation and could be used to screen other compounds to rank and predict PXR activity.

Conformational energies and hydrogen bond geometries were not taken into account in this research project. This could be a potential future area to explore when the need arises.

## **Chapter 7**

### **Conclusion**

#### **7.1 Major findings**

##### **7.1.1 Ligand-based study**

The QSAR model being developed had a consensus model of 2 Naïve Bayes and 3 support vector machine (SVM) modelling methods as these modelling methods are less sensitive and hence have a lower capacity for the model to be biased due to differences in distance between each object or datapoint.

The model being developed was also able to determine that 1D and 2D descriptors were sufficient to build a model that was able to predict the compounds in the validation sets well without including additional 3D descriptors.

With the additional criteria of including multiple thresholds to define the applicability domain of the model, and to perform a more rigorous 5-fold external cross-validation of the model by splitting the dataset into 5 equal parts and using a different dataset each time to train the model and test on the validation set, the

additional methods were able to predict each compound in the validation set, testing set and ZINC chemical library with a certain degree of confidence, and the additional results of using a different consensus model to screen through 5 different validation sets enable a more representative view and analysis of the results of the ligand-based study.

Among the purchasable ZINC database compounds that were screened in the ligand-based study, 45673 compounds were identified for further analysis in structure-based studies.

### **7.1.2 Structure-based study of ZINC database compounds on JIP binding site**

The X-ray crystal structure of JNK1 with the truncated form of JIP1 was used to dock the 45673 compounds from the earlier ligand-based study on the JIP1 binding pocket of JNK1. A pharmacophore model was developed where the hydrogen bond interactions between Thr 159 in JIP1 with the residue Arg 127 in JNK1 was used to rapidly dock 45673 compounds and identify suitable poses as a starting point for molecular dynamics simulations. The ligand root-mean square deviation is used to measure the stability of the molecular dynamics trajectory and clustering of the trajectory was carried out to select a representative pose for analysis of binding interactions. A refined list of 1 compound was prioritized for

*in vitro* testing after showing a stable trajectory with potential hydrogen bonds that were formed with Arg 127.

### **7.1.3 Structure-based study of ZINC database compounds on DFG binding site**

In the docking of DFG site compounds, as there were an insufficient number of compounds to perform a ligand-based study, a subset of purchasable compounds were chosen based on a diversity cutoff of 0.7 based on ChemAxon fingerprints. This became the starting number of compounds to use for the structure-based study. The compound list was refined and promising compounds were purchased and tested for *in vitro* activity.

The selected list of compounds that were tested for inhibitory activity on JNK2 did not show any significant binding at concentrations of up to 40 $\mu$ M. 1 compound showed 19 % inhibition at 40  $\mu$ M on JNK2, and 1 compound showed 20 % inhibition at 40  $\mu$ M on JNK1.

### **7.1.4 *in silico* model of human and zebrafish PXR activators**

Using the X-ray crystal structure of human PXR and a homology model of zebrafish PXR, an *in silico* model was developed to correlate the amount of

CYP3A4 gene expression with binding orientation of various approved drugs to human and zebrafish PXR.

Through a workflow process of combining docking with molecular dynamics, the binding orientation for 15 compounds were identified where structural interactions in each compound and 3 conserved residues in human and zebrafish PXR were identified to be important for PXR activation. These results could potentially be used to model other drug or chemical compounds in human and other species' PXR to predict the extent of metabolism of the compound and CYP3A4 expression and activity and reduce the costs in carrying out *in vitro* and *in vivo* assays.

## **7.2 Limitations**

Among the limitations of the ligand-based study, the relative lack of positive compounds in the training set and validation set and the unknown inhibitory activity profile in the putative negatives dataset affect the accuracy of the model and hence, the actual number of true positives and false positives identified from screening the ZINC chemical library are not known.

Structure-based studies were carried out to identify JNK inhibitors and PXR activators. Although molecular dynamics are able to form more accurate prediction models than docking with and without pharmacophore constraints, it is

highly dependent on available 3-dimensional structural information and site-directed mutagenesis studies that are available in the literature.

### **7.3 Future work**

The next phase of this research would be to carry out *in vitro* tests on prioritized compounds, and additional molecular dynamics studies on compounds which met the docking with pharmacophore constraints filter, but were not selected for molecular dynamics studies and *in vitro* testing due to time and budget constraints. As only 1 compound out of the 17833934 compounds in the ZINC database were chosen for *in vitro* tests, with an ever increasing number of diverse chemicals being added to the database every month, there remains a very large chemical space to sample and a very large knowledge gap with regard to identifying potential JNK1 and JNK2 inhibitors, targeting two different active sites.

One possible approach to consider in future work would be to use ligand-based 3D pharmacophore models as an additional prefiltering tool to refine the model and to sample the large and diverse chemical space more accurately as this method has been reported in the literature to be useful as a complementary method to HTS and existing 2D ligand-based and 3D structure-based computational methods[91,101,189,190].

In addition, as the initial goals of the *in vitro* tests were to carry out tests for biological activity, there remains this knowledge gap where the identified positive compounds are not known which active site it binds to and in which conformation. Hence, further work might possibly include finding additional collaborators to carry out X-ray crystallography studies or nuclear magnetic resonance imaging studies to determine the location of the active site and the pose that the compound adopts to bind to the protein and induce the biological activity for identified JNK inhibitors or PXR activators.

Another possible further work will also include carrying out similar virtual screening studies on different JNK isoforms JNK2 and JNK3 to further screen and identify novel selective inhibitors that target only one JNK isoform. This is because JNK1, JNK2 and JNK3 are located in varying amounts in different parts of the human body as revealed by Northern blot analysis on different mammalian cell lines, and current research show that inhibition or gene-knockout of only one JNK isoform has a therapeutic effect in reducing the severity of a disease in *in vitro* and *in vivo* experiments, but gene-knockout of two or more could be lethal instead[36,191,192]. Hence, this aspect of research will address the safety aspect of developing a selective inhibitor of only one or two JNK isoforms, and hence also allow the possibility of having multiple patents being developed at the end of the process.

Another additional further work would include working further with collaborators to obtain biological activity on PXR to refine and improve on the model to be able to predict additional FDA approved drugs or xenobiotics more accurately by attempting to determine the amount of agonist or antagonist activity, together with the degree of binding and possible binding poses through additional docking and molecular dynamics studies.

#### **7.4 Dual-targeting approach and polypharmacology**

The dual-targeting approach based on the concept of polypharmacology is gaining more interest in recent years. There are two possible interpretations to this approach. The first interpretation refers to the design of compounds that target two or more proteins involved in a disease process simultaneously. An example would be to design inhibitors that could inhibit both the JNK and p38 enzymes for possible treatment of cancer where it involves multiple different enzymes and pathways. To consider such an approach for a computational study, a disease must be chosen first, and then the docking of compounds must be carried out to several proteins involved in the disease. However, docking to several proteins is not a trivial task as each protein is unique and requires separate refinement to the docking procedure. Thus, based on the time available, the design of inhibitors for JNK will be carried out first.



The second interpretation refers to the design of compounds that can bind to two binding sites in the same protein simultaneously. Among the 191 compounds that target the JIP binding site of JNK, the set of compounds published by Chen et al[37] were discovered to bind to both the ATP binding site, as well as the JIP binding site. Technically, it is possible to design such dual targeting drugs in this study. However, unlike the current design approach being used in this study, which involves screening of commercially available compounds, which could be easily bought and tested using kinase assay kits, testing such dual targeting drugs will not be as straightforward due to the need to find additional collaborators that are able to synthesize and test the compounds. For this reason, this approach is not given priority at this stage in the research project. However, this approach might be included in further studies when the opportunity arises.

## Bibliography

1. G. Manning, D. B. Whyte, R. Martinez, T. Hunter and S. Sudarsanam, "The protein kinase complement of the human genome" *Science*, vol. 298, no. 5600, pp. 1912-1934, Dec 6. 2002.
2. P. Cohen, "Protein kinases--the major drug targets of the twenty-first century?" *Nature reviews Drug discovery*, vol. 1, no. 4, pp. 309-315, Apr. 2002.
3. L. K. Gavrin and E. Saiah, "Approaches to discover non-ATP site kinase inhibitors" *MedChemComm*, vol. 4, no. 1, pp. 41-51, 2013.
4. X. Chen, Z. L. Ji and Y. Z. Chen, "TTD: Therapeutic Target Database" *Nucleic acids research*, vol. 30, no. 1, pp. 412-415, Jan 1. 2002.
5. F. Zhu, B. Han, P. Kumar, X. Liu, X. Ma, X. Wei, L. Huang, Y. Guo, L. Han, C. Zheng and Y. Chen, "Update of TTD: Therapeutic Target Database" *Nucleic acids research*, vol. 38, no. suppl 1(Database issue), pp. D787-791, Jan. 2010.
6. F. Zhu, Z. Shi, C. Qin, L. Tao, X. Liu, F. Xu, L. Zhang, Y. Song, J. Zhang, B. Han, P. Zhang and Y. Chen, "Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery" *Nucleic acids research*, vol. 40, no. D1(Database issue), pp. D1128-1136, Jan. 2012.
7. M. Kanehisa and S. Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes" *Nucleic acids research*, vol. 28, no. 1, pp. 27-30, Jan 1. 2000.
8. M. Kanehisa, S. Goto, Y. Sato, M. Furumichi and M. Tanabe, "KEGG for integration and interpretation of large-scale molecular data sets" *Nucleic acids research*, vol. 40, no. D1(Database issue), pp. D109-114, Jan. 2012.
9. R. Caspi, T. Altman, K. Dreher, C. A. Fulcher, P. Subhraveti, I. M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, A. Pujar, A. G. Shearer, M. Travers, D. Weerasinghe, P. Zhang and P. D. Karp, "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases" *Nucleic acids research*, vol. 40, no. D1(Database issue), pp. D742-753, Jan. 2012.
10. M. A. Bogoyevitch and P. G. Arthur, "Inhibitors of c-Jun N-terminal kinases: JuNK no more?" *Biochimica et biophysica acta*, vol. 1784, no. 1, pp. 76-93, Jan. 2008.
11. M. A. Bogoyevitch, K. R. Ngoei, T. T. Zhao, Y. Y. Yeap and D. C. Ng, "c-Jun N-terminal kinase (JNK) signaling: recent advances and challenges" *Biochimica et biophysica acta*, vol. 1804, no. 3, pp. 463-475, Mar. 2010.
12. A. M. Manning and R. J. Davis, "Targeting JNK for therapeutic benefit: from junk to gold?" *Nature reviews Drug discovery*, vol. 2, no. 7, pp. 554-565, Jul. 2003.
13. C. D. Klaassen, J. B. Watkins and L. J. Casarett (2010) Casarett & Doull's essentials of toxicology. 2nd ed. McGraw-Hill Medical, New York

14. A. P. Li, "Biomarkers and human hepatocytes" *Biomarkers in medicine*, vol. 8, no. 2, pp. 173-183, Feb. 2014.
15. J. L. Barwick, L. C. Quattrochi, A. S. Mills, C. Potenza, R. H. Tukey and P. S. Guzelian, "Trans-species gene transfer for analysis of glucocorticoid-inducible transcriptional activation of transiently expressed human CYP3A4 and rabbit CYP3A6 in primary cultures of adult rat and rabbit hepatocytes" *Molecular pharmacology*, vol. 50, no. 1, pp. 10-16, Jul. 1996.
16. D. R. Nelson, L. Koymans, T. Kamataki, J. J. Stegeman, R. Feyereisen, D. J. Waxman, M. R. Waterman, O. Gotoh, M. J. Coon, R. W. Estabrook, I. C. Gunsalus and D. W. Nebert, "P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature" *Pharmacogenetics*, vol. 6, no. 1, pp. 1-42, Feb. 1996.
17. J. Chen and K. Raymond, "Roles of rifampicin in drug-drug interactions: underlying molecular mechanisms involving the nuclear pregnane X receptor" *Annals of clinical microbiology and antimicrobials*, vol. 5, no. pp. 3, 2006.
18. J. Chen, K. N. Zhao and C. Chen, "The role of CYP3A4 in the biotransformation of bile acids and therapeutic implication for cholestasis" *Annals of translational medicine*, vol. 2, no. 1, pp. 7, Jan. 2014.
19. B. Goodwin, E. Hodgson and C. Liddle, "The orphan human pregnane X receptor mediates the transcriptional activation of CYP3A4 by rifampicin through a distal enhancer module" *Molecular pharmacology*, vol. 56, no. 6, pp. 1329-1339, Dec. 1999.
20. S. A. Kliewer, J. T. Moore, L. Wade, J. L. Staudinger, M. A. Watson, S. A. Jones, D. D. McKee, B. B. Oliver, T. M. Willson, R. H. Zetterstrom, T. Perlmann and J. M. Lehmann, "An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway" *Cell*, vol. 92, no. 1, pp. 73-82, Jan 9. 1998.
21. W. Xie, J. L. Barwick, M. Downes, B. Blumberg, C. M. Simon, M. C. Nelson, B. A. Neuschwander-Tetri, E. M. Brunt, P. S. Guzelian and R. M. Evans, "Humanized xenobiotic response in mice expressing nuclear receptor SXR" *Nature*, vol. 406, no. 6794, pp. 435-439, Jul 27. 2000.
22. Z. Zhang, P. E. Burch, A. J. Cooney, R. B. Lanz, F. A. Pereira, J. Wu, R. A. Gibbs, G. Weinstock and D. A. Wheeler, "Genomic analysis of the nuclear receptor family: new insights into structure, regulation, and evolution from the rat genome" *Genome research*, vol. 14, no. 4, pp. 580-590, Apr. 2004.
23. P. Cozzini and G. E. Kellogg (2012) *Computational Approaches to Nuclear Receptors*. Volume 30. Royal Society of Chemistry, Cambridge, UK
24. S. Gupta, T. Barrett, A. J. Whitmarsh, J. Cavanagh, H. K. Sluss, B. Derijard and R. J. Davis, "Selective interaction of JNK protein kinase isoforms with transcription factors" *The EMBO journal*, vol. 15, no. 11, pp. 2760-2770, Jun 3. 1996.

25. B. Derijard, M. Hibi, I. H. Wu, T. Barrett, B. Su, T. Deng, M. Karin and R. J. Davis, "JNK1: a protein kinase stimulated by UV light and Ha-Ras that binds and phosphorylates the c-Jun activation domain" *Cell*, vol. 76, no. 6, pp. 1025-1037, Mar 25. 1994.
26. Z. Wu, J. Wu, E. Jacinto and M. Karin, "Molecular cloning and characterization of human JNKK2, a novel Jun NH2-terminal kinase-specific kinase" *Molecular and cellular biology*, vol. 17, no. 12, pp. 7407-7416, Dec. 1997.
27. M. Ito, K. Yoshioka, M. Akechi, S. Yamashita, N. Takamatsu, K. Sugiyama, M. Hibi, Y. Nakabeppu, T. Shiba and K. I. Yamamoto, "JSAP1, a novel jun N-terminal protein kinase (JNK)-binding protein that functions as a Scaffold factor in the JNK signaling pathway" *Molecular and cellular biology*, vol. 19, no. 11, pp. 7539-7548, Nov. 1999.
28. N. Kelkar, S. Gupta, M. Dickens and R. J. Davis, "Interaction of a mitogen-activated protein kinase signaling module with the neuronal protein JIP3" *Molecular and cellular biology*, vol. 20, no. 3, pp. 1030-1043, Feb. 2000.
29. R. J. Davis, "Signal transduction by the JNK group of MAP kinases" *Cell*, vol. 103, no. 2, pp. 239-252, Oct 13. 2000.
30. J. Hirosumi, G. Tuncman, L. Chang, C. Z. Gorgun, K. T. Uysal, K. Maeda, M. Karin and G. S. Hotamisligil, "A central role for JNK in obesity and insulin resistance" *Nature*, vol. 420, no. 6913, pp. 333-336, Nov 21. 2002.
31. Z. Han, D. L. Boyle, L. Chang, B. Bennett, M. Karin, L. Yang, A. M. Manning and G. S. Firestein, "c-Jun N-terminal kinase is required for metalloproteinase expression and joint destruction in inflammatory arthritis" *The Journal of clinical investigation*, vol. 108, no. 1, pp. 73-81, Jul. 2001.
32. X. G. Xia, T. Harding, M. Weller, A. Bieneman, J. B. Uney and J. B. Schulz, "Gene transfer of the JNK interacting protein-1 protects dopaminergic neurons in the MPTP model of Parkinson's disease" *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 18, pp. 10433-10438, Aug 28. 2001.
33. O. Potapova, M. Gorospe, F. Bost, N. M. Dean, W. A. Gaarde, D. Mercola and N. J. Holbrook, "c-Jun N-terminal kinase is essential for growth of human T98G glioblastoma cells" *The Journal of biological chemistry*, vol. 275, no. 32, pp. 24767-24775, Aug 11. 2000.
34. O. Potapova, A. Haghghi, F. Bost, C. Liu, M. J. Birrer, R. Gjerset and D. Mercola, "The Jun kinase/stress-activated protein kinase pathway functions to regulate DNA repair and inhibition of the pathway sensitizes tumor cells to cisplatin" *The Journal of biological chemistry*, vol. 272, no. 22, pp. 14041-14044, May 30. 1997.
35. T. Borsello, P. G. Clarke, L. Hirt, A. Vercelli, M. Repici, D. F. Schorderet, J. Bogousslavsky and C. Bonny, "A peptide inhibitor of c-Jun N-terminal kinase protects against excitotoxicity and cerebral ischemia" *Nature medicine*, vol. 9, no. 9, pp. 1180-1186, Sep. 2003.

36. V. Waetzig and T. Herdegen, "Context-specific inhibition of JNKs: overcoming the dilemma of protection and damage" *Trends in pharmacological sciences*, vol. 26, no. 9, pp. 455-461, Sep. 2005.
37. T. Chen, N. Kablaoui, J. Little, S. Timofeevski, W. R. Tschantz, P. Chen, J. Feng, M. Charlton, R. Stanton and P. Bauer, "Identification of small-molecule inhibitors of the JIP-JNK interaction" *The Biochemical journal*, vol. 420, no. 2, pp. 283-294, Jun 1. 2009.
38. B. L. Bennett, D. T. Sasaki, B. W. Murray, E. C. O'Leary, S. T. Sakata, W. Xu, J. C. Leisten, A. Motiwala, S. Pierce, Y. Satoh, S. S. Bhagwat, A. M. Manning and D. W. Anderson, "SP600125, an anthrapyrazolone inhibitor of Jun N-terminal kinase" *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 24, pp. 13681-13686, Nov 20. 2001.
39. M. Alam, R. E. Beevers, T. Ceska, R. J. Davenport, K. M. Dickson, M. Fortunato, L. Gowers, A. F. Haughan, L. A. James, M. W. Jones, N. Kinsella, C. Lowe, J. W. Meissner, A. L. Nicolas, B. G. Perry, D. J. Phillips, W. R. Pitt, A. Platt, A. J. Ratcliffe, A. Sharpe and L. J. Tait, "Synthesis and SAR of aminopyrimidines as novel c-Jun N-terminal kinase (JNK) inhibitors" *Bioorganic & medicinal chemistry letters*, vol. 17, no. 12, pp. 3463-3467, Jun 15. 2007.
40. Y. Asano, S. Kitamura, T. Ohra, K. Aso, H. Igata, T. Tamura, T. Kawamoto, T. Tanaka, S. Sogabe, S. Matsumoto, M. Yamaguchi, H. Kimura and F. Itoh, "Discovery, synthesis and biological evaluation of isoquinolones as novel and highly selective JNK inhibitors (1)" *Bioorganic & medicinal chemistry*, vol. 16, no. 8, pp. 4715-4732, Apr 15. 2008.
41. Y. Asano, S. Kitamura, T. Ohra, F. Itoh, M. Kajino, T. Tamura, M. Kaneko, S. Ikeda, H. Igata, T. Kawamoto, S. Sogabe, S. Matsumoto, T. Tanaka, M. Yamaguchi, H. Kimura and S. Fukumoto, "Discovery, synthesis and biological evaluation of isoquinolones as novel and highly selective JNK inhibitors (2)" *Bioorganic & medicinal chemistry*, vol. 16, no. 8, pp. 4699-4714, Apr 15. 2008.
42. S. D. Chamberlain, A. M. Redman, J. W. Wilson, F. Deanda, J. B. Shotwell, R. Gerding, H. Lei, B. Yang, K. L. Stevens, A. M. Hassell, L. M. Shewchuk, M. A. Leesnitzer, J. L. Smith, P. Sabbatini, C. Atkins, A. Groy, J. L. Rowand, R. Kumar, R. A. Mook, Jr., G. Moorthy and S. Patnaik, "Optimization of 4,6-bis-anilino-1H-pyrrolo[2,3-d]pyrimidine IGF-1R tyrosine kinase inhibitors towards JNK selectivity" *Bioorganic & medicinal chemistry letters*, vol. 19, no. 2, pp. 360-364, Jan 15. 2009.
43. P. Gaillard, I. Jeanclaude-Etter, V. Ardissonne, S. Arkinstall, Y. Cambet, M. Camps, C. Chabert, D. Church, R. Cirillo, D. Gretener, S. Halazy, A. Nichols, C. Szyndralewicz, P. A. Vitte and J. P. Gotteland, "Design and synthesis of the first generation of novel potent, selective, and in vivo active (benzothiazol-2-yl)acetonitrile inhibitors of the c-Jun N-terminal kinase" *Journal of medicinal chemistry*, vol. 48, no. 14, pp. 4596-4607, Jul 14. 2005.

44. P. S. Humphries, J. A. Lafontaine, C. S. Agree, D. Alexander, P. Chen, Q. Q. Do, L. Y. Li, E. A. Lunney, R. J. Rajapakse, K. Siegel, S. L. Timofeevski, T. Wang and D. M. Wilhite, "Synthesis and SAR of 4-substituted-2-aminopyrimidines as novel c-Jun N-terminal kinase (JNK) inhibitors" *Bioorganic & medicinal chemistry letters*, vol. 19, no. 8, pp. 2099-2102, Apr 15. 2009.
45. R. Jiang, D. Duckett, W. Chen, J. Habel, Y. Y. Ling, P. LoGrasso and T. M. Kamenecka, "3,5-Disubstituted quinolines as novel c-Jun N-terminal kinase inhibitors" *Bioorganic & medicinal chemistry letters*, vol. 17, no. 22, pp. 6378-6382, Nov 15. 2007.
46. T. Kamenecka, J. Habel, D. Duckett, W. Chen, Y. Y. Ling, B. Frackowiak, R. Jiang, Y. Shin, X. Song and P. LoGrasso, "Structure-activity relationships and X-ray structures describing the selectivity of aminopyrazole inhibitors for c-Jun N-terminal kinase 3 (JNK3) over p38" *The Journal of biological chemistry*, vol. 284, no. 19, pp. 12853-12861, May 8. 2009.
47. M. Liu, S. Wang, J. E. Clampit, R. J. Gum, D. L. Haasch, C. M. Rondinone, J. M. Trevillyan, C. Abad-Zapatero, E. H. Fry, H. L. Sham and G. Liu, "Discovery of a new class of 4-anilinopyrimidines as potent c-Jun N-terminal kinase inhibitors: Synthesis and SAR studies" *Bioorganic & medicinal chemistry letters*, vol. 17, no. 3, pp. 668-672, Feb 1. 2007.
48. M. Liu, Z. Xin, J. E. Clampit, S. Wang, R. J. Gum, D. L. Haasch, J. M. Trevillyan, C. Abad-Zapatero, E. H. Fry, H. L. Sham and G. Liu, "Synthesis and SAR of 1,9-dihydro-9-hydroxypyrazolo[3,4-b]quinolin-4-ones as novel, selective c-Jun N-terminal kinase inhibitors" *Bioorganic & medicinal chemistry letters*, vol. 16, no. 10, pp. 2590-2594, May 15. 2006.
49. P. LoGrasso and T. Kamenecka, "Inhibitors of c-jun-N-terminal kinase (JNK)" *Mini reviews in medicinal chemistry*, vol. 8, no. 8, pp. 755-766, Jul. 2008.
50. T. Ruckle, M. Biamonte, T. Grippi-Vallotton, S. Arkininstall, Y. Cambet, M. Camps, C. Chabert, D. J. Church, S. Halazy, X. Jiang, I. Martinou, A. Nichols, W. Sauer and J. P. Gotteland, "Design, synthesis, and biological activity of novel, potent, and selective (benzoylaminomethyl)thiophene sulfonamide inhibitors of c-Jun-N-terminal kinase" *Journal of medicinal chemistry*, vol. 47, no. 27, pp. 6921-6934, Dec 30. 2004.
51. Y. Shin, W. Chen, J. Habel, D. Duckett, Y. Y. Ling, M. Koenig, Y. He, T. Vojkovsky, P. LoGrasso and T. M. Kamenecka, "Synthesis and SAR of piperazine amides as novel c-jun N-terminal kinase (JNK) inhibitors" *Bioorganic & medicinal chemistry letters*, vol. 19, no. 12, pp. 3344-3347, Jun 15. 2009.
52. M. J. Stocks, S. Barber, R. Ford, F. Leroux, S. St-Gallay, S. Teague and Y. Xue, "Structure-driven HtL: design and synthesis of novel aminoindazole inhibitors of c-Jun N-terminal kinase activity" *Bioorganic & medicinal chemistry letters*, vol. 15, no. 14, pp. 3459-3462, Jul 15. 2005.

53. B. G. Szczepankiewicz, C. Kosogof, L. T. Nelson, G. Liu, B. Liu, H. Zhao, M. D. Serby, Z. Xin, M. Liu, R. J. Gum, D. L. Haasch, S. Wang, J. E. Clampit, E. F. Johnson, T. H. Lubben, M. A. Stashko, E. T. Olejniczak, C. Sun, S. A. Dorwin, K. Haskins, C. Abad-Zapatero, E. H. Fry, C. W. Hutchins, H. L. Sham, C. M. Rondinone and J. M. Trevillyan, "Aminopyridine-based c-Jun N-terminal kinase inhibitors with cellular activity and minimal cross-kinase activity" *Journal of medicinal chemistry*, vol. 49, no. 12, pp. 3563-3580, Jun 15. 2006.
54. H. Zhao, M. D. Serby, Z. Xin, B. G. Szczepankiewicz, M. Liu, C. Kosogof, B. Liu, L. T. Nelson, E. F. Johnson, S. Wang, T. Pederson, R. J. Gum, J. E. Clampit, D. L. Haasch, C. Abad-Zapatero, E. H. Fry, C. Rondinone, J. M. Trevillyan, H. L. Sham and G. Liu, "Discovery of potent, highly selective, and orally bioavailable pyridine carboxamide c-Jun NH<sub>2</sub>-terminal kinase inhibitors" *Journal of medicinal chemistry*, vol. 49, no. 15, pp. 4455-4458, Jul 27. 2006.
55. Discovery Studio Modeling Environment. 4.0. San Diego, CA, USA: Accelrys Software Inc.; 2013
56. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, "The Protein Data Bank" *Nucleic acids research*, vol. 28, no. 1, pp. 235-242, Jan 1. 2000.
57. Y. S. Heo, S. K. Kim, C. I. Seo, Y. K. Kim, B. J. Sung, H. S. Lee, J. I. Lee, S. Y. Park, J. H. Kim, K. Y. Hwang, Y. L. Hyun, Y. H. Jeon, S. Ro, J. M. Cho, T. G. Lee and C. H. Yang, "Structural basis for the selective inhibition of JNK1 by the scaffolding protein JIP1 and SP600125" *The EMBO journal*, vol. 23, no. 11, pp. 2185-2195, Jun 2. 2004.
58. K. Stierand and M. Rarey, "Drawing the PDB: Protein-Ligand Complexes in Two Dimensions" *ACS medicinal chemistry letters*, vol. 1, no. 9, pp. 540-545, Dec 9. 2010.
59. P. M. Traxler, "Protein tyrosine kinase inhibitors in cancer treatment" *Expert opinion on therapeutic patents*, vol. 7, no. 6, pp. 571-588, 1997/06/01. 1997.
60. R. K. Barr, T. S. Kendrick and M. A. Bogoyevitch, "Identification of the critical features of a small peptide inhibitor of JNK activity" *The Journal of biological chemistry*, vol. 277, no. 13, pp. 10987-10997, Mar 29. 2002.
61. R. K. Barr, I. Boehm, P. V. Attwood, P. M. Watt and M. A. Bogoyevitch, "The critical features and the mechanism of inhibition of a kinase interaction motif-based peptide inhibitor of JNK" *The Journal of biological chemistry*, vol. 279, no. 35, pp. 36327-36338, Aug 27. 2004.
62. J. L. Stebbins, S. K. De, T. Machleidt, B. Becattini, J. Vazquez, C. Kuntzen, L. H. Chen, J. F. Cellitti, M. Riel-Mehan, A. Emdadi, G. Solinas, M. Karin and M. Pellecchia, "Identification of a new JNK inhibitor targeting the JNK-JIP interaction site" *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 43, pp. 16809-16813, Oct 28. 2008.

63. P. J. Hajduk, J. R. Huth and S. W. Fesik, "Druggability indices for protein targets derived from NMR-based screening data" *Journal of medicinal chemistry*, vol. 48, no. 7, pp. 2518-2525, Apr 7. 2005.
64. C. Pargellis, L. Tong, L. Churchill, P. F. Cirillo, T. Gilmore, A. G. Graham, P. M. Grob, E. R. Hickey, N. Moss, S. Pav and J. Regan, "Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site" *Nature structural biology*, vol. 9, no. 4, pp. 268-272, Apr. 2002.
65. A. Kuglstatter, M. Ghate, S. Tsing, A. G. Villasenor, D. Shaw, J. W. Barnett and M. F. Browner, "X-ray crystal structure of JNK2 complexed with the p38alpha inhibitor BIRB796: insights into the rational design of DFG-out binding MAP kinase inhibitors" *Bioorganic & medicinal chemistry letters*, vol. 20, no. 17, pp. 5217-5220, Sep 1. 2010.
66. B. Klebl, G. Müller and M. Hamacher (2011) Protein kinases as drug targets. Wiley-VCH, Weinheim
67. "A unified nomenclature system for the nuclear receptor superfamily" *Cell*, vol. 97, no. 2, pp. 161-163, Apr 16. 1999.
68. S. Kakizaki, Y. Yamazaki, D. Takizawa and M. Negishi, "New insights on the xenobiotic-sensing nuclear receptors in liver diseases--CAR and PXR" *Current drug metabolism*, vol. 9, no. 7, pp. 614-621, Sep. 2008.
69. J. M. Pascussi, L. Drocourt, S. Gerbal-Chaloin, J. M. Fabre, P. Maurel and M. J. Vilarem, "Dual effect of dexamethasone on CYP3A4 gene expression in human hepatocytes. Sequential role of glucocorticoid receptor and pregnane X receptor" *European journal of biochemistry / FEBS*, vol. 268, no. 24, pp. 6346-6358, Dec. 2001.
70. B. D. Wallace, L. Betts, G. Talmage, R. M. Pollet, N. S. Holman and M. R. Redinbo, "Structural and functional analysis of the human nuclear xenobiotic receptor PXR in complex with RXRalpha" *Journal of molecular biology*, vol. 425, no. 14, pp. 2561-2577, Jul 24. 2013.
71. T. M. Devlin (2011) Textbook of biochemistry : with clinical correlations. 7th ed. John Wiley & Sons, Hoboken, NJ
72. Y. D. Gao, S. H. Olson, J. M. Balkovec, Y. Zhu, I. Royo, J. Yabut, R. Evers, E. Y. Tan, W. Tang, D. P. Hartley and R. T. Mosley, "Attenuating pregnane X receptor (PXR) activation: a molecular modelling approach" *Xenobiotica; the fate of foreign compounds in biological systems*, vol. 37, no. 2, pp. 124-138, Feb. 2007.
73. G. Luo, T. Guenther, L. S. Gan and W. G. Humphreys, "CYP3A4 induction by xenobiotics: biochemistry, experimental methods and impact on drug discovery and development" *Current drug metabolism*, vol. 5, no. 6, pp. 483-505, Dec. 2004.
74. A. J. Lau, G. Yang, C. W. Yap and T. K. Chang, "Selective agonism of human pregnane X receptor by individual ginkgolides" *Drug metabolism and disposition: the biological fate of chemicals*, vol. 40, no. 6, pp. 1113-1121, Jun. 2012.
75. R. E. Watkins, G. B. Wisely, L. B. Moore, J. L. Collins, M. H. Lambert, S. P. Williams, T. M. Willson, S. A. Kliewer and M. R. Redinbo, "The human nuclear xenobiotic receptor PXR: structural determinants of



- directed promiscuity" *Science*, vol. 292, no. 5525, pp. 2329-2333, Jun 22. 2001.
76. S. Mani, W. Dou and M. R. Redinbo, "PXR antagonists and implication in drug metabolism" *Drug metabolism reviews*, vol. 45, no. 1, pp. 60-72, Feb. 2013.
  77. L. E. Anderson, A. M. Dring, L. D. Hamel and M. A. Stoner, "Modulation of constitutive androstane receptor (CAR) and pregnane X receptor (PXR) by 6-arylpyrrolo[2,1-d][1,5]benzothiazepine derivatives, ligands of peripheral benzodiazepine receptor (PBR)" *Toxicology letters*, vol. 202, no. 2, pp. 148-154, Apr 25. 2011.
  78. S. Ekins, E. J. Reschly, L. R. Hagey and M. D. Krasowski, "Evolution of pharmacologic specificity in the pregnane X receptor" *BMC evolutionary biology*, vol. 8, no. pp. 103, 2008.
  79. R. Lahana, "How many leads from HTS?" *Drug discovery today*, vol. 4, no. 10, pp. 447-448, Oct. 1999.
  80. C. G. Wermuth (2008) *The practice of medicinal chemistry*. 3rd ed. Elsevier/Academic Press, Amsterdam ; Boston
  81. S. K. De, E. Barile, V. Chen, J. L. Stebbins, J. F. Cellitti, T. Machleidt, C. B. Carlson, L. Yang, R. Dahl and M. Pellecchia, "Design, synthesis, and structure-activity relationship studies of thiophene-3-carboxamide derivatives as dual inhibitors of the c-Jun N-terminal kinase" *Bioorganic & medicinal chemistry*, vol. 19, no. 8, pp. 2582-2588, Apr 15. 2011.
  82. S. K. De, L. H. Chen, J. L. Stebbins, T. Machleidt, M. Riel-Mehan, R. Dahl, V. Chen, H. Yuan, E. Barile, A. Emdadi, R. Murphy and M. Pellecchia, "Discovery of 2-(5-nitrothiazol-2-ylthio)benzo[d]thiazoles as novel c-Jun N-terminal kinase inhibitors" *Bioorganic & medicinal chemistry*, vol. 17, no. 7, pp. 2712-2717, Apr 1. 2009.
  83. S. K. De, V. Chen, J. L. Stebbins, L. H. Chen, J. F. Cellitti, T. Machleidt, E. Barile, M. Riel-Mehan, R. Dahl, L. Yang, A. Emdadi, R. Murphy and M. Pellecchia, "Synthesis and optimization of thiadiazole derivatives as a novel class of substrate competitive c-Jun N-terminal kinase inhibitors" *Bioorganic & medicinal chemistry*, vol. 18, no. 2, pp. 590-596, Jan 15. 2010.
  84. S. K. De, J. L. Stebbins, L. H. Chen, M. Riel-Mehan, T. Machleidt, R. Dahl, H. Yuan, A. Emdadi, E. Barile, V. Chen, R. Murphy and M. Pellecchia, "Design, synthesis, and structure-activity relationship of substrate competitive, selective, and in vivo active triazole and thiadiazole inhibitors of the c-Jun N-terminal kinase" *Journal of medicinal chemistry*, vol. 52, no. 7, pp. 1943-1952, Apr 9. 2009.
  85. Y. Sui, N. Ai, S. H. Park, J. Rios-Pilier, J. T. Perkins, W. J. Welsh and C. Zhou, "Bisphenol A and its analogues activate human pregnane X receptor" *Environmental health perspectives*, vol. 120, no. 3, pp. 399-405, Mar. 2012.
  86. J. E. Chrencik, J. Orans, L. B. Moore, Y. Xue, L. Peng, J. L. Collins, G. B. Wisely, M. H. Lambert, S. A. Kliewer and M. R. Redinbo, "Structural disorder in the complex of human pregnane X receptor and the macrolide

- antibiotic rifampicin" *Molecular endocrinology*, vol. 19, no. 5, pp. 1125-1134, May. 2005.
87. J. Kirchmair, S. Distinto, D. Schuster, G. Spitzer, T. Langer and G. Wolber, "Enhancing drug discovery through in silico screening: strategies to increase true positives retrieval rates" *Current medicinal chemistry*, vol. 15, no. 20, pp. 2040-2053, 2008.
  88. C. Liao, M. Sitzmann, A. Pugliese and M. C. Nicklaus, "Software and resources for computational medicinal chemistry" *Future medicinal chemistry*, vol. 3, no. 8, pp. 1057-1085, Jun. 2011.
  89. I. Muegge, "Synergies of virtual screening approaches" *Mini reviews in medicinal chemistry*, vol. 8, no. 9, pp. 927-933, Aug. 2008.
  90. G. L. Patrick and J. Spencer (2009) An introduction to medicinal chemistry. 4th ed. Oxford University Press, New York
  91. C. Sotriffer (2011) Virtual screening : principles, challenges, and practical guidelines. Volume 48. Wiley-VCH, Weinheim
  92. W. P. Walters, M. T. Stahl and M. A. Murcko, "Virtual screening—an overview" *Drug discovery today*, vol. 3, no. 4, pp. 160-178, 1998.
  93. C. Hansch and T. Fujita, "p- $\sigma$ - $\pi$  Analysis. A Method for the Correlation of Biological Activity and Chemical Structure" *Journal of the American Chemical Society*, vol. 86, no. 8, pp. 1616-1626, 1964/04/01. 1964.
  94. C. G. Wermuth, C. R. Ganellin, P. Lindberg and L. A. Mitscher, "Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998)" *Pure and applied Chemistry*, vol. 70, no. 5, pp. 1129-1143, 1998.
  95. I. Ijjaali, F. Petitet, E. Dubus, O. Barberan and A. Michel, "Assessing potency of c-Jun N-terminal kinase 3 (JNK3) inhibitors using 2D molecular descriptors and binary QSAR methodology" *Bioorganic & medicinal chemistry*, vol. 15, no. 12, pp. 4256-4264, Jun 15. 2007.
  96. V. Schattel, G. Hinselmann, A. Jahn, A. Zell and S. Laufer, "Modeling and benchmark data set for the inhibition of c-Jun N-terminal kinase-3" *Journal of chemical information and modeling*, vol. 51, no. 3, pp. 670-679, Mar 28. 2011.
  97. P. Sharma and N. Ghoshal, "Exploration of a binding mode of benzothiazol-2-yl acetonitrile pyrimidine core based derivatives as potent c-Jun N-terminal kinase-3 inhibitors and 3D-QSAR analyses" *Journal of chemical information and modeling*, vol. 46, no. 4, pp. 1763-1774, Jul-Aug. 2006.
  98. P. Yi and M. Qiu, "3D-QSAR and docking studies of aminopyridine carboxamide inhibitors of c-Jun N-terminal kinase-1" *European journal of medicinal chemistry*, vol. 43, no. 3, pp. 604-613, Mar. 2008.
  99. S. Ekins, S. Kortagere, M. Iyer, E. J. Reschly, M. A. Lill, M. R. Redinbo and M. D. Krasowski, "Challenges predicting ligand-receptor interactions of promiscuous proteins: the nuclear receptor PXR" *PLoS computational biology*, vol. 5, no. 12, pp. e1000594, Dec. 2009.
  100. A. Khandelwal, M. D. Krasowski, E. J. Reschly, M. W. Sinz, P. W. Swaan and S. Ekins, "Machine learning methods and docking for predicting

- human pregnane X receptor activation" *Chemical research in toxicology*, vol. 21, no. 7, pp. 1457-1467, Jul. 2008.
101. A. R. Leach, V. J. Gillet, R. A. Lewis and R. Taylor, "Three-dimensional pharmacophore methods in drug discovery" *Journal of medicinal chemistry*, vol. 53, no. 2, pp. 539-558, Jan 28. 2010.
  102. I. J. Enyedy and W. J. Egan, "Can we use docking and scoring for hit-to-lead optimization?" *Journal of computer-aided molecular design*, vol. 22, no. 3-4, pp. 161-168, Mar-Apr. 2008.
  103. G. L. Warren, C. W. Andrews, A. M. Capelli, B. Clarke, J. LaLonde, M. H. Lambert, M. Lindvall, N. Nevins, S. F. Semus, S. Senger, G. Tedesco, I. D. Wall, J. M. Woolven, C. E. Peishoff and M. S. Head, "A critical assessment of docking programs and scoring functions" *Journal of medicinal chemistry*, vol. 49, no. 20, pp. 5912-5931, Oct 5. 2006.
  104. T. Cheng, X. Li, Y. Li, Z. Liu and R. Wang, "Comparative Assessment of Scoring Functions on a Diverse Test Set" *Journal of chemical information and modeling*, vol. 49, no. 4, pp. 1079-1093, 2009/04/27. 2009.
  105. D. Plewczynski, M. Lazniewski, R. Augustyniak and K. Ginalski, "Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database" *Journal of computational chemistry*, vol. 32, no. 4, pp. 742-755, Mar. 2011.
  106. M. Karplus, "Molecular dynamics simulations of biomolecules" *Accounts of chemical research*, vol. 35, no. 6, pp. 321-323, Jun. 2002.
  107. R. Schleif, "Modeling and studying proteins with molecular dynamics" *Methods in enzymology*, vol. 383, no. pp. 28-47, 2004.
  108. K. R. Ngoei, D. C. Ng, P. R. Gooley, D. P. Fairlie, M. J. Stoermer and M. A. Bogoyevitch, "Identification and characterization of bi-thiazole-2,2'-diamines as kinase inhibitory scaffolds" *Biochimica et biophysica acta*, vol. 1834, no. 6, pp. 1077-1088, Jun. 2013.
  109. Y. Cheng and M. R. Redinbo, "Activation of the human nuclear xenobiotic receptor PXR by the reverse transcriptase-targeted anti-HIV drug PNU-142721" *Protein science : a publication of the Protein Society*, vol. 20, no. 10, pp. 1713-1719, Oct. 2011.
  110. W. Wang, W. W. Prosser, J. Chen, S. S. Taremi, H. V. Le, V. Madison, X. Cui, A. Thomas, K. C. Cheng and C. A. Lesburg, "Construction and characterization of a fully active PXR/SRC-1 tethered protein with increased stability" *Protein engineering, design & selection : PEDS*, vol. 21, no. 7, pp. 425-433, Jul. 2008.
  111. R. E. Watkins, P. R. Davis-Searles, M. H. Lambert and M. R. Redinbo, "Coactivator binding promotes the specific interaction between ligand and the pregnane X receptor" *Journal of molecular biology*, vol. 331, no. 4, pp. 815-828, Aug 22. 2003.
  112. R. E. Watkins, J. M. Maglich, L. B. Moore, G. B. Wisely, S. M. Noble, P. R. Davis-Searles, M. H. Lambert, S. A. Kliewer and M. R. Redinbo, "2.1 A crystal structure of human PXR in complex with the St. John's wort compound hyperforin" *Biochemistry*, vol. 42, no. 6, pp. 1430-1438, Feb 18. 2003.

113. Y. Xue, E. Chao, W. J. Zuercher, T. M. Willson, J. L. Collins and M. R. Redinbo, "Crystal structure of the PXR-T1317 complex provides a scaffold to examine the potential for receptor antagonism" *Bioorganic & medicinal chemistry*, vol. 15, no. 5, pp. 2156-2166, Mar 1. 2007.
114. R. E. Watkins, S. M. Noble and M. R. Redinbo, "Structural insights into the promiscuity and function of the human pregnane X receptor" *Current opinion in drug discovery & development*, vol. 5, no. 1, pp. 150-158, Jan. 2002.
115. A. Berthier, F. Oger, C. Gheeraert, A. Boulahtouf, R. Le Guevel, P. Balaguer, B. Staels, G. Salbert and P. Lefebvre, "The novel antibacterial compound walrycin A induces human PXR transcriptional activity" *Toxicological sciences : an official journal of the Society of Toxicology*, vol. 127, no. 1, pp. 225-235, May. 2012.
116. S. Ekins, C. Chang, S. Mani, M. D. Krasowski, E. J. Reschly, M. Iyer, V. Kholodovych, N. Ai, W. J. Welsh, M. Sinz, P. W. Swaan, R. Patel and K. Bachmann, "Human pregnane X receptor antagonists and agonists define molecular requirements for different binding sites" *Molecular pharmacology*, vol. 72, no. 3, pp. 592-603, Sep. 2007.
117. V. Sepe, R. Ummarino, M. V. D'Auria, G. Lauro, G. Bifulco, C. D'Amore, B. Renga, S. Fiorucci and A. Zampella, "Modification in the side chain of solomonsterol A: discovery of cholestan disulfate as a potent pregnane-X-receptor agonist" *Organic & biomolecular chemistry*, vol. 10, no. 31, pp. 6350-6362, Aug 21. 2012.
118. D. G. Teotico, M. L. Frazier, F. Ding, N. V. Dokholyan, B. R. Temple and M. R. Redinbo, "Active nuclear receptors exhibit highly correlated AF-2 domain motions" *PLoS computational biology*, vol. 4, no. 7, pp. e1000111, 2008.
119. M. A. Bogoyevitch, "Therapeutic promise of JNK ATP-noncompetitive inhibitors" *Trends in molecular medicine*, vol. 11, no. 5, pp. 232-239, 2005.
120. M. J. Schnieders, T. S. Kaoud, C. Yan, K. N. Dalby and P. Ren, "Computational insights for the discovery of non-ATP competitive inhibitors of MAP kinases" *Current pharmaceutical design*, vol. 18, no. 9, pp. 1173-1185, 2012.
121. L. Svecova, R. Vrzal, L. Burysek, E. Anzenbacherova, L. Cerveny, J. Grim, F. Trejtnar, J. Kunes, M. Pour, F. Staud, P. Anzenbacher, Z. Dvorak and P. Pavek, "Azole antimycotics differentially affect rifampicin-induced pregnane X receptor-mediated CYP3A4 gene expression" *Drug metabolism and disposition: the biological fate of chemicals*, vol. 36, no. 2, pp. 339-348, Feb. 2008.
122. W. Alderton, S. Berghmans, P. Butler, H. Chassaing, A. Fleming, Z. Golder, F. Richards and I. Gardner, "Accumulation and metabolism of drugs and CYP probe substrates in zebrafish larvae" *Xenobiotica; the fate of foreign compounds in biological systems*, vol. 40, no. 8, pp. 547-557, Aug. 2010.

123. H. M. Bolt, "Rifampicin, a keystone inducer of drug metabolism: from Herbert Remmer's pioneering ideas to modern concepts" *Drug metabolism reviews*, vol. 36, no. 3-4, pp. 497-509, Oct. 2004.
124. H. T. Chng, H. K. Ho, C. W. Yap, S. H. Lam and E. C. Chan, "An investigation of the bioactivation potential and metabolism profile of Zebrafish versus human" *Journal of biomolecular screening*, vol. 17, no. 7, pp. 974-986, Aug. 2012.
125. H. P. Tseng, T. H. Hseu, D. R. Buhler, W. D. Wang and C. H. Hu, "Constitutive and xenobiotics-induced expression of a novel CYP3A gene from zebrafish larva" *Toxicology and applied pharmacology*, vol. 205, no. 3, pp. 247-258, Jun 15. 2005.
126. H. T. Chng, R. F. S. Lee, S. H. Lam, Z. Gong, H. K. Ho and E. C. Y. Chan. Zebrafish: An *in-vivo* High-throughput Screening Model for CYP3A4 Induction and Inhibition. 19th MDO Meeting and 12th European ISSX Meeting. The Netherlands, 2012
127. K. Yao, Y. Y. Cho, A. M. Bode, A. Vummenthala, J. G. Park, K. Liu, Y. P. Pang and Z. Dong, "A selective small-molecule inhibitor of c-Jun N-terminal kinase 1" *FEBS letters*, vol. 583, no. 13, pp. 2208-2212, Jul 7. 2009.
128. C. Y. Liew, X. H. Ma, X. Liu and C. W. Yap, "SVM model for virtual screening of Lck inhibitors" *Journal of chemical information and modeling*, vol. 49, no. 4, pp. 877-885, Apr. 2009.
129. I. V. Tetko, I. Sushko, A. K. Pandey, H. Zhu, A. Tropsha, E. Papa, T. Oberg, R. Todeschini, D. Fourches and A. Varnek, "Critical assessment of QSAR models of environmental toxicity against *Tetrahymena pyriformis*: focusing on applicability domain and overfitting by variable selection" *Journal of chemical information and modeling*, vol. 48, no. 9, pp. 1733-1746, Sep. 2008.
130. ChemDraw Pro. 12.0. Waltham, MA, USA: PerkinElmer Informatics; 2011
131. C. W. Yap, "PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints" *Journal of computational chemistry*, vol. 32, no. 7, pp. 1466-1474, May. 2011.
132. B. Hoffman, S. J. Cho, W. Zheng, S. Wyrick, D. E. Nichols, R. B. Mailman and A. Tropsha, "Quantitative structure-activity relationship modeling of dopamine D(1) antagonists using comparative molecular field analysis, genetic algorithms-partial least-squares, and K nearest neighbor methods" *Journal of medicinal chemistry*, vol. 42, no. 17, pp. 3217-3226, Aug 26. 1999.
133. A. Tropsha and A. Golbraikh, "Predictive QSAR modeling workflow, model applicability domains, and virtual screening" *Current pharmaceutical design*, vol. 13, no. 34, pp. 3494-3504, 2007.
134. R. Nisbet, J. F. Elder and G. Miner (2009) Handbook of statistical analysis and data mining applications. Academic Press/Elsevier, Amsterdam ; Boston

135. A. Bender, H. Y. Mussa, R. C. Glen and S. Reiling, "Similarity searching of chemical databases using atom environment descriptors (MOLPRINT 2D): evaluation of performance" *Journal of chemical information and computer sciences*, vol. 44, no. 5, pp. 1708-1718, Sep-Oct. 2004.
136. C. C. Chang and C. J. Lin, "LIBSVM: A Library for Support Vector Machines" *Acm Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 2011.
137. C. W. Hsu, C. C. Chang and C. J. Lin. A Practical Guide to Support Vector Classification <http://www.csie.ntu.edu.tw/~cjlin/>. Accessed 2013. National Taiwan University, Taipei, Taiwan, 2010
138. G. Fumera, F. Roli and G. Giacinto, "Reject option with multiple thresholds" *Pattern Recognition*, vol. 33, no. 12, pp. 2099-2101, Dec. 2000.
139. Y. He, S. W. Lim and C. W. Yap, "Determination of torsade-causing potential of drug candidates using one-class classification and ensemble modelling approaches" *Current drug safety*, vol. 7, no. 4, pp. 298-308, Sep. 2012.
140. N. Sharma and C. W. Yap, "Consensus QSAR model for identifying novel H5N1 inhibitors" *Molecular diversity*, vol. 16, no. 3, pp. 513-524, Aug. 2012.
141. C. Y. Liew, X. H. Ma and C. W. Yap, "Consensus model for identification of novel PI3K inhibitors in large chemical library" *Journal of computer-aided molecular design*, vol. 24, no. 2, pp. 131-141, Feb. 2010.
142. C. W. Yap and Y. Z. Chen, "Prediction of cytochrome P450 3A4, 2D6, and 2C9 inhibitors and substrates by using support vector machines" *Journal of chemical information and modeling*, vol. 45, no. 4, pp. 982-992, Jul-Aug. 2005.
143. J. J. Irwin and B. K. Shoichet, "ZINC--a free database of commercially available compounds for virtual screening" *Journal of chemical information and modeling*, vol. 45, no. 1, pp. 177-182, Jan-Feb. 2005.
144. J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad and R. G. Coleman, "ZINC: A Free Tool to Discover Chemistry for Biology" *Journal of chemical information and modeling*, vol. no. pp. Jun 15. 2012.
145. N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, "Open Babel: An open chemical toolbox" *Journal of cheminformatics*, vol. 3, no. pp. 33, 2011.
146. Molecular Descriptors Guide: Description of the Molecular Descriptors Appearing in the Toxicity Estimation Software Tool Version 1.02. <http://www.epa.gov/nrmrl/std/qsar/MolecularDescriptorsGuide-v102.pdf>. Accessed 2013. United States Environmental Protection Agency, Washington, D. C., 2008
147. A. K. Ghose and G. M. Crippen, "Atomic Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure-Activity Relationships I. Partition Coefficients as a Measure of Hydrophobicity" *Journal of computational chemistry*, vol. 7, no. 4, pp. 565-577, 1986.

148. L. B. Kier and L. H. Hall (1999) *Molecular Structure Description: The Electrotopological State*. Academic Press, San Diego
149. G. Moreau and P. Broto, "The autocorrelation of the topological molecular structure: a new molecular descriptor" *Nouveau Journal de Chimie*, vol. 4, no. pp. 359-360, 1980.
150. K. Roy and G. Ghosh, "Introduction of Extended Topochemical Atom (ETA) Indices in the Valence Electron Mobile (VEM) Environment as Tools for QSAR/QSPR Studies" *Internet Electronic Journal of Molecular Design*, vol. 2, no. 9, pp. 599-620, 2003.
151. R. Todeschini and V. Consonni (2009) *Molecular Descriptors for Cheminformatics*. Wiley-VCH, Weinheim
152. P. Labute, "Protonate3D: Assignment of ionization states and hydrogen coordinates to macromolecular structures" *Proteins: Structure, Function, and Bioinformatics*, vol. 75, no. 1, pp. 187-205, 2009.
153. *Molecular Operating Environment*. 2012.10. Montreal, Quebec, Canada: Chemical Computing Group Inc.; 2013
154. J. W. Ponder and D. A. Case, "Force fields for protein simulations" *Advances in protein chemistry*, vol. 66, no. pp. 27-85, 2003.
155. H. Edelsbrunner. Weighted alpha shapes. Technical paper of the Department of Computer Science of the University of Illinois at Urbana-Champaign. Urbana, Illinois
156. C. R. Corbeil, C. I. Williams and P. Labute, "Variability in docking success rates due to dataset preparation" *Journal of computer-aided molecular design*, vol. 26, no. 6, pp. 775-786, Jun. 2012.
157. M. Naim, S. Bhat, K. N. Rankin, S. Dennis, S. F. Chowdhury, I. Siddiqi, P. Drabik, T. Sulea, C. I. Bayly, A. Jakalian and E. O. Purisima, "Solvated interaction energy (SIE) for scoring protein-ligand binding affinities. 1. Exploring the parameter space" *Journal of chemical information and modeling*, vol. 47, no. 1, pp. 122-133, Jan-Feb. 2007.
158. P. Labute, "The generalized Born/volume integral implicit solvent model: Estimation of the free energy of hydration using London dispersion instead of atomic surface area" *Journal of computational chemistry*, vol. 29, no. 10, pp. 1693-1698, 2008.
159. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, "Comparison of simple potential functions for simulating liquid water" *The Journal of chemical physics*, vol. 79, no. 2, pp. 926-935, 1983.
160. D. A. Case, T. A. Darden, T. E. I. Cheatham, 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. AMBER 12. University of California, San Francisco, 2012
161. Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman,

- "A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations" *Journal of computational chemistry*, vol. 24, no. 16, pp. 1999-2012, Dec. 2003.
162. J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes" *Journal of Computational Physics*, vol. 23, no. 3, pp. 327-341, 1977.
163. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, "UCSF Chimera--a visualization system for exploratory research and analysis" *Journal of computational chemistry*, vol. 25, no. 13, pp. 1605-1612, Oct. 2004.
164. T. A. Halgren, "Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94" *Journal of computational chemistry*, vol. 17, no. 5-6, pp. 490-519, 1996.
165. P. Labute. An Integrated Application in MOE for the Visualization and Analysis of Protein Active Sites with Molecular Surfaces, Contact Statistics and Electrostatic Maps  
[https://www.chemcomp.com/journal/f\\_surfmap.htm](https://www.chemcomp.com/journal/f_surfmap.htm). Accessed 2016. Chemical Computing Group Inc., Montreal, Quebec, Canada, 1996
166. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility" *Journal of computational chemistry*, vol. 30, no. 16, pp. 2785-2791, Dec. 2009.
167. R. A. Laskowski and M. B. Swindells, "LigPlot+: multiple ligand-protein interaction diagrams for drug discovery" *Journal of chemical information and modeling*, vol. 51, no. 10, pp. 2778-2786, Oct 24. 2011.
168. JChem. Budapest, Hungary: ChemAxon Kft.; 2012
169. SciFinder. Columbus, Ohio, USA: Chemical Abstracts Service; 2012
170. D. A. Dougherty, "Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp" *Science*, vol. 271, no. 5246, pp. 163-168, Jan 12. 1996.
171. J. P. Gallivan and D. A. Dougherty, "Cation-pi interactions in structural biology" *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 17, pp. 9459-9464, Aug 17. 1999.
172. I. Y. Torshin, I. T. Weber and R. W. Harrison, "Geometric criteria of hydrogen bonds in proteins and identification of "bifurcated" hydrogen bonds" *Protein engineering*, vol. 15, no. 5, pp. 359-363, May. 2002.
173. D. Shaw, S. M. Wang, A. G. Villasenor, S. Tsing, D. Walter, M. F. Browner, J. Barnett and A. Kuglstatter, "The crystal structure of JNK2 reveals conformational flexibility in the MAP kinase insert and indicates its involvement in the regulation of catalytic activity" *Journal of molecular biology*, vol. 383, no. 4, pp. 885-893, Nov 21. 2008.
174. J. Wang, P. Cieplak and P. A. Kollman, "How well does a restrained electrostatic potential (RESP) model perform in calculating



- conformational energies of organic and biological molecules?" *Journal of computational chemistry*, vol. 21, no. 12, pp. 1049-1074, 2000.
175. N. Foloppe and J. A. D. MacKerell, "All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data" *Journal of computational chemistry*, vol. 21, no. 2, pp. 86-104, 2000.
176. J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad and R. G. Coleman, "ZINC: a free tool to discover chemistry for biology" *Journal of chemical information and modeling*, vol. 52, no. 7, pp. 1757-1768, Jul 23. 2012.
177. M. A. Fabian, W. H. Biggs, 3rd, D. K. Treiber, C. E. Atteridge, M. D. Azimioara, M. G. Benedetti, T. A. Carter, P. Ciceri, P. T. Edeen, M. Floyd, J. M. Ford, M. Galvin, J. L. Gerlach, R. M. Grotzfeld, S. Herrgard, D. E. Insko, M. A. Insko, A. G. Lai, J. M. Lelias, S. A. Mehta, Z. V. Milanov, A. M. Velasco, L. M. Wodicka, H. K. Patel, P. P. Zarrinkar and D. J. Lockhart, "A small molecule-kinase interaction map for clinical kinase inhibitors" *Nature biotechnology*, vol. 23, no. 3, pp. 329-336, Mar. 2005.
178. T. J. Hou, K. Xia, W. Zhang and X. J. Xu, "ADME evaluation in drug discovery. 4. Prediction of aqueous solubility based on atom contribution approach" *Journal of chemical information and computer sciences*, vol. 44, no. 1, pp. 266-275, Jan-Feb. 2004.
179. R. van Deursen, L. C. Blum and J. L. Reymond, "Visualisation of the chemical space of fragments, lead-like and drug-like molecules in PubChem" *Journal of computer-aided molecular design*, vol. 25, no. 7, pp. 649-662, Jul. 2011.
180. Y. H. Liu, S. L. Mo, H. C. Bi, B. F. Hu, C. G. Li, Y. T. Wang, L. Huang, M. Huang, W. Duan, J. P. Liu, M. Q. Wei and S. F. Zhou, "Regulation of human pregnane X receptor and its target gene cytochrome P450 3A4 by Chinese herbal compounds and a molecular docking study" *Xenobiotica; the fate of foreign compounds in biological systems*, vol. 41, no. 4, pp. 259-280, Apr. 2011.
181. D. Sharma, A. J. Lau, M. A. Sherman and T. K. Chang, "Agonism of human pregnane X receptor by rilpivirine and etravirine: comparison with first generation non-nucleoside reverse transcriptase inhibitors" *Biochemical pharmacology*, vol. 85, no. 11, pp. 1700-1711, Jun 1. 2013.
182. Drug Development and Drug Interactions: Table of Substrates, Inhibitors and Inducers  
<http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DevelopmentResources/DrugInteractionsLabeling/ucm093664.htm#4>. Accessed. US. Department of Health and Human Services, 2012
183. C. The UniProt, "Reorganizing the protein space at the Universal Protein Resource (UniProt)" *Nucleic acids research*, vol. 40, no. D1, pp. D71-D75, 2012.
184. Maestro. 10.3. New York, NY, USA: Schrödinger, LLC; 2015
185. J. Kyte (2007) Structure in protein chemistry. 2nd ed. Garland Science, New York

186. G. W. Milligan, "An examination of the effect of six types of error perturbation on fifteen clustering algorithms" *Psychometrika*, vol. 45, no. 3, pp. 325-342, 1980/09/01. 1980.
187. J. H. Ward, "Hierarchical Grouping to Optimize an Objective Function" *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236-244, 1963/03/01. 1963.
188. JMP. 10. Cary, NC, USA: SAS Institute Inc.; 2013
189. P. Ripphausen, B. Nisius, L. Peltason and J. Bajorath, "Quo vadis, virtual screening? A comprehensive survey of prospective applications" *Journal of medicinal chemistry*, vol. 53, no. 24, pp. 8461-8467, Dec 23. 2010.
190. A. Varnek and A. Tropsha (2008) *Chemoinformatics Approaches to Virtual Screening*. The Royal Society of Chemistry, Cambridge
191. C.-Y. Kuan, D. D. Yang, D. R. S. Roy, R. J. Davis, P. Rakic and R. A. Flavell, "The Jnk1 and Jnk2 Protein Kinases Are Required for Regional Specific Apoptosis during Early Brain Development" *Neuron*, vol. 22, no. 4, pp. 667-676, 1999.
192. L. Resnick and M. Fennell, "Targeting JNK3 for the treatment of neurodegenerative disorders" *Drug discovery today*, vol. 9, no. 21, pp. 932-939, Nov 1. 2004.