

**QSAR METHODS DEVELOPMENT, VIRTUAL AND
EXPERIMENTAL SCREENING FOR CANNABINOID LIGAND
DISCOVERY**

by

Kyaw Zeyar Myint

BS, Biology, BS, Computer Science, Hampden-Sydney College, 2007

Submitted to the Graduate Faculty of
School of Medicine in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Kyaw Zeyar Myint

It was defended on

August 20th, 2012

and approved by

Dr. Ivet Bahar, Professor, Department of Computational and Systems Biology

Dr. Billy W. Day, Professor, Department of Pharmaceutical Sciences

Dr. Christopher Langmead, Associate Professor, Department of Computer Science, CMU

Dissertation Advisor: Dr. Xiang-Qun Xie, Professor, Department of Pharmaceutical Sciences

Copyright © by Kyaw Zeyar Myint

2012

**QSAR METHODS DEVELOPMENT, VIRTUAL AND EXPERIMENTAL
SCREENING FOR CANNABINOID LIGAND DISCOVERY**

Kyaw Zeyar Myint, PhD

University of Pittsburgh, 2012

G protein coupled receptors (GPCRs) are the largest receptor family in mammalian genomes and are known to regulate wide variety of signals such as ions, hormones and neurotransmitters. It has been estimated that GPCRs represent more than 30% of current drug targets and have attracted many pharmaceutical industries as well as academic groups for potential drug discovery. Cannabinoid (CB) receptors, members of GPCR superfamily, are also involved in the activation of multiple intracellular signal transductions and their endogenous ligands or cannabinoids have attracted pharmacological research because of their potential therapeutic effects. In particular, the cannabinoid subtype-2 (CB2) receptor is known to be involved in immune system signal transductions and its ligands have the potential to be developed as drugs to treat many immune system disorders without potential psychotic side-effects. Therefore, this work was focused on discovering novel CB2 ligands by developing novel quantitative structure-activity relationship (QSAR) methods and performing virtual and experimental screenings. Three novel QSAR methods were developed to predict biological activities and binding affinities of ligands. In the first method, a traditional fragment-based approach was improved by introducing a fragment similarity concept that enhanced the prediction accuracy remarkably. In the second method, pharmacophoric and morphological descriptors were incorporated to derive a novel QSAR regression model with good prediction accuracy. In the third method, a novel fingerprint-based artificial neural network QSAR model

was developed to overcome the similar scaffold requirement of many fragment-based and other 3D-QSAR methods. These methods provide a foundation for virtual screening and hit ranking of chemical ligands from large chemical space. In addition, several novel CB2 selective ligands within nM binding affinities were discovered. These ligands were proven to be inverse agonists as validated by functional assays and could be useful probes to study CB2 signaling as well as potential drug candidates for autoimmune diseases.

TABLE OF CONTENTS

PREFACE.....	XV
1.0 INTRODUCTION.....	1
1.1 RECENT ADVANCES IN QSAR METHODS	1
1.1.1 Introduction.....	1
1.1.2 Fragment-based 2D-QSAR methods.....	5
1.1.3 3D-QSAR	11
1.1.4 Comparison of 2D or fragment-based QSAR versus 3D or nD-QSAR methods 22	
1.2 VIRTUAL SCREENING APPROACHES.....	25
1.2.1 Structure-based method	25
1.2.2 Ligand-based method	26
1.3 CANNABINOID RECEPTORS AND THEIR LIGANDS	32
1.3.1 Background and significance	32
1.3.2 Computational design of CB1 and CB2 receptors	33
1.3.3 Computational design and screening of cannabinoid ligands	37
1.4 OUTLINE OF THE DISSERTATION.....	41
2.0 FRAGMENT-SIMILARITY-BASED QSAR (FS-QSAR) ALGORITHM FOR LIGAND BIOLOGICAL ACTIVITY PREDICTIONS	44

2.1	INTRODUCTION	44
2.2	METHODS	46
2.2.1	Data sets	46
2.2.2	Computational method	46
2.3	CALCULATIONS	50
2.3.1	Partial charge calculation and fragment generation	50
2.3.2	Parameter tuning using leave-one-out cross-validation (LOOCV).....	51
2.3.3	Generation of training and testing data sets.....	52
2.4	RESULTS AND DISCUSSION	52
2.4.1	FS-QSAR modeling on COX2 inhibitor analogs.....	52
2.4.2	FS-QSAR modeling on bis-sulfone analogs	53
2.4.3	BCUT-similarity score analysis	53
2.4.4	Comparisons with different approaches.....	55
2.5	CONCLUSION	58
3.0	NEW QSAR PREDICTION MODELS DERIVED FROM GPCR CB2- ANTAGONISTIC TRIARYL BIS-SULFONE ANALOGS BY A COMBINED MOLECULAR MORPHOLOGICAL AND PHARMACOPHORIC APPROACH	73
3.1	INTRODUCTION	73
3.2	METHODS	75
3.2.1	Pharmacophore-based molecular similarity calculation.....	75
3.2.2	Morphology-based molecular similarity calculation	76
3.2.3	QSAR model generation.....	77
3.3	RESULTS AND DISCUSSION	79

3.3.1	Generation of the pharmacophore model and score of the pharmacophoric match.....	79
3.3.2	Generation of hypermolecular alignment and scoring of shape-based molecular similarity	82
3.3.3	Development of the PharmShape algorithm based on the QSAR prediction model.....	84
3.4	CONCLUSION	89
4.0	FINGERPRINT-BASED ARTIFICIAL NEURAL NETWORKS QSAR (FANN-QSAR) FOR LIGAND BIOLOGICAL ACTIVITY PREDICTIONS.....	105
4.1	INTRODUCTION	105
4.2	METHODS.....	106
4.2.1	Data sets	106
4.2.2	Fingerprint generation	110
4.2.3	Fingerprint-based Artificial Neural Network QSAR	110
4.2.4	Radioligand competition binding assay	112
4.3	RESULTS AND DISCUSSION	113
4.3.1	Comparisons with other 3D and 2D-QSAR methods	114
4.3.2	Prediction of Cannabinoid receptor binding activity using FANN-QSAR method.....	117
4.3.3	Generalization ability of FANN-QSAR method on newly reported cannabinoid ligands	120
4.3.4	An application of FANN-QSAR: virtual screening of the NCI compound database.....	123

4.4	CONCLUSION	127
5.0	MOLECULAR MODELING, DISCOVERY AND QSAR STUDY OF NOVEL CANNABINOID LIGANDS.....	129
5.1	INTRODUCTION	129
5.2	EXPERIMENTAL AND COMPUTATIONAL METHODS	131
5.2.1	Chemistry.....	131
5.2.2	Radioligand competition binding assays.....	132
5.2.3	cAMP assays	133
5.2.4	Osteoclast formation assay	134
5.2.5	Cytotoxicity assay on human mononuclear cells.....	135
5.2.6	Molecular docking and 3D-QSAR model development	135
5.3	RESULTS AND DISCUSSION	138
5.3.1	Pharmacology and SAR analysis.....	138
5.3.2	Functional bioactivity at CB2 receptors in vitro.....	142
5.3.3	Osteoclast Formation Assay.....	143
5.3.4	Cytotoxicity analysis using normal human cells	143
5.3.5	Molecular docking and 3D-QSAR studies.....	144
5.4	CONCLUSION	147
6.0	SUMMARY AND FUTURE WORK	164
	APPENDIX A.....	166
	BIBLIOGRAPHY.....	169

LIST OF TABLES

Table 1-1. Summary of different QSAR methods and source information.	23
Table 1-2. List of selected structure-based docking programs	29
Table 1-3. List of selected ligand-based programs	30
Table 2-1. COX2 analogs used for FS-QSAR modeling.....	60
Table 2-2. Bis-sulfone analogs (I)	62
Table 2-3. Bis-sulfone analogs (II).....	63
Table 2-4. Bis-sulfone analogs (III).....	64
Table 2-5. A list of COX2 analogs for 5 different testing sets	65
Table 2-6. A list of bis-sulfone analogs for 5 different testing sets	65
Table 2-7. Results of leave-one-out cross-validation (LOOCV) on COX2 analogs	66
Table 2-8. Results of leave-one-out cross-validation (LOOCV) on bis-sulfone analogs	66
Table 2-9. Result summary on five COX2 testing sets.....	66
Table 2-10. Result summary on five bis-sulfone testing sets	67
Table 2-11. Fragment pairs and their BCUT-similarity scores for each bis-sulfone testing set... 68	
Table 2-12. Fragment pairs chosen in Figure 2-1 and their corresponding similarity scores..... 72	
Table 3-1. The structures and K_i values of the Triaryl bis-sulfones used as a training set to generate QSAR prediction model.....	90

Table 3-2. The structures and K_i values of the Triaryl bis-sulfones used to test the generated QSAR prediction model.....	91
Table 3-3. Pharmacophore-based molecular similarity score, shape-based molecular similarity score and the calculated molecular properties for the training set of triaryl bis-sulfones listed in Table 3-1.	94
Table 3-4. Pharmacophore-based molecular similarity score, shape-based molecular similarity score and the calculated molecular properties for the testing set of triaryl bis-sulfones listed in Table 3-2.	96
Table 3-5. Partial Least Squares Analysis to generate relationships between bioactive data and combinations of their pharmacophore-based molecular similarity, shape-based molecular similarity and calculated molecular property descriptors	98
Table 4-1. Numbers of training, validation and test set compounds in each data set.	109
Table 4-2. FANN-QSAR performance comparisons with other QSAR methods	115
Table 4-3. A summary of the performance of each FANN-QSAR model on CB ₂ ligand data set	119
Table 4-4. Identified NCI hit compounds with CB ₂ binding activities.....	125
Table 5-1. Radioligand Competition Binding Affinity (K _i) Data of PAM Derivatives (I)	158
Table 5-2. Radioligand Competition Binding Affinity (K _i) Data of PAM Derivatives (II).....	159
Table 5-3. Radioligand Competition Binding Affinity (K _i) Data of PAM Derivatives (III)	160
Table 5-4. Radioligand Competition Binding Affinity (K _i) Data of PAM Derivatives (IV)	160
Table 5-5. Radioligand Competition Binding Affinity (K _i) Data of PAM Derivatives (V).....	161
Table 5-6. Experimental (expt) and predicted (pred) pK _i values of PAM derivatives in the training set and test set.....	162

LIST OF FIGURES

Figure 1-1. A general scheme of a QSAR model development which includes systematic training and testing processes.	4
Figure 1-2. Hologram-QSAR (HQSAR) model development, which includes molecular hologram generation and partial least square analysis to derive a final predictive HQSAR equation.....	8
Figure 1-3. A general CoMFA workflow.	13
Figure 1-4. Categorization of structure-based virtual screening methods and sample programs.	28
Figure 1-5. Hierarchical descriptions of ligand-based design and some sample programs.....	28
Figure 1-6. Representative structures of major classes of cannabinoid ligands	38
Figure 2-1. A plot of Tanimoto (Tc) vs. BCUT similarity scores of random fragment pairs from bis-sulfone analogs.....	71
Figure 3-1. (Left) Four representative GALAHAD-generated hypermolecular superimpositions and correspondent pharmacophore models with four triaryl bis-sulfones 2, 5, 8, and 11. (Right) Four representative Surfex-Sim generated hypermolecular superimposition with four triaryl bis-sulfones 2, 5, 8, and 11.	102
Figure 3-2. (Left) The modified pharmacophore query defined according to the SAR studies. (Right) Combination of the MOLCAD-generated molecular surface based on molecular shape	

hypothesis and the modified pharmacophore query based on GALAHAD-generated pharmacophore model.....	103
Figure 3-3. Plot of the corresponding QSAR calculated and experimental values of binding affinity of triaryl bis-sulfones in the training and testing sets at the CB2 receptor.	104
Figure 4-1. Representative compounds from five QSAR data sets: (A) enalaprat (ACE); (B) E2020 (AchE); (C) Ro14-5974 (BZR); (D) celecoxib (COX2); (E) methotrexate (DHFR).	108
Figure 4-2. Representative CB ₂ compounds from CBID data set, reflecting the structural diversity of the data set.	109
Figure 4-3. The architecture of fingerprint-based ANN-QSAR (FANN-QSAR) model.....	112
Figure 4-4. Cross-validation results of each FANN-QSAR method on CB2 ligand data set.	118
Figure 4-5. Probability density function (left) and cumulative distribution function (right) plots of residual values of 295 newly reported cannabinoid compounds.....	121
Figure 4-6. Scatter plot between experimental pKi and predicted pKi values of 278 test cannabinoid ligands after the removal of 17 outliers.....	122
Figure 4-7. Experimental and predicted pKi values of 278 test cannabinoid ligands	122
Figure 4-8. CB2 receptor binding affinity Ki values of four NCI hit compounds measured by [3H]CP-55940 radioligand competition binding assay using human CB2 receptors harvested from transfected CHO-CB2 cells.....	126
Figure 5-1. Representative CB ₂ selective compounds with various chemical scaffolds.	148
Figure 5-2. Structures of the lead compound 1 (or XIE95 or PAM) and the modified target compounds 9, 17, 18, 21 and 50.	149
Figure 5-3. Comparisons of LANCE signal of different CB ₂ receptor ligands in stably transfected CHO cells expressing human CB ₂ receptors in a concentration-dependent fashion.	150

Figure 5-4. A predicted binding pocket around important binding residues of the CB2 receptor	151
Figure 5-5. CP55940, SR144528, compound 1 and compound 17 (ball and stick models) docked to the CB2 binding pocket displaying hydrogen-bonding and hydrophobic interactions with the receptor	152
Figure 5-6. Overall alignments of training set molecules (A) and test set molecules (B) to the compound 17 as well as CoMFA contour maps of compound 17 showing steric and electrostatic (C) interactions.....	153
Figure 5-7. Scatter plots of CoMFA-calculated and experimental binding affinity values (pKi) for the training and test sets.	154
Figure 5-8 Inhibition of human osteoclastogenesis by CB2 ligands.	155
Figure 5-9 Cytotoxic effects of PAM compounds on normal human mononuclear cells.	156

PREFACE

I would like to gratefully thank my advisor Dr. Xiang-Qun Xie for his support and guidance throughout my studies. His motivation and enthusiasm in studying several different biological problems have been inspirational to me. I am particularly grateful for the freedom and opportunity he granted me to pursue and explore many computational biology problems. Under his watch, I have developed further confidence in myself and matured into a well-rounded independent scientist.

I would also like to sincerely thank the members of my thesis committee, Dr. Ivet Bahar, Dr. Billy Day and Dr. Christopher Langmead for their insight, comments, and expert advice that helped me shape my dissertation.

I am very grateful and honored to receive a predoctoral training grant from the NIH as part of the HHMI-NIBIB Interfaces Initiative under the Joint CMU-Pitt computational biology PhD program for the last two years of my studies. It provided me with the financial support and motivation to excel in my research field.

I also owe a great deal of thanks to my colleagues in the lab, both past and present who helped in various ways. I am grateful to Dr. Haiping Cao, Dr. Peng Yang and Abdulrahman Almehezia for our several collaborative projects. Their expertise in medicinal chemistry and efforts in synthesizing several lead compounds have been instrumental to achieve our research goals. I am thankful to Drs. Yuxun Zhang, Qin Tong and Rentian Feng for their help in

experimental validation assays throughout my studies. I am grateful to Dr. Jianzhong Chen and Dr. Lirong Wang as well as Chao Ma for their computational knowledge and support for several of my research projects.

Finally, I would like to thank my father, mother and brother for their love and support throughout my studies. Although they are thousands of miles away from me, they have always made me feel as if they are with me supporting me. I am forever grateful to my beloved partner Yee Yee Myint for being a part of my life and for her unconditional love and support.

1.0 INTRODUCTION

1.1 RECENT ADVANCES IN QSAR METHODS

1.1.1 Introduction

Quantitative structure-activity relationship (QSAR) analysis is based on the general principle of medicinal chemistry that the biological activity of a ligand or compound is related to its molecular structure or properties, and structurally similar molecules may have similar biological activities [1]. Such molecular structural information is encoded in molecular descriptors and a QSAR model defines mathematical relationships between descriptors and biological activities of known ligands to predict unknown ligands' activities. QSAR methods have been applied in several scientific studies including chemistry, biology, toxicology and drug discovery to predict and classify biological activities of virtual or newly-synthesized compounds [2-5]. QSAR models can also be used in designing new chemical entities (NCEs) and are now regarded as essential tools in pharmaceutical industries to identify promising hits and generate high quality leads in the early stages of drug discovery [4, 6]. In other words, QSAR studies can reduce the costly failures of drug candidates by identifying the most promising hit compounds and reducing the number of costly experiments.

In general, QSAR modeling (Figure 1-1) involves a systematic process with multiple steps. These include dataset preparation, molecular descriptors selection and generation, mathematical or statistical models derivation, model training and validation using a training dataset and model testing on a testing dataset.

During the first step, or dataset preparation, it is important to pay attention to the quality of data to develop a reliable QSAR model. Data should come from the same bioassay protocols and it is preferable to collect and use the data generated from a single lab or source in order to avoid data inconsistencies and interlaboratory variability. Moreover, the dataset should have a large enough number of compounds to ensure statistical stability of a QSAR model and the bioactivity should cover a range of values with a good distribution [4].

The second step in QSAR modeling is the selection and generation of molecular descriptors for ligands in the dataset. There are many descriptors available and only some of them are significantly correlated with the activity. Therefore, selection of appropriate descriptors, which best capture the structural variation and information is important to derive a robust QSAR model. Several methods such as evolutionary algorithms (for example, genetic algorithm) and machine learning techniques (for example, forward selection) can be used for descriptor/variable/feature selection.

After molecular descriptors are defined and generated for all ligands in the dataset, the next step is to decide a suitable statistical or mathematical model to find the relationship between such descriptors and biological activities. For instance, linear approaches such as multiple linear regression (MLR) or partial least square (PLS) and non-linear methods such as neural networks or support vector machine can be used as correlation or mapping functions.

Once a model is chosen, it is then trained on a training dataset which contains a subset of randomly selected compounds from a known dataset, leaving the remaining to be used as testing compounds. During the model training, validation methods such as leave-one-out cross-validation (LOOCV) are often performed to ensure the statistical stability of the QSAR model. The training process is repeated until a satisfactory training performance is achieved. Finally, a testing process is performed in which the trained model is used to predict activity values of those compounds in the testing set.

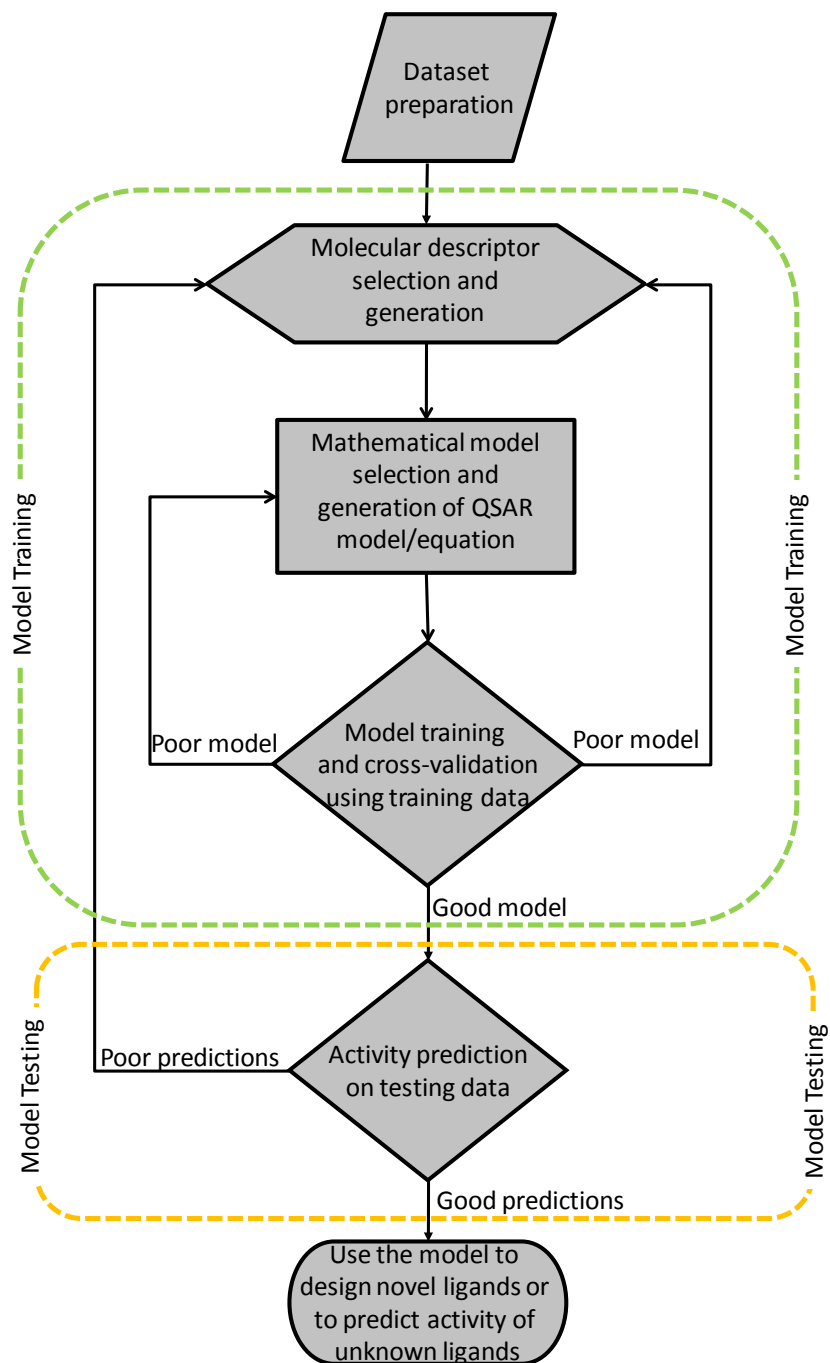


Figure 1-1. A general scheme of a QSAR model development which includes systematic training and testing processes.

A wide range of QSAR methodologies have been invented since the concept was first introduced by Free, Wilson, Hansch, and Fujita [7, 8] in 1964. Traditional 2D-QSAR methods

such as the Free-Wilson and Hansch-Fujita approaches use 2D molecular substituents or fragments and their physicochemical properties to perform quantitative predictions. Since then, QSAR has experienced a fast development and the first novel 3D-QSAR method called comparative molecular field analysis (CoMFA) was introduced by Cramer *et al.* in 1988. The CoMFA method brought a foundation for the development of other 3D-QSAR methods such as CoMSIA, SOMFA, CoMMA as well as multidimensional (nD)-QSAR methods such as 4D-QSAR, 5D-QSAR, *etc.*, to tackle known 3D-QSAR problems such as subjective molecular alignment and bioactive conformation problems. In recent years, fragment-based methods have attracted some attention because predicting molecular properties and activities based on molecular fragments is simple, fast and robust. In this chapter, we present recently available fragment-based QSAR methods and multidimensional (nD)-QSAR methods developed over the past few decades.

1.1.2 Fragment-based 2D-QSAR methods

Over the years, improved methods—that are based on such traditional QSAR methods—have been introduced. 2D methods allow modeling of a wide variety of ligands or compounds including cases where 3D crystal receptor or target structures are not available [6].

Hologram-QSAR (HQSAR)

One earlier example of a fragment-based method is HQSAR (Hologram QSAR) from Tripos [9, 10]. Given a method based on 2D molecular fragments, HQSAR does not require molecular alignment and therefore allows for automated analyses of large data sets without manual intervention. The first step in the HQSAR method is to generate molecular holograms

which contain counts of molecular fragments and can be related to 2D fingerprints. As depicted in Figure 1-2, the input dataset contains 2D structures of compounds and they are split into all possible linear and branched fragments. Then each unique fragment is assigned to a specific large positive integer by using a cyclic redundancy check (CRC) algorithm. All fragments generated are then hashed into array (hologram) bins in the range from 1 to L (total length of hologram). Bin occupancies represent counts of fragments in each bin. In other words, they are structural descriptors, which contain topological and compositional molecular information. During the second step, such fragment counts or hologram bins are correlated to corresponding biological activities (dependent variables) in a form of mathematical equation. Leave-one-out cross-validation (LOOCV) is performed to identify an optimal number of explanatory variables or components which yields an optimal model. Then by using standard partial least square (PLS) analysis, a following mathematical regression equation is derived to correlate hologram bin values or components with corresponding biological activities:

$$BA_i = const + \sum_{j=1}^L x_{ij} C_j \quad (1.1)$$

where BA_i is the biological activity of the i^{th} compound such as K_i or IC_{50} , x_{ij} is the occupancy value of the molecular hologram of the i^{th} compound at position or bin j , C_j is the coefficient for the bin j derived from the PLS analysis, and L is the length of the hologram.

One drawback of HQSAR is a phenomenon called a fragment collision problem which happens during the hashing process of fragments. Although hashing reduces the length of the hologram, it causes bins to have different fragments in the same bin. The hologram length, a user-definable parameter, controls the number of bins in the hologram and alteration of hologram

length can cause the pattern of bin occupancies to change. The program provides 12 default lengths which have been found to give good predictive models on different datasets. Each of these default lengths provides a unique set of fragment collisions [10].

Several HQSAR models for different ligand datasets including cases where 3D crystal structures of receptor targets or proteins are unavailable have been developed in recent years [11-14]. For example, HQSAR was used to study a set of 9-substituted-9-deazaguanine analogs which inhibit the human purine nucleoside phosphorylase (PNP) enzyme. HQSAR was used to identify structural features with poor and favorable contributions towards molecular interactions in the active site [11]. In addition, HQSAR has been used in virtual screening to identify hits [15-17]. For instance, Salum *et al.* studied a set of 180 indole derivatives having potent anticancer activity. They developed several HQSAR models and compared them to determine optimal cutoff values in virtual screening procedures [6].

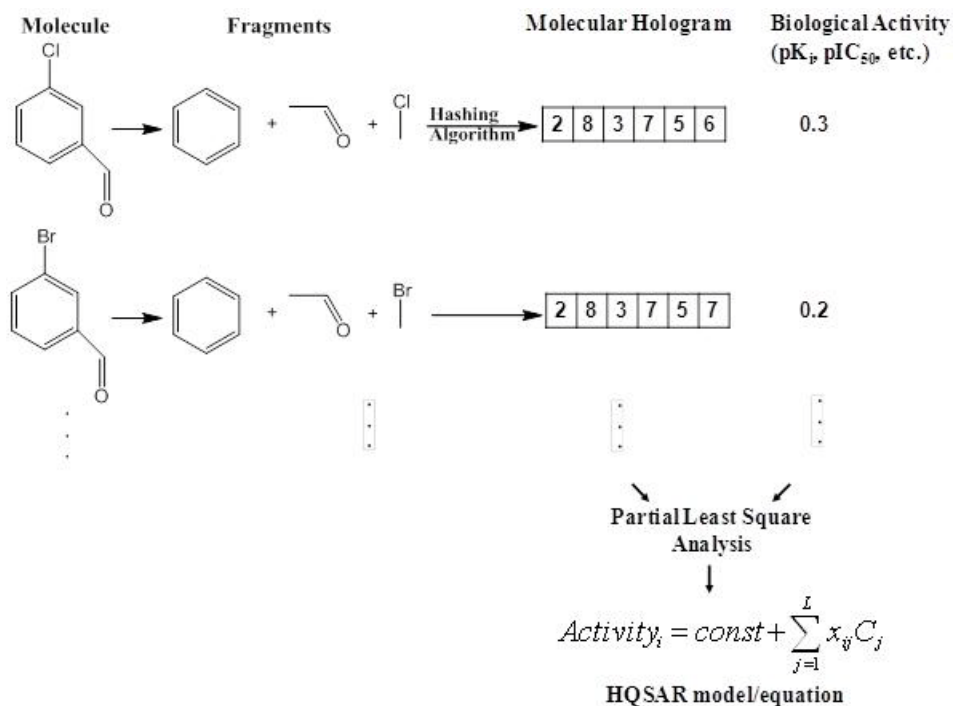


Figure 1-2. Hologram-QSAR (HQSAR) model development, which includes molecular hologram generation and partial least square analysis to derive a final predictive HQSAR equation.

Fragment-Based QSAR (FB-QSAR)

Recently, Du *et al.* [18] introduced a 2D-QSAR method based on molecular fragments. The method uses a mixed Hansch-Fujita [8] linear free energy equation and Free-Wilson [7] equation. In particular, molecular fragments are first generated from ligands and the total binding free energy ΔG_i^o between ligand i and the receptor is considered as the sum of contributions $\Delta g_{i,\alpha}$ from all fragments:

$$\Delta G_i^o = \sum_{\alpha=1}^M b_{\alpha} \Delta g_{i,\alpha} \quad (1.2)$$

where $\Delta g_{i,\alpha}$ is the free energy contribution of fragment $F_{i,\alpha}$ and b_{α} is a weight coefficient for each fragment. The binding free energy of a fragment $\Delta g_{i,\alpha}$, is described by a set of physical and chemical properties of the fragment:

$$\Delta g_{i,\alpha} = \sum_{l=1}^L a_l p_{i,\alpha,l} \quad (1.3)$$

where $p_{i,\alpha,l}$ is the l -th property of fragment $F_{i,\alpha}$ in molecule m_i and a_l is the coefficient of l -th property of the fragment.

In their studies, a total of 48 neuraminidase (NA) inhibitor analogs were used to train and test the model. Ten physicochemical properties were calculated for each substituent. Using an iterative double least square (IDLS) procedure, two sets of coefficients, one for fragments (b_α from Equation 1.2) and another for physicochemical properties (a_l from Equation 1.3), in the linear equation were solved alternately and iteratively until the model met the convergence criterion. After 176 iterations, the model converged and both sets of coefficients were solved. Such converged coefficients were used for the test calculation and the correlation coefficient (r) was 0.9525 (or $r^2 = 0.91$). They also tested on Free-Wilson and Hansch-Fujita models, which achieved r values of 0.2488 ($r^2 = 0.06$) and 0.9373 ($r^2 = 0.88$), respectively. The quantitative results proved the IDLS procedure enhanced the predictive power, and, given a novel method, more applications are necessary to fully explore its predictive potential.

Fragment-Similarity Based QSAR (FS-QSAR)

More recently, Myint *et al.* developed a fragment-similarity based QSAR (FS-QSAR) method to solve the major limitation of the original Free-Wilson method by introducing the fragment-similarity concept in the linear regression equation [19]. Such a similarity concept was applied for the first time to improve the traditional Free-Wilson equation instead of using physicochemical properties which often produce non-unique solutions. In this approach, the fragment similarity calculation was carried out by the similarity. It used the lowest or highest eigen values calculated from BCUT-matrices [20, 21], which contained partial charges of individual atoms and their atomic connection information in each individual fragments. In Section 2.0, the method is described in details.

Top Priority Fragment QSAR

Casalegno *et al.* [22] introduced a fragment-based QSAR approach to predict pesticide aquatic toxicity to the rainbow trout. The method prioritizes fragments' contributions to toxicity with the assumption that one fragment among others present in a compound is mainly responsible for the toxicity. They used 282 carefully selected pesticides which were partitioned into 240 training and 42 testing molecules. In the first stage, all 282 molecules were broken into small substructures or atomic centered units (ACUs). Then, a numerical criterion based on the training set toxicity data was applied to assign one fragment or top-priority fragment (TPF), made up of one or more ACUs, to each training molecule. Once the TPFs were extracted, a 'priority matrix' was used to extract all priority relationships. A priority matrix contains information among training TPFs and can be used to find out which TPF has a priority to be assigned to a testing molecule. In the last stage, testing molecules were submitted to check for the presence of TPFs and information from the priority matrix was used to identify the ones(s) with highest priority, and final prediction was made based on average fragment toxicity. The final r^2 for the training set was 0.85 and 0.75 for the test set proving the model's effectiveness.

Other Fragment-Related QSAR Studies

In recent years, some new fragment-based QSAR methods have been discovered as well as applications to biological interests. Zhokhova *et al.* [23] introduced a method which uses fragmental descriptors with labeled atoms and applied it to their QSAR/QSPR (quantitative structure-property relationship) studies. In their approach, the fast stepwise multiple linear regression (FSMLR) and three-layer artificial neural network (ANN) methods implemented in

the NASAWIN program [24] were used to generate fragmental descriptors with labeled atoms and to construct QSAR/QSPR models. Andrade *et al.* [25] used HQSAR and other 2D-QSAR programs to study a series of hydrazides as antituberculosis agents. They used DRAGON 5.4 [26], BuildQSAR [27], PIROUETTE [28] programs for generation and selection of 2D molecular descriptors. Tsygankova *et al.* [29] also did the QSAR studies of barbituric acid derivatives using 2D fragments as descriptors with different regression approaches such as step-by-step regression to construct correlation equations.

1.1.3 3D-QSAR

3D-QSAR methods have been developed to improve the prediction accuracies of 2D methods. 3D methods are computationally more complex and demanding than 2D approaches. In general, there are two families of 3D-QSAR methods: alignment-dependent methods and alignment-independent methods. Both families need experimentally or computationally derived bioactive conformations of ligands as templates for studies. Such 3D conformers are one of the most important factors to produce reliable 3D-QSAR models and are also the major drawbacks of 3D methods. Examples of both families are discussed below.

Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA)

One of well-known methods is a three dimensional QSAR method called CoMFA developed by Cramer *et al.* [30]. It is a method to describe 3D structure-activity relationship quantitatively by considering 3D structures, and steric and electrostatic fields of ligands which are superimposed to generate such molecular fields. In other words, CoMFA is an alignment-

dependent method in which molecular field interaction energy terms are correlated with biological activities/responses using multivariate statistical analyses. Figure 1-3 illustrates a general CoMFA modeling process where active molecules are first placed in a 3D grid. Using a probe atom, steric and electrostatic energies are measured at each grid point for each molecule. Partial least square (PLS) analysis is then performed to correlate such field energy terms to activity values and make predictions. Such features and calculations make CoMFA an improved and different method from other traditional QSAR approaches.

Another 3D QSAR method named CoMSIA by Klebe *et al.* is similar to CoMFA in terms of using a probe atom along grid points. However, additional molecular fields have been implemented in the CoMSIA approach. In particular, electrostatic, steric, hydrophobic, hydrogen bond acceptor (HBA), and hydrogen bond donor (HBD) properties are generated using a Gaussian distance function [31]. Using such a Gaussian-type potential function instead of Lennard-Jones and Coulombic functions provides accurate information at grid points for calculating molecular fields [32].

However, the major drawback of both methods is that all molecules have to be aligned and such alignment can affect the final CoMFA/CoMSIA model and predictions. A good alignment is necessary and quality of such alignment can be subjective, time-consuming [33] and CoMFA/CoMSIA models are sometimes non-reproducible [32]. Nevertheless, several CoMFA/CoMSIA models have been developed for many drug design and molecular modeling studies [5, 34-38].

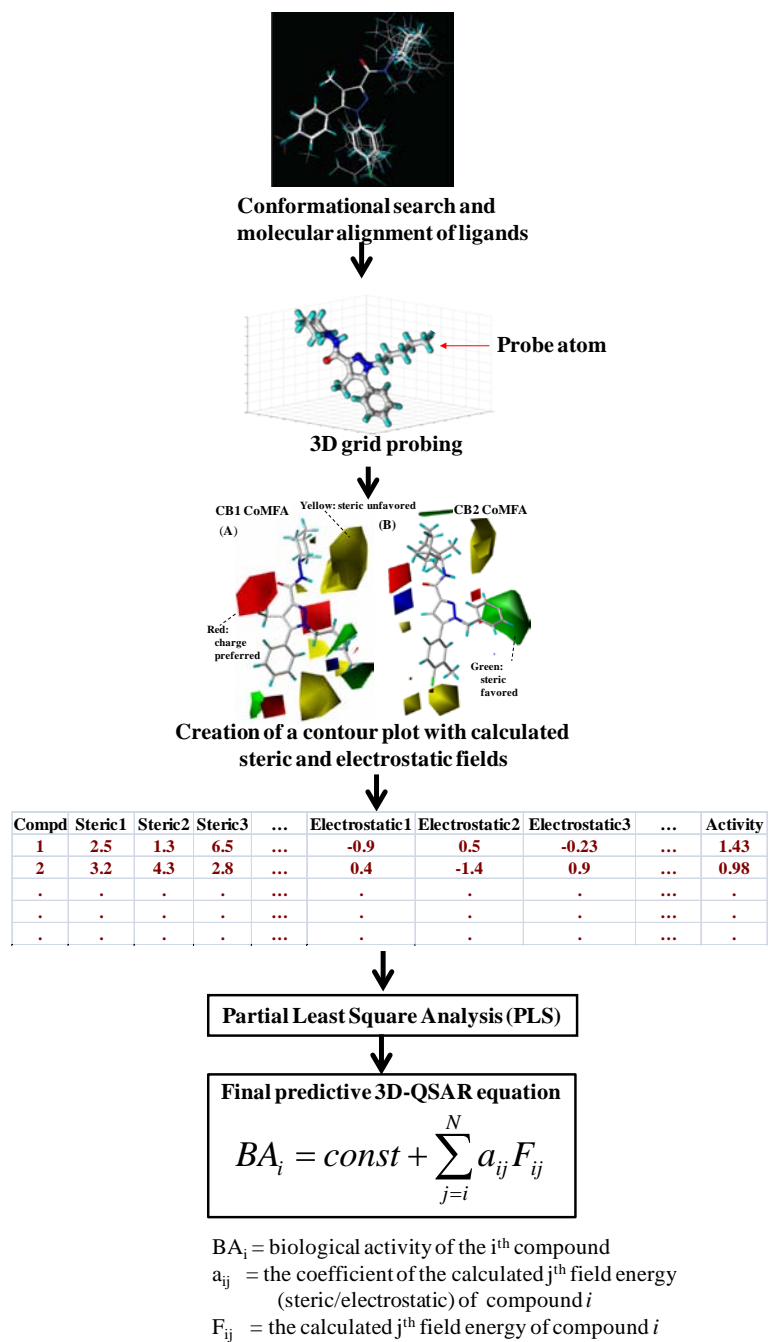


Figure 1-3. A general CoMFA workflow.

Topomer CoMFA

Recently, Cramer *et al.* introduced a new QSAR method named the Topomer CoMFA [39] which is a rapid fragment-based 3D-QSAR method to predict significant R-groups, which can optimize the biological activities as well as optimized structural changes for lead scaffold hopping. It uses the compound library collection as a source of molecular fragments to identify such substituents or R-groups. The Topomer CoMFA method, unlike CoMFA, does not require the subjective alignment of 3D ligand conformers and uses automated alignment rules. A topomer describes both a conformation and orientation of a molecular fragment and it is generated based on 2D structure without any relation to a receptor site or other ligands [33, 39]. After such topomers are generated, CoMFA analysis is then carried out where electrostatic and steric fields are calculated using a probe atom around the 3D grid. Subsequently, partial least square (PLS) with leave-one-out cross-validation is performed to generate a predictive model. 15 3D-QSAR analyses retrieved from the literature yielded an average q^2 of 0.520 compared to literature average q^2 of 0.636 [39]. Topomer CoMFA has the potential to optimize biological activities of ligands via fragments and has been used in lead-optimization and R-groups virtual screening studies [33, 39].

Self-Organizing Molecular Field Analysis (SOMFA)

Robinson *et al.* [40] introduced another alignment-dependent 3D-QSAR method called SOMFA, which is based on both molecular shape and electrostatic potentials. Briefly, 3D grids are created as in other 3D-QSAR methods and for each grid point, molecular shape and electrostatic potential values are calculated. Shape values are binary meaning 1 for being inside the van der Waals envelope and 0 outside. The key step is that the electrostatic potential value at each grid point is multiplied by the mean centered activity for that molecule as a weighing factor which causes the most active and least active molecules to have higher values than other

common and less interesting molecules which are closer to the mean activity. The SOMFA grid value at a given x,y,z is defined as:

$$SOMFA_{x,y,z} = \sum_i^{Training_Set} Pr\,operty_i(x, y, z) Mean_Centered_Activity \quad (1.4)$$

Using such a property master grid, an estimate of the activity of the i^{th} molecule as defined by a certain property can be derived as:

$$SOMFA_{property,i} = \sum_x \sum_y \sum_z Pr\,operty_i(x, y, z) SOMFA_{x,y,z} \quad (1.5)$$

In the final stage, correlations between calculated SOMFA property values ($SOMFA_{property,i}$) and biological activities are derived via multiple linear regression and a final predictive model is produced. Robinson *et al.* tested the model using two datasets : 31 steroid compounds and 35 sulfonamides. The corresponding correlation coefficient values (r^2) of 0.5776 ($r = 0.76$) and 0.5329 ($r = 0.73$) were achieved, respectively. Compared to other methods such as CoMFA [30], MS-WHIM [41] and few others on steroid dataset, SOMFA had the lowest standard deviation of errors of prediction (SDEP), which is the root-mean-square error of the predictions. In short, SOMFA is similar to CoMFA in terms of using grids and necessity of molecular alignment but is not as statistically rigorous as CoMFA [1], as the SOMFA model is conceptually simple without heavy statistical elements such as partial least square (PLS).

Alignment-Free 3D-QSAR Methods

In the last few decades, other 3D-QSAR methods which do not rely on alignments were introduced. Some examples include autocorrelation of molecular surfaces properties (AMSP)

[42], comparative molecular moment analysis (CoMMA) [43], WHIM (Weighted Holistic Invariant Molecular) method [44, 45], Molecular surface (MS)-WHIM [41], and GRIND [46].

Autocorrelation of Molecular Surfaces Properties (AMSP)

Wagener *et al.* introduced the AMSP method to map the physical properties of ligands to a van der Waals surface and individual atoms, respectively. It uses a 3D descriptor based on spatial autocorrelation of molecular properties at distinct points on the molecular surface. The points are randomly distributed to have a continuous surface and the autocorrelation coefficient is obtained by summing the products of property values at various pairs of points at particular distances. For a series of distance intervals (d_{lower} , d_{upper}), a vector of autocorrelation coefficients is obtained as follows:

$$A(d_{lower}, d_{upper}) = \frac{1}{L} \sum_{ij} p_i p_j \quad (d_{lower} < d_{ij} < d_{upper}) \quad (1.6)$$

where p_i is the molecular property value at point i , p_j is the molecular property value at point j and L is the total number of distances in the interval [42].

Therefore, the vector contains a compressed expression of the distribution of a property on the molecular surface. After autocorrelation vectors were obtained, a multilayer neural network was then trained using such vectors to derive a predictive model of biological activity of 31 steroid compounds. The correlation coefficient value, r , of 0.82 ($r^2 = 0.6724$) was achieved with a cross-validated r^2 of 0.63. In summary, the advantages of such autocorrelation vectors are the facts that they are shown to be invariant to translation and rotation since only spatial distances are used and have condensed description of molecular surface. However, original information cannot be reconstructed from such condensed vectors and the pharmacophore nature of a ligand may not be clear or interpretable [42].

Comparative Molecular Moment Analysis (CoMMA)

Silverman *et al.* [43] introduced the CoMMA method, which calculates the zeroth-, first-, and second-order spatial moments of the charge (such as quadrupolar moments) and the mass distribution (such as moments of inertia). Such molecular moment descriptors may be classified in three different categories: descriptors relating solely to molecular shape, descriptors relating only to molecular charge and descriptors relating to both shape and charge. The authors calculated 13 such descriptors and used them in partial least square analysis to generate predictive QSAR models for 31 steroid compounds. A range of statistical performance was obtained depending on different partial charge calculation methods used to derive electrostatic moments. Cross-validated r^2 values ranging from 0.412 to 0.828 were obtained using electrostatic moment descriptors calculated from Gasteiger charges or Gaussian molecular orbital *ab initio* methods. The results showed that using quantum chemistry calculation-based moments produced better predictive models than using only Gasteiger charge-based moments. Despite CoMMA's comparable statistical performances to CoMFA's, there are some limitations which may account for the limited number of published CoMMA applications. One reason is that the value of these descriptors, which measures the displacement between the center of mass and center of dipole with respect to the principal inertial axes, equals infinity for symmetric molecules whose dipole moment is zero [4].

Weighted Holistic Invariant Molecular (WHIM) Descriptor-Based QSAR

WHIM descriptors contain 3D molecular information such as molecular size, shape, symmetry and distribution of molecular surface point coordinates [44, 45]. Molecular surface

(MS)-WHIM is a WHIM-based 3D descriptor derived directly from molecular surface properties [41]. For WHIM descriptors, two types of matrices are defined: a molecular matrix containing cartesian coordinates of the n atoms and diagonal matrices containing the weights which are physicochemical properties associated with the n atoms of the molecule [41]. Each element of the diagonal matrix is defined as:

$$S_{jk} = \frac{\sum_{i=1}^n w_i (q_{ij} - \bar{q}_j)(q_{ik} - \bar{q}_k)}{\sum_{i=1}^n w_i} \quad (1.7)$$

where n is the number of atoms, w_i is the weight of i th atom, q_{ij} is the j^{th} coordinate of the i^{th} atom and \bar{q}_j is the average of the j^{th} coordinates [44].

In this expression, atoms can be weighted by mass, van der Waals volume, atomic electronegativity, electrotopological index of Kier and Hall, atomic polarizability and molecular electrostatic potential [32]. Elements in each diagonal matrix are subjected to principal component analysis (PCA) to obtain the scoring matrix, which is used to calculate PCA eigen values and eigen value proportion. Such values and proportions are then correlated with the molecular size and shape, respectively. One major advantage of the WHIM approach is that it provides a 3D QSAR descriptor which is invariant to translation and rotation of 3D molecular structures. In MS-WHIM, properties associated with the molecular surface points are used as different weighting schemes to compute statistical parameters. In particular, the unitary value and molecular electrostatic potential (MEP) are computed at each point of the Connolly molecular surface [47], and they are considered as weights. The unitary value contains information about the molecular surface shape and MEP provides the electrostatic information

about the electron density distribution [41]. Although the WHIM approach is not sensitive to molecular orientation, MS-WHIM descriptor values are affected by the facts that the Connolly surface points are dependent on the 3D orientation of the molecule and indices for different weighting schemes are sensitive to surface point density [41]. The authors tested both WHIM and MS-WHIM on 31 steroid compounds and achieved the SDEP (standard deviation error of prediction) values of 1.750 and 0.742, respectively while CoMFA's SDEP was 0.837. The results suggested that MS-WHIM prediction performance was comparable to CoMFA's. SDEP was defined as follows:

$$\text{SDEP} = \sqrt{\frac{\sum (y_{\text{pred}} - y_{\text{obs}})^2}{n}} \quad (1.8)$$

WHIM/MS-WHIM descriptors are invariant to 3D molecular orientation but both methods, like other 3D-QSAR methods, rely on ligand conformation, which may be subjective if ligand-receptor co-crystal structures are not known for the target of interest.

Grid-Independent Descriptors (GRIND)-Based QSAR

In an attempt to provide alignment-free descriptors which are easy to understand and interpret, Pastor *et al.* introduced grid-independent descriptors [46]. The method utilizes specific probes such as the O probe (carbonyl oxygen) and N1 probe (amide nitrogen) to calculate molecular interaction fields (MIFs) at grid points. At each node of the grid, the energy between the probe and target ligand (E) is calculated as:

$$E = \sum E_{es} + \sum E_{hb} + \sum E_{lj} \quad (1.9)$$

where E_{es} is the electrostatic energy, E_{hb} is the hydrogen-bonding energy, and E_{lj} is the Lennard-Jones potential energy [48].

In this method, electrostatic interactions, hydrophobic interactions, hydrogen bond acceptor and hydrogen bond donor fields are considered to get a set of positions which defines a ‘virtual receptor site’ (VRS). VRS regions are then encoded into GRIND via an auto- and cross-correlation transform so that those regions are no longer dependent upon their positions in the 3D space. In other words, autocorrelation descriptors of the fields are calculated and only the highest products of molecular interaction energies are stored while others are discarded. This difference is responsible for the ‘reversibility’ of GRIND and the descriptors can be back-projected in 3D space using another related program called ALMOND [49]. The statistical performance of GRIND is comparable to other methods, but the advantage is that it is alignment-free and easy to interpret. However, bioactive conformations of ligands are valuable information to derive the virtual receptor site (VRS) and limitations on such information may affect final predictive models like other 3D methods.

Multi-Dimensional (nD) QSAR Methods

Multi-dimensional (nD) QSAR methods are essentially extensions of 3D-QSAR methods. These methods incorporate additional physical characteristics or properties (or a new dimension) to tackle the drawbacks of 3D-QSAR methods. One example is 4D-QSAR by Hopfinger *et al.* [50] which samples molecular conformations and alignments during the generation of a QSAR model. While incorporating some CoMFA features, it introduces the fourth dimension, which is the conformational Boltzmann sampling, and enables the method to be used as a receptor-independent (RI) method as well as receptor-dependent (RD) method in which the geometry of

the receptor is known. It should be noted that their 4D-QSAR method does not solve the alignment problem but it allows a rapid evaluation of individual trial alignments [50]. Such 4D-QSAR implementation can be found in XMAP program [50, 51]. Recently, it has been shown that 5D- and 6D-QSAR can be used for multiple representations of the receptor as well as its solvation states [52-54]. In the reported 5D-QSAR method, Vedani *et al.* introduced a multiple representation of induced-fit hypotheses, *i.e.*, the adaptation of the receptor binding pocket to the individual ligand topology, as the fifth dimension. In other words, they generated a family of quasi-atomistic receptor surrogates [55] which are optimized by using a genetic algorithm. The binding energy was calculated as:

$$E_{binding} \approx E_{ligand-receptor} - E_{solvation,ligand} - T\Delta S - E_{internal\ strain} - E_{induced\ fit} \quad (1.10)$$

where $E_{ligand-receptor}$ is the force field energy of the ligand-receptor interaction, $E_{solvation,ligand}$ is the ligand desolvation energy, $T\Delta S$ is the change in the ligand entropy upon receptor binding, $E_{internal\ strain}$ is the change in ligand internal energy upon receptor binding, and $E_{induced\ fit}$ is the energy uptake required for adapting the receptor surrogate [53].

The 5D-QSAR method was tested on a set of 65 *NK-1* receptor antagonists and a set of 131 *Ah* receptor ligands, achieving predictive r^2 values of 0.837 and 0.832, while 4D-QSAR model resulted in 0.834 and 0.795, respectively [53]. They concluded that the binding affinities of new molecules were predicted more accurately with 5D-QSAR than with other lower dimension models. In the reported 6D-QSAR model, the simultaneous consideration of different solvation models was introduced by mapping parts of the surface area with different solvent properties [54]. 3D, 4D, 5D and 6D models were explored as comparison studies and the results

showed the 6D-QSAR model produced the best predictive r^2 of 0.885 [54]. Both 5D- and 6D-QSAR methods are implemented in the *Quasar* and *VirtualToxLab* software [55, 56].

1.1.4 Comparison of 2D or fragment-based QSAR versus 3D or nD-QSAR methods

In general, the predictive quality of 3D-QSAR methods depends on several factors such as the quality of molecular alignments/superimpositions, and information on ligand bioactive conformations. Especially molecular superimpositions are subjective and ligand bioactive conformations always remain unclear when there is no structural information on the corresponding receptor-ligand complexes. Conventional CoMFA results may often be non-reproducible because the model depends on the orientation of alignment of molecules, which can be varied and subjective. Although various improved methods and other procedures, which were discussed earlier in the paper, have been introduced to overcome major limitations of 3D-QSAR methods, *i.e.*, subjective molecular alignment and bioactive conformation problems, many of them still require manual interventions and superimpositions [57, 58]. From this prospect, 2D fragment-based QSAR methods have certain advantages over multi-dimensional QSAR methods since fragment-based or 2D-QSAR methods are simple and robust and do not require subjective (or time consuming) molecular alignment or putative binding conformation or determination of 3D structures. However, the disadvantage is that some of 2D-QSAR methods such as Hansch-Fujita method may provide non-unique solutions and the overall predictive quality may not be as good as some multi-dimensional methods which are computationally more complex and demanding. A summary of QSAR methods discussed in the paper is listed in Table 1-1. It should be noted that the performance of each QSAR model depends on the choice of dataset and different datasets can result in different predictive q^2 or r^2 or SDEP value.

Table 1-1. Summary of different QSAR methods and source information.

Method	nD	Dataset	Statistical model	Performance	Reference/Website
HQSAR	2D	21 Steroids	PLS	$q^2 = 0.71$; $r^2 = 0.85$ [10]	[10] http://www.tripos.com
FB-QSAR	2D	48 NA analogs	IDLS	$r = 0.95$ ($r^2 = 0.91$) [18]	[18]
FS-QSAR	2D	85 bis-sulfone analogs; 83 COX2 analogs	MLR	$r^2 = 0.68$; $r^2 = 0.62$ [19]	[19]
TPF-QSAR	2D	282 pesticides	PM-based prediction	$r^2 = 0.75$ [22]	[22]
CoMFA	3D	21 Steroids 54 HIV-1PR inhibitors	PLS	$q^2 = 0.75$; $r^2 = 0.96$ [10] $q^2 = 0.68$; $r^2 = 0.69$ [59]	[30] http://www.tripos.com [59]
CoMSIA	3D	Thermolysin inhibitors 54 HIV-1PR inhibitors	PLS	$q^2 = [0.59, 0.64]$ [31] $q^2 = 0.65$; $r^2 = 0.73$ [59]	[60, 61] http://www.tripos.com [59]
Topomer CoMFA	3D	15 datasets from literature	PLS	average $q^2 = 0.636$ [39]	[39] http://www.tripos.com
SOMFA	3D	31 steroids; 35 sulfonamides	MLR	$r^2 = 0.58$; $r^2 = 0.53$ [40]	[40]
AMSP	3D	31 steroids	MNN	$q^2 = 0.63$; $r^2 = 0.67$ [42]	[42]
CoMMA	3D	31 steroids	PLS	$q^2 = [0.41, 0.82]$ [43]	[43]
WHIM	3D	31 steroids	PCA	SDEP = 1.750 [41]	[44] http://www.vcclab.org/lab/indexhlp/whimdes.html
MS-WHIM	3D	31 steroids	PCA	SDEP = 0.742 [41]	[41]
GRIND	3D	31 steroids 175 hERG inhibitors	PLS; PCA PLS; SVM	$q^2 = 0.64$; SDEP = 0.26 [46] $q^2 = 0.41$; $r^2 = 0.57$; SDEP = 0.72 [62]	[46] http://www.moldiscovery.com/soft_grid.php [62]

Method	nD	Dataset	Statistical model	Performance	Reference/Website
4D-QSAR	4D	20 DHFR inhibitors; 42 PGF ₂ a analogs; 40 2-substituted dipyridodiazepione inhibitors 33 p38-MAPK inhibitors	PLS GL-PLS	$r^2 = [0.90, 0.95]$; $r^2 = [0.73, 0.86]$; $r^2 = [0.67, 0.76]$ [50] $q^2 = [0.67, 0.85]$ [63]	[50] http://www.seascapelearning.com/4DsgiSW/ [63]
5D-QSAR	5D	65 NK-1 antagonists; 131 Ah ligands	MLR	$r^2 = 0.84$; $r^2 = 0.83$ [53]	[53] http://www.biograf.ch
6D-QSAR	6D	106 estrogen receptor ligands	MLR	$q^2 = 0.90$; $r^2 = 0.89$ [54]	[54] http://www.biograf.ch
HQSAR = Hologram QSAR FB-QSAR = Fragment-based QSAR FS-QSAR = fragment-similarity-based QSAR TPF-QSAR = Top priority fragment QSAR CoMFA = Comparative molecular field analysis CoMSIA = Comparative molecular similarity indices analysis SOMFA = Self-organizing molecular field analysis AMSP = Autocorrelation of molecular surface properties CoMMA = Comparative molecular moment analysis WHIM = Weighted holistic invariant molecular QSAR MS-WHIM = Molecular surface WHIM GRIND = Grid independent descriptor				PLS = Partial least square IDLS = Iterative double least square PM = Priority matrix MNN = Multilayer neural networks MLR = Multiple linear regression PCA = Principal component analysis	$q^2 =$ cross-validated r^2 SDEP = standard deviation of errors of prediction

1.2 VIRTUAL SCREENING APPROACHES

1.2.1 Structure-based method

Detailed crystal structure coordinate data are used to design novel small molecules for structure-based virtual screening approaches. There are several structure-based methods and two basic subclasses of methods are: fragment position method and molecule growth method [64]. Two well-known programs for fragment-position methods are GRID [65] and MCSS (Multiple copy simultaneous search) [66] [67]. Generally, these methods find binding pockets or sites which are energetically favorable for various interacting fragments, position them appropriately and link these fragments together by other linker fragments [68]. For growth methods, a seed atom or fragment is placed in the binding site first. Then a ligand is grown by bonding atoms or fragments. There are several important issues that one should be wary of in *de novo* designs. For example, the algorithms to some extent should be able to reproduce known chemotypes for different targets without having prior knowledge of known solutions as well as novel structures [68]. Some programs that uses the growth methods are SMOG (small molecule growth) [69], GrowMol [70], GenStar [71] and GROW [72].

Two important components in structure-based virtual screening are molecular docking and scoring. Docking is the process which brings two molecular species together with a predicted conformation and orientation (or posing). Scoring is a method of accessing the binding affinity of such two species. In other words, docking and scoring complements each other. Historically, docking of small molecules to protein/receptor binding sites was started in the early

1980s [73] and many docking programs such as DOCK [74], GOLD [75], AUTODOCK [76] and Surflex-Dock [77] have been developed. Many docking programs assume the protein to be rigid because of the high computational cost required by the flexibility. However, most of current docking programs consider ligand and protein flexibility since flexible docking produce better results than rigid body docking [78]. Categorization of different docking methods and sample programs are listed in Figure 1-4 and Table 1-2. Given different docking methods, the success of docking molecules into a target site depends on the accuracy of scoring functions which rank the hit compounds based on how well they bind to the receptor site. There are four main classes of scoring functions: force field-based, knowledge-based, empirical and consensus scoring functions. All of them aim to approximate the binding free energy. Force field-based scoring functions use classical molecular mechanics energy calculations such as van der Waals and electrostatic interactions. However, force fields are computationally more costly than the other types of scoring functions. Empirical scoring functions estimate the binding free energy by a weighted sum of ligand-receptor interaction parameters. The weights are obtained by a regression analysis of experimental binding constants of protein-ligand complexes. Knowledge-based scoring functions use a sum of protein-ligand atom-pair interactions to calculate the binding affinity and frequencies of each possible pair of atoms in contact to each other are determined [78, 79]. Consensus scoring combines scores from different scoring functions to compensate errors in single scores.

1.2.2 Ligand-based method

Ligand-based design methods result from a lack of a 3D crystal structure of a target protein. Such lack of information leads to an alternative solution of using active compounds of the target

protein as the basis to further generate novel structures. Ligand-based methods make use of pharmacophoric, structural or topological features from known ligands/compounds and derive putative ligands as well as queries for database screening. Particularly, in a pharmacophore-based design, a three dimensional ligand pharmacophore model is derived first and then the model is used to screen compound databases or design structures which are complimentary to the primary target constraints. Another subclass of ligand-based screening is based on structural similarity of test molecules to a reference molecule(s). In this approach, a test molecule is superimposed or aligned to the reference molecule and similarity is measured. Such similarity value can then be used as a scoring factor to rank hits from database searches. However, such alignment techniques require manual intervention and a considerable amount of time and, therefore, descriptor based representations were introduced for faster database screening. There are 1D, 2D and 3D descriptors based on the number of properties used to represent molecules. 1D descriptors are molecular properties such as molecular weight and log P whereas 2D descriptors are classified as linear descriptor and tree-type descriptor. 3D descriptors consider conformational information of ligands in addition to other molecular properties. 2D Linear descriptor has two cases: binary descriptor and real value descriptor. Structural keys and molecular fingerprints belong to binary descriptors and Feature tree and MTree are considered as tree-type descriptors [80, 81]. Some ligand-based methods use molecular fragments to screen compound databases. These methods use different approaches to break molecules into smaller fragments and use such fragments to do virtual screening. One example of a fragment-based method is called Topomer Search from Tripos [82] which can perform R-group search, scaffold search or whole molecule search. Figure 1-5 illustrates a hierarchical description of ligand-based methods and Table 1-3 contains a list of selected ligand-based programs.

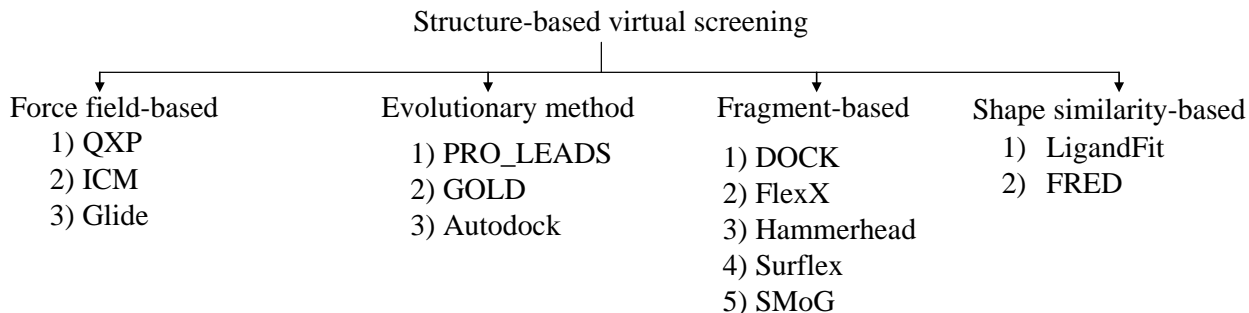


Figure 1-4. Categorization of structure-based virtual screening methods and sample programs

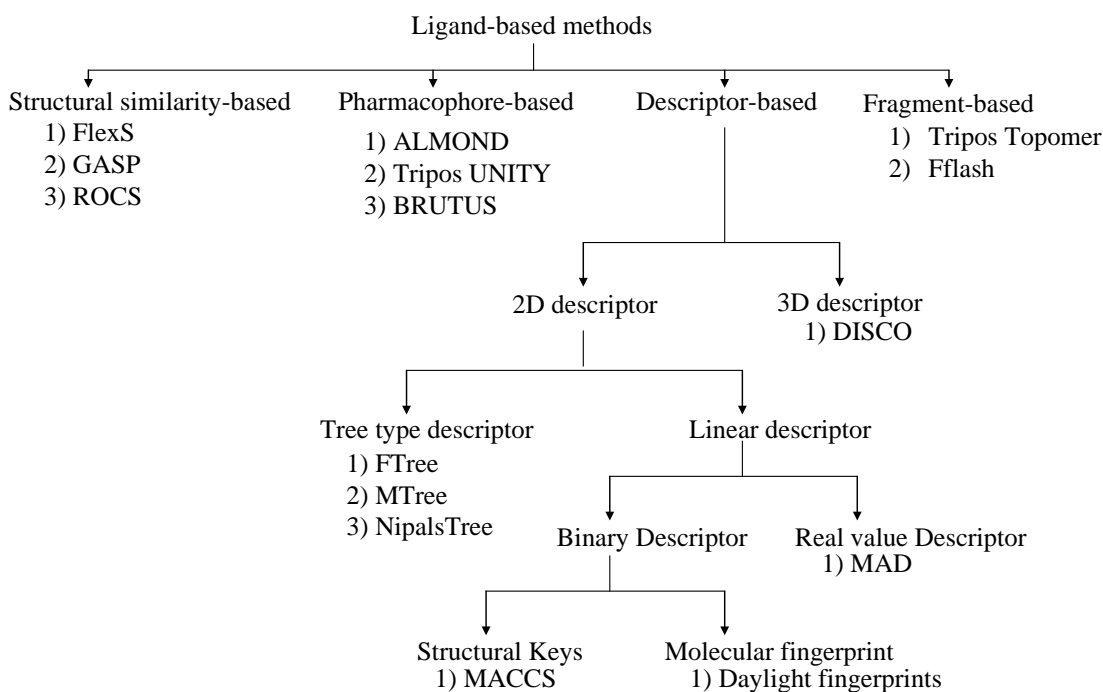


Figure 1-5. Hierarchical descriptions of ligand-based design and some sample programs (adapted from Reddy *et al.*)

Table 1-2. List of selected structure-based docking programs

Docking Method	Programs	Description	Reference
Force field-based	QXP	QXP (quick explore) uses a fast Monte Carlo (MC) search method to match proposed molecules to a template and dock flexibly to a binding site.	[83]
	ICM	ICM (Internal Coordinate Modeling) performs the optimization of ligand conformation during docking by changing internal coordinate system variables and Monte Carlo minimization procedure.	[84]
	Glide	Glide (Grid based ligand docking with energetics) uses a series of hierarchical filters to look for possible ligand locations in the binding region. After a systematic search of conformation, orientation and position of docked ligands, it uses a Monte Carlo sampling method to further refine the best candidates.	[85]
Evolutionary method	PRO_LEADS	It uses an evolutionary method called a tabu search algorithm which calculates the root mean square deviation between current molecular coordinate and previous conformations of other molecules in order to decide if a molecular conformation will be rejected or accepted.	[73]
	GOLD	GOLD (Genetic Optimization for Ligand Docking) docks a flexible ligand to a semi-flexible receptor by using a genetic algorithm (GA). It uses a force-field-based scoring function.	[75]
	Autodock	It uses not only an evolutionary/genetic algorithm but also the Monte Carlo simulated annealing methods to predict the binding interaction of small molecules to the protein.	[76]
Fragment-based	DOCK	It uses a clique-detection procedure to fit compounds into the binding site. Hydrophobicity descriptors are considered in scoring the docked ligand-protein complexes.	[74]
	FlexX	The program, after breaking the ligand at rotatable bonds into fragments, puts one fragment after another at the binding site and rebuilds the whole ligand. It uses a pose-clustering algorithm to classify the docked ligand conformers.	[81, 86, 87]
	Hammerhead	It is a fast and automated flexible docking program which can screen large databases. The program is fully automated: from the binding site search, small molecules docking and final compound selection for in vitro assays.	[88]
	SMoG	SMoG (Small Molecule Growth) uses a metropolis Monte Carlo growth algorithm and its scoring function is derived from a knowledge-based pair-wise atom potentials.	[69]

Table 1-3. List of selected ligand-based programs (adapted from Reddy *et al*)

Method	Programs	Description	Reference
Shape similarity-based	LigandFit	It uses a grid-based approach to examine shapes of the ligand and the target.	[78, 89]
	FRED	It applies Gaussian functions to fit the ligand shape to the negative image of the shape of the protein.	[78, 90]
Structural similarity-based	FlexS	It superimposes a test molecule which is kept flexible to a rigid reference molecule. It can be used to compare small and medium sized molecules.	[91]
	GASP	Unlike FlexS, both test and reference molecules can be kept flexible during alignments. But it is slower than FlexS because of such flexibility.	[92]
	ROCS	ROCS (Rapid Overlay of Chemical Structures) is a shape similarity-based algorithm that uses the Gaussian representation of molecular volume to screen for lead compounds.	[93, 94]
Pharmacophore-based	ALMOND	It can calculate the locations of pharmacophore features in 3D space and provide geometric correlation between such features. It also provides information related to pharmaco-kinetic properties of molecules.	[46, 81]
	BRUTUS	This is a fast grid-based algorithm which uses charge distribution to do rigid body molecular alignment.	[95]
	Tripods UNITY	It can define a pharmacophore model by using multiple pharmacophore feature points and use such model for 3D database screening.	[96]
Descriptor-based	DISCO	DISCO is a three dimensional pharmacophore descriptor program. It identifies 3D pharmacophore features from template compounds, generates interpoint distances among these features and uses the Bron-Kerbosch clique detection algorithm for interdistance comparisons.	[97, 98]
	Ftrees	It is a graph-based algorithm which condenses molecular descriptions into a graph object and search for active compounds from large molecular databases.	[99]
	NipalsTree	It is a hierarchical clustering algorithm which is based on the principle component analysis (PCA) using NIPALS method.	[100]

	MAD	MAD (Mapping to Activity-class-specific Descriptor value ranges) is a real-value descriptor based mapping method which searches for test compounds which has similar activity-class-specific descriptor values as reference compounds.	[101]
	MACCS	MACCS keys are the structural keys and it considers not only the presence or absence of each fragment but also the frequency of its occurrence.	[102]
	Daylight Fingerprint	It is a structural path-based fingerprint method which uses a hashing algorithm to generate distinct fingerprints.	[103]
Fragment-based	Tripos Topomer	It can perform various fragment-based database queries such as R-group search, scaffold search or whole molecule search based on topomer (specific pose or alignment of a molecular fragment) similarity.	[82]
	Fflash	Like Topomer, it can perform fragment-based similarity search using graph based clique detection procedure.	[104]

1.3 CANNABINOID RECEPTORS AND THEIR LIGANDS

1.3.1 Background and significance

GPCRs (G-protein-coupled receptors), with more than 1000 different members, are the largest receptor family in mammalian genomes and represent, to the best of our knowledge, more than 60% of all receptors which is the reason why the superfamily of GPCRs is one of the largest families of proteins in the human genome [105]. They are known to regulate wide variety of signals such as ions, hormones and neurotransmitters. It has been estimated that GPCRs represent more than 30% of current drug targets [106, 107] and have attracted many pharmaceutical industries as well as academic groups for potential drug discoveries. Cannabinoid (CB) receptors, members of G-protein coupled receptor (GPCR) superfamily, also involve in the activation of multiple intracellular signal transductions and their endogenous ligands or cannabinoids have attracted pharmacological research because of their potential therapeutic effects. In particular, the CB2 receptor is known to be involved in immune system signal transductions and its ligands have the potential to be developed as drugs to treat many immune system disorders. Currently there is no CB2 drug on the market yet although a CB1 selective drug namely Rimonabant has been withdrawn from the market recently because of its several psychotic side effects. Therefore, CB2 selective drugs may have better therapeutic potential because CB2 receptors are mainly present in the immune system

instead of the nervous system. In fact, CB2 ligands have been proven to be useful in treating autoimmune and immunological disorders such as multiple sclerosis [108]. They could be used as anti-inflammatory agents that alleviate inflammatory pain [109, 110] and as anti-cancer agents that inhibit the growth of tumors of immune origin [111]. Therefore, it is valuable to study CB2 ligands and identify novel CB2 selective ligands with the potential to be developed as therapeutic drugs. However, up until 2011 only 6 GPCR crystal structures were available, but, in 2012, additional 7 crystal structures have been resolved. This leaves us with 98% of GPCR structures remaining to be solved [112]. No experimental crystal structure of cannabinoid receptors are available yet to be used in a receptor-based drug design. Many groups including ours have developed three dimensional CB2 receptor models using homology modeling approaches based on known GPCR crystal structures. Below I will discuss some methods which have been used to model CB receptors.

1.3.2 Computational modeling of CB1 and CB2 receptors

There are two main types of cannabinoid receptors: CB1 cloned in 1990 [113], and CB2 cloned in 1993 [114]. The CB2 receptor shows 44% identity with the CB1 receptor [114, 115]. In modeling CB receptors in the past decade, the bovine rhodopsin crystal structure [116] was often used as a template. Such modeling was done via homology/comparative modeling in which rhodopsin and CB receptor sequences are compared to analyze highly conserved residues among GPCRs and such information is used to build the receptor model. The only problem with using the reported structure of rhodopsin as a template is that the structure was obtained in its inactive form without any native ligand bound to it

and, therefore, it may not provide the best representation of ligand binding site(s) of an active form of this GPCR.

One of early work on a 3D GPCR model before the availability of high resolution (2 Å) crystal structure of bovine rhodopsin was done by Joyce Baldwin [117]. The author examined a probable arrangement of the seven transmembrane helices by visually aligning 204 sequences of GPCRs and incorporating them in a database using the homologous sequence editor HOMED. As a result, the author established constraints such as each helix must be positioned next to its neighbors in the sequence, and helices I, IV, and V must be exposed to the lipid while helix III is the least exposed. His work agreed with a CB1 model proposed by Bramblett *et al* [118] at around the same time. They determined the length and orientations of the transmembrane helices of the CB1 receptor by using a variety of mathematical formulae in order to build a 3D model. During that time, the bovine rhodopsin structure at 9 Å resolution [119] became available and the structure supported the validity of a seven transmembrane alpha helix model for GPCRs. In their work, Bramblett *et al* aimed to get the ends and orientations of the seven transmembrane helices of the CB1 receptor. They first aligned the amino acid sequence of the CB1 receptor with the sequences of fifty-eight other GPCRs. Based on the alignment, they found out highly conserved residues in the GPCRs. Then using Fourier transform methods, they calculated alpha periodicity index (AP) which was used to determine the amino acids which form each of the seven transmembrane helices. Next they calculated variability moment vectors to find out the position of each helix in the membrane. Based on these vectors, they constructed a tentative helix bundle which was found to be largely consistent with the model proposed by Baldwin.

In the last decade, many modeling software packages became available and several research groups have taken advantage of them. One of the groups was Montero *et al* [115] who built homology models of the CB1 and CB2 receptors in their docking analysis study. They started out by sequential alignment of CB1 and CB2 with the Rhodopsin (Rho) as a template structure using CLUSTALW program [120]. Then they performed a manual adjustment of multiple alignment sequence with the program SEAVIEW and formatted using the program JOY [121]. Based on the best alignment, 3D models were automatically built using MODELLER [115, 122, 123]. A 3D model was obtained after the optimization of a molecular probability density function (pdf) and minimization of the models were done using the CHARMM force field with conjugate gradient method with a convergence gradient of 0.01. Out of 15 models generated by MODELLER, the model with the lowest value of the pdf and fewest restraints violations was chosen for further refinement. Short loop regions were built using MODELLER and long loops were refined using the Search-Loop option in the SYBYL program [82] software suite coupled database. The database contained protein structural fragments. After minimization of final models and refinement process, the models were validated using the VERIFY [124], PROCHECK [125] and COMPARER [126] programs. One of their results indicated that the major sequence difference between the Rhodopsin and the CB receptors lie in the transmembrane region 5 where CB receptors show strong aromatic environment. Another finding was that the main differences between CB1 and CB2 lie in the N-terminal, extracellular loop 2 (EL2), C-terminal of transmembrane helix 7 (TM7) and the C-terminal. The CB2 receptor shows a rich proline region in EL2 while CB1 has a serine located in the intracellular loop 6 (IL3) which is a phosphorylate site of

protein kinase C (PKC) [115, 127]. They also found out that only CB1 of all members of the class I of GPCR family has a long N-terminal.

Xie *et al* [128] also built the comparative CB2 model using multiple sequence alignment methods and the crystal structure of bovine rhodopsin. They used nine complete GPCR sequences from the GPCR database to align with the CB2 sequence using InsightII homology module [129]. Then they performed manual multiple sequence alignments using MSI Biosym software to locate homology aligned regions for all 10 GPCR sequences including the sequence of CB2 receptor. The aligned sequences were evaluated with the hydrophobicity scoring matrix [129] to assess the transmembrane, intra- and extracellular segments of the CB2 receptor. They analyzed conserved motifs and residues, hydrogen-bond networks and helix-tilt angles and their model was shown to be consistent with the known experimental findings derived from site-directed mutagenesis, X-ray crystallographic and NMR studies. Their modeling results indicated that a salt bridge is formed between R3.50(131) and D6.30(240), an interaction responsible for acting as an ionic lock for rhodopsin-like GPCR activation [130]. They identified the presence of a cytoplasmic helix from computer modeling and NMR experiments. They, in contrast to earlier studies, also reported novel putative binding sites of the CB2 receptor.

Gonzalez *et al* [131] has recently performed the computational modeling study of microdomains in the CB1. In their study, they built the CB1 receptor model (R) using the crystal structure of bovine rhodopsin as template for comparative/homology modeling. In this method, CB1 and rhodopsin sequences are aligned based on known highly conserved residues among GPCRs. Once the alignment information was obtained, the authors used

several software packages such as MODELLER [122] to generate comparative models and PROSA II [132] and Verify 3D [133] to evaluate those models. The best model was further refined by optimizing the loop regions via a simulated annealing (SA) technique in which backbone residues of transmembrane helices were constrained and loop regions were optimized. A conjugate gradient energy minimization was followed using CHARMM22 force field. They also generated a second conformation of CB1 model (R*) which could be a putative active state of CB1 by perturbing transmembrane helices 3 and 6. Choosing these particular helices was based on empirical evidence that these two helices are proved to have an important role in the activation of various class A GPCRs. The authors used the rotating constraint (RC) module implemented in NAMD v2.5 [134] for such perturbation. The simulation was then carried out using the CHARMM22 force field for 1 ns in vacuum. Their simulation results indicated that their CB1 model is a useful starting point for further structural characterization of CB1 receptor.

1.3.3 Computational design and screening of cannabinoid ligands

More than five major classes of structurally different, but biologically similar CB ligands have been discovered (Figure 1-6). Four structural classes of representative CB agonists: Δ^9 -THC, CP-55940, AEA, and WIN5212-2 are known to bind on both CB1 and CB2 receptors. A few antagonists, such as SR141716A and SR144528 [135], have been discovered for specific binding of only one of each receptor subtype. SR141716A is the first CB1 selective antagonist displaying nanomolar affinity to the CB1 receptor but has a low activity to the CB2 receptors [136] whereas SR144528 is a CB2 antagonist ligand.

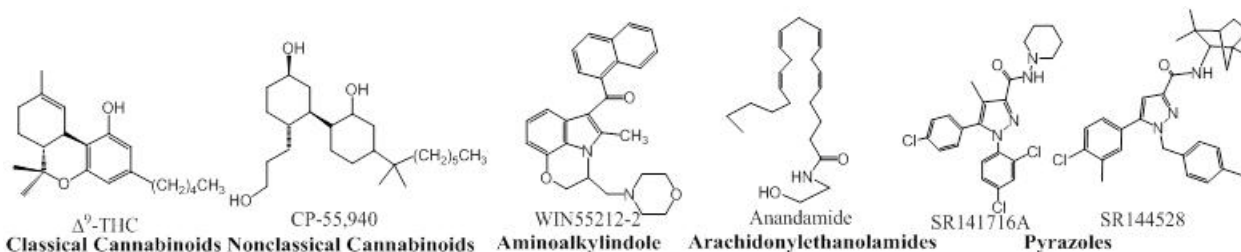


Figure 1-6. Representative structures of major classes of cannabinoid ligands

Our group has reported 3D QSAR (Quantitative Structure-Activity Relationship) study of CB1 and CB2 pyrazole-type agonistic and antagonistic specificities using comparative molecular field analysis (CoMFA) approach [5, 30]. (-)- Δ^9 -tetrahydrocannabinol (THC), a classical cannabinoid (CC), is the most active constituent in marijuana [137]. Early SAR studies [138-141] generated a 3-point pharmacophore model with a free phenolic OH, an optimal 1, 1'-dimethylheptyl side chain, and a hydroxyl in the cyclohexyl ring that enhances CB activity. Anandamide, also known as *N*-arachidonylethanolamine (AEA), was isolated from porcine brain and identified as an endogenous CB ligand [142]. AEA binds to both CB receptor subtypes and induces certain CB activities [143, 144]. Win55212-2, one of the most potent aminoalkylindoles (AAIs), was synthesized by Sterling Winthrop, Inc. [145, 146]. AAIs were originally synthesized as cyclooxygenase inhibitors [147, 148]. Some other CB ligands have been discovered, such as 2-AG (2-arachidonyl-glycerol) [149, 150], biphenylic CB analogs [151, 152], triaryl bis-sulfones [153-155] and as well as other novel inverse agonists [156, 157]. Computational design of CB ligands is mainly based on experimental SAR studies of known CB ligands and partially based on 3D models of CB receptors. Several

methods including ligand-based design, structure-based design and de novo algorithms which have been used for the design and screening of CB ligands are discussed below.

Rogers-Evans and co-workers [158] applied one of de novo design algorithms called TOPAS (*TOP*ology-Assigning System) [159] to identify novel CB1 ligands. TOPAS is based on an evolution algorithm which uses molecular fragments as building blocks to design novel compounds. In their work, they focused on the indole cannabinoids derived from pyrrole series to generate novel structures via scaffold hopping using TOPAS algorithm. To construct new molecules, they first collected 1381 known GPCR modulators which underwent fragmentation using the RECAP procedure [160]. Then using TOPAS algorithm, they constructed new generation of virtual molecules iteratively by taking the best of the previous generation as the ‘parent’. They computed the fitness of molecules as the pair-wise similarity to the seed compound. Finally, they tested their compounds in vitro and found out two novel selective hits which could undergo further refinement and lead optimization.

An example of a structure-based virtual screening approach to find CB2 antagonists was done by our lab [161]. In this work, the Xie group’s previous 3D CB2 receptor homology model was used and a 3D database query algorithm was developed to examine the receptor model and predict binding pocket. First, an antagonist (SR144528)-bound CB2 receptor model was developed via flexible docking and molecular dynamic/mechanic simulation using important binding residues information obtained from site mutagenesis data. The generated ligand-bound CB2 model was then used to analyze the antagonist binding site of CB2. The predicted binding pocket was then evaluated using a testing compound database. The FlexX-Pharm docking algorithm was

then used for virtual screening and consensus scoring (CScore) procedure to rank hits from five different scoring functions: FlexX [86, 87], PMF [162], ChemScore [163], D_Score [164], and G_Score [165]. Finally, the enrichment factor was calculated to evaluate the performance of different CScore functions for the CB2 antagonists. The results indicated that the CB2 model can be a promising model for future structure-based CB2 leads screening.

The Xie lab also developed ligand-based virtual screening protocol [5] in which 3D-QSAR studies of arylpyrazole antagonists of CB1 and CB2 in combination with 2D NMR spectroscopy were performed. First, the solution conformation of *N*-(piperidin-1-yl)-5-phenyl-1-(*n*-pentyl)-4-methyl-1*H*-pyrazole-3-carboxamide (AM263), a high affinity arylpyrazole analogue, was determined using NOESY NMR data and computer-assisted conformational search in which molecular dynamic (MD) and mechanic (MM) simulations were performed using the Tripos Sybyl software package [82]. The starting structure of AM263 was built using Sybyl and performed conformation simulation by first relaxing the initial compound structure through minimization. MD simulations were then carried at 1000 K using Tripos force field to sample different conformations that were then minimized using a combination of the steepest descent and conjugate gradient methods. Twenty-nine arylpyrazole compounds in the training set for parallel CoMFA (comparative molecular field analysis) analyses were used to build 3D-QSAR models. These models were based on experimental binding affinity values of arylpyrazole analogues for CB1 and CB2 receptors. Different conformations of AM263 from previous MD simulations were used as templates to align compounds in the training set and nine other compounds from literature were used in the test set to evaluate 3D-QSAR models.

In such evaluation, cross-validated r^2 values were calculated via CoMFA Partial Least-Squares (PLS) analysis. 3D-QSAR/CoMFA study suggested a bioactive conformation of arylparazole that was congruent to its solution conformation. Therefore, CoMFA analyses gave necessary information to better understand the binding characteristics of arylpyrazoles to the CB1 and CB2 receptors.

1.4 OUTLINE OF THE DISSERTATION

There are two main parts in this dissertation: novel QSAR methods development and discovery of new CB2 ligands. While fulfilling those two major goals, our research work has been reported in 5 manuscripts including 4 research articles and 1 review article. There are 4 manuscripts related to QSAR methods development and 1 manuscript for discovery of novel CB ligands which are discussed in greater details in later chapters. Specifically, I and Prof. Xie reviewed recent fragment-based and other novel QSAR methods in:

- Myint, K-Z. and Xie, X-Q. Recent Advances in Fragment-Based QSAR and Multi-Dimensional QSAR Methods, *International Journal of Molecular Sciences* 11 (2010), pp. 3846-3866.

In addition, I and others in the Xie lab have published our results from a novel fragment-based QSAR method in:

- Myint, K-Z., Ma, C., Wang, L. and Xie, X.Q. Fragment-Similarity-Based QSAR (FS-QSAR): A Novel 2D-QSAR Method to Predict Biological Activities of

Triaryl Bis-sulfone and COX2 Analogs, *SAR and QSAR in Environmental Research* 22 (2011), pp. 385-410.

Moreover, the results from combining pharmacophoric and morphological descriptors to derive a novel QSAR model were reported in:

- Chen, J-Z., Myint, K-Z. and Xie, X-Q. A Novel QSAR Approach for GPCR CB₂-Antagonistic Triaryl Bis-Sulfone Analogs: A Combined Molecular Morphological and Pharmacophoric Approach, *SAR and QSAR in Environmental Research* 22 (2011), pp. 525-44.

In order to tackle similar-scaffold requirement limitations of fragment-based, CoMFA and other 3D-QSAR methods, a novel fingerprint-based artificial neural networks QSAR (FANN-QSAR), which no longer requires compounds to have similar or same scaffold to make effective ligand activity predictions, was developed. This work has been accepted for publication in the journal *Molecular Pharmaceutics*:

- Myint, K-Z., Wang, L. and Xie, X-Q. Fingerprint-based Artificial Neural Networks QSAR (FANN-QSAR) for Ligand Biological Activity Predictions, *Molecular Pharmaceutics* (2012). Accepted.

Moreover, the Xie group's discovery of novel CB ligands has been reported in a recently submitted manuscript:

- Yang, P.*, Myint, K-Z.*, Cao, H., Tong, Q., Almehizia, A., Alqarni, M., Feng, R., Wang, L. and Xie, X-Q. Lead discovery, chemistry optimization and biological

evaluation studies of novel bi-amide derivatives as CB2 receptor inverse agonists and osteoclast inhibitors, *Journal of Medicinal Chemistry* (2012). Submitted. (*: *equal contribution*)

These developed methods, results and discussions are further described in details in Chapters 2-5.

2.0 FRAGMENT-SIMILARITY-BASED QSAR (FS-QSAR) ALGORITHM FOR LIGAND BIOLOGICAL ACTIVITY PREDICTIONS

2.1 INTRODUCTION

Various QSAR methods have been developed and they differ in which statistical method is used and which molecular descriptors are used [166]. As previously discussed, 3D-QSAR has some known problems such as subjective molecular alignment and bioactive conformations whereas 2D-QSAR is attractive because predicting molecular properties and activities based on 2D molecular structures is simple, fast and robust. 2D-QSAR methods allow modeling of a wide variety of ligands or compounds including cases where 3D crystal receptor or target structures are not available [6]. As described previously, QSAR methods were first introduced by Free, Wilson, Hansch, and Fujita [7, 8]. In Free-Wilson QSAR models [7], biological activity values were correlated with the presence and absence of molecular substituents, whereas Hansch-Fujita model [8] predicted biological activities based on physicochemical parameters or properties. Typically, such molecular properties used in the Hansch-Fujita equation were calculated based on a whole molecule structure. On the other hand, the Free-Wilson method focused on the effect of substituent changes on the overall biological activity. In general, it is often noted that Hansch-Fujita models are not unique, i.e. many models can be derived

for the same dataset [167], whereas Free-Wilson models are unique but predictions are only possible for combinations of existing chemical substituents that appear in the training set, but not for any other substituents. In other words, the traditional QSAR model often has poor predictions if substituents in the test set do not exist in the training set.

Therefore, a novel fragment-similarity-based QSAR method (FS-QSAR) was developed to address such major limitations of the traditional method by introducing a fragment-similarity concept. In this FS-QSAR method, any testing substituent or fragment not found in the training set were compared to those training fragments, and the most similar training fragment was selected along with its corresponding regression coefficient. Regression coefficients were then weighted by the similarity values between testing and training fragments, and overall biological activity of a compound was predicted by considering such corrected fragment contributions. The FS-QSAR method was examined and validated in various cases studies, including a dataset of COX2 inhibitors [168] and a dataset of CB2 triaryl bis-sulfone antagonist analogs [153-155]. Quantitative predictions demonstrate that the FS-QSAR method improved the traditional QSAR method remarkably and affords a robust and effective model that can be used to predict biological activities of target compounds.

2.2 METHODS

2.2.1 Data sets

The FS-QSAR model was trained and tested on two independent datasets: COX2 inhibitors and triaryl bis-sulfone analogs. The COX2 inhibitor dataset is a well-known dataset in QSAR studies [168, 169]. There are several analogs of COX2 inhibitors and Table 2-1 lists all 83 compounds used in this work. The second dataset is a set of 85 triaryl bis-sulfone analogs (Table 2-2, Table 2-3, Table 2-4) that have strong CB2 binding affinities with K_i 's less than 1000 nM. The bis-sulfone analogs were obtained from three different papers that were reports of studies conducted by the same group of authors [153-155].

2.2.2 Computational method

Fragment-similarity based QSAR (FS-QSAR)

The fragment-similarity-based QSAR (FS-QSAR) method is an improvement of the traditional Free-Wilson method which is a linear combination of the effects of substituents. In order to solve the limitation of the Free-Wilson model discussed above, we developed the fragment-similarity-based QSAR method. In this approach, each testing fragment not present in any of training compound is compared to those existing training fragments. Then the most similar fragment is picked for each substituent position and contributions of such fragments are used to predict the overall biological activity of the compound. In the original method, fragments of a new compound have to be in the

training set for an effective activity prediction. In this new method, such requirement is no longer necessary and the most similar training fragments will be used if they are not present in the training set. The method is based on the hypothesis which implies similar functional groups may have similar chemical and physical properties which attribute to similar biological activities. The hypothesis is induced based on a central premise of medicinal chemistry which states structurally similar molecules have similar physicochemical properties and possibly similar biological activities and has been stated in many computational chemistry and similarity-based virtual screening drug design studies [170-177].

The FS-QSAR equation can be described as:

$$\log(BA) = const + \sum_{j=1}^N [\max_{k=1}^{P_j} \{Sim(F_{jk}, F_{jg})\}] \times A_j^{MSF} \quad (2.1)$$

BA = biological activity of a given compound (such as K_i or IC_{50}).

N = a total number of substituent positions.

P_j = a total number of possible substituents at the j^{th} substituent position.

\max = a maximum function which picks the maximum score among similarity scores.

F_{jk} = the k^{th} fragment (a known fragment in the training set) at the j^{th} substituent position.

F_{jg} = a given fragment (the fragment from a testing/unknown compound) at the j^{th} substituent position.

$Sim[F_{jk}, F_{jg}]$ = the fragment similarity function compares F_{jg} to F_{jk} and calculates a similarity score.

A_j^{MSF} = the coefficient of the most similar fragment (MSF) at the j^{th} substituent position.

Multiple linear regression (MLR) implemented in Matlab® computing software [178] was used to solve for all the variables in the equation. Molecular substituents or fragments found in training molecules represent independent variables and no duplicate training fragments were included. Corresponding regression coefficients (A_j^{MSF}) for training fragments and the constant (*const*) were solved via MLR. $Sim[F_{ik}, F_{ig}]$ represents either BCUT-similarity or Tanimoto similarity function, which calculates a similarity score between a training and testing fragment. The score is used as a weighting factor for the regression coefficient of the most similar training fragment.

BCUT-similarity function

The chemistry space BCUT-similarity function is defined as:

$$Sim(F_{jk}, F_{jg}) = e^{-\left|EV(F_{jk}) - EV(F_{jg})\right|} \quad (2.2)$$

$EV(F_{jk})$ = lowest/highest eigen value of BCUT matrix of a fragment (F_{jk}).

BCUT matrix [20] of a fragment is defined as below. For an $n * n$ square matrix, each atom in a fragment from atom 1 to atom n is represented in each row and column of the matrix. s represents the scaling factor to scale the off-diagonal elements. A diagonal element (m_{xx}) is equal to the partial charge of the x^{th} atom. An off-diagonal element (m_{xy}) is equal to the number of bonds between the x^{th} atom and y^{th} atom, i.e. m_{xy} is equal

	1	2	...	n		
1	$\begin{bmatrix} m_{11} & m_{12}S & \dots & m_{1n}S \\ m_{21}S & m_{22} & \dots & m_{2n}S \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ m_{n1}S & m_{n2}S & \dots & m_{nn} \end{bmatrix}$				to 1 for single bond, 2 for double bonds, 3 for	
2					triple bonds, 1.5 for conjugated bonds and 0.01	
.						if there are no direct bond(s) between two (<i>x</i>
.						and <i>y</i>) atoms. There are two other parameters
.						used in constructing BCUT matrices: whether
n					hydrogens should be kept (K) or removed (R),	
					and whether to use the lowest (L) or highest	

(**H**) eigen value. **K** (keep hydrogens) or **R** (remove hydrogens) defines if hydrogen atoms are considered when constructing the matrix. Lowest (**L**) or highest (**H**) eigen values of BCUT matrices are used for similarity measurement. In this work, it was decided to keep hydrogens (**K**) since hydrogen bonds are important for ligand-protein binding. **L** (lowest eigen value) or **H** (highest eigen value) and scaling factor were decided by the leave-one-out cross validation (LOOCV) for each training dataset. The parameters which gave the best cross-validated r^2 (q^2) value were used to test the model.

Tanimoto similarity function

In Tanimoto similarity calculation [177, 179], each fragment is represented as strings of 0 and 1. Based on these bits (fingerprint), the Tanimoto similarity function is defined as:

$$Tc(\text{fragment 1}, \text{fragment 2}) = \frac{c}{a + b - c} \quad (2.3)$$

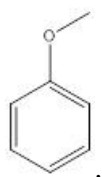
where *a* = number of '1' bits in fragment 1; *b* = number of '1' bits in fragment 2; *c* = number of '1' bits common in both fragments. Tripos Selector program [77] was used

for the fingerprint generation [180] and Tanimoto similarity calculations. **Error!**
Bookmark not defined.

2.3 CALCULATIONS

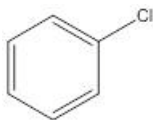
2.3.1 Partial charge calculation and fragment generation

Before fragments were generated, partial charges of individual atoms in a totally connected compound were calculated. The type of partial charge is the Gasteiger-Hückel type calculated using Sybyl8.0 from Tripos [77]. Then desired fragments were cleaved. It is important to calculate charges before cleaving the fragments since charges will not be the same if they are calculated after cleavage. By calculating as a whole molecule, the effects of neighboring atoms and their electron environment are considered. When fragments were retrieved from COX2 inhibitor analogs, a benzene ring was also regarded as part of the fragment. For example, when an R₂ fragment of a COX2 inhibitor analog was 4-OMe, not only the oxygen-methyl (-OMe) group but also the benzene ring attached to the fragment were considered as follows:



The benzene ring was included as part of a fragment to prevent from having a fragment with only one atom when a substituent is an atom such as chlorine (Cl) instead of a functional group. This rule was applied consistently to all fragments generated.

Similarly, fragments for bis-sulfone analogs were generated as follows. For fragment linkers at L₂ position, two neighbor carbon atoms were included as part of any given L₂ fragment. For example, when L₂ linker fragment was mentioned as an oxygen (O), a C-O-C fragment was used instead of only O, since a fragment should have more than one atom. To be consistent, all other L₂ fragments were considered in the same manner. Similarly, for X and Y substituent positions, we included the benzene-ring as part of the fragment. For example, when an X-fragment was *4-Cl*, then instead of using Cl (chlorine) atom alone, the following fragment was considered:



2.3.2 Parameter tuning using leave-one-out cross-validation (LOOCV)

Two parameters (scaling factor and **H** or **L** parameters) involved in constructing BCUT matrices were determined using the LOOCV method. For each BCUT matrix derived from each fragment, eigen values were calculated using the JAMA/C++ library from TNT (Template Numerical Toolkit) package [181] and the lowest or highest eigen value was kept for each fragment. For a given training set, scaling factors were varied from 0.005 to 1.0 with 0.005 increment and **H** and **L** were used alternatively. For each pair of parameters (scaling factor and **H** or **L**), corresponding BCUT matrices and eigen values were calculated for every fragment in the training set. Then FS-QSAR model was trained using the LOOCV method and the pair of parameters which gave the best cross-validated correlation coefficient value (q^2 value) was chosen for model testing. During the LOOCV

process, for a dataset containing N compounds, the model was trained using $N-1$ compounds and tested on the excluded n^{th} compound. The process was repeated for $n = 1, 2, 3, \dots, N$ and the cross-validated correlation coefficient (q^2) was then calculated.

2.3.3 Generation of training and testing data sets

For each dataset (COX2 inhibitor analogs and Bis-sulfone analogs) the training and testing sets were generated by randomly picking compounds from the original dataset. For statistical FS-QSAR modeling, the process was repeated five times resulting in five different pairs of training and testing sets (Table 2-5 and Table 2-6). During the training process, the constant ($const$) and the regression coefficients (A_j^{MSF}) from Equation 2.1 were solved and were used in subsequent model testing.

2.4 RESULTS AND DISCUSSION

2.4.1 FS-QSAR modeling on COX2 inhibitor analogs

COX2 inhibitor dataset is a well-known dataset used in many QSAR studies [168, 169]. A total of 83 COX2 analogs having the same scaffold were used, 69 of which were used as training data and 14 were used as testing data. All q^2 values (Table 2-7) provided good correlations and the average r^2 of five different testing sets was 0.62 (Table 2-9, “FS-QSAR (BCUT)” column) with the variance of 0.01, which inferred the consistency of the model in predicting activities of random testing compound sets.

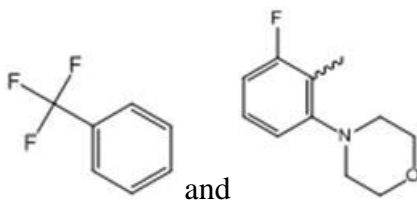
2.4.2 FS-QSAR modeling on bis-sulfone analogs

The FS-QSAR regression model was generated and tested on bis-sulfone analogs. Eighty-five bis-sulfone analogs with strong CB₂ binding affinities ($K_i < 1000$ nM) were selected. All q^2 values (Table 2-8) provided good correlations and the average r^2 of five different testing sets was 0.68 (Table 2-10, “FS-QSAR (BCUT)” column), which implied that the model predicted biological activities effectively for the bis-sulfone dataset as well as the COX2 dataset. In other words, the model worked effectively for the well-known and standard QSAR dataset such as COX2 inhibitor analogs as well as a real world cannabinoid ligand dataset such as bis-sulfone analogs.

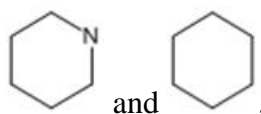
2.4.3 BCUT-similarity score analysis

As mentioned, BCUT similarity function calculates fragment similarity values and finds a training fragment which is most similar to a given/unknown testing fragment if the testing fragment does not exist in the training set. Results from such similarity calculation for bis-sulfone fragments can be found in Table 2-11. Specifically, column 1 contains testing compounds found in one of five testing sets (column 2). Column 3 refers to fragments of the testing compound at a particular substituent position which were not found in the training fragments and was necessary to match and find the most similar training fragment at the substituent position with a non-zero regression coefficient. Those similar training fragments are listed in column 4. Column 5 lists substituent positions where the testing and training fragments are found and column 6 contains two parameters (scaling factor and H or L) which were used in calculating BCUT similarity scores.

Column 7 lists the BCUT-similarity score between the testing fragment and the training fragment calculated using the Equation 2.2. It can be observed that the majority of training fragments selected by the BCUT similarity function resembled their counterpart testing fragments. There were, however, some drawbacks of the method. For example, one would expect the following pair of fragments (for “1,29” testing compound, Table 2-11):



to have a lower similarity value than the following pair of fragments (for “3,50” testing compound, Table 2-11):



This is because of the presence of intrinsic drawback of the method; this is further explained in the following section by comparing to the Tanimoto method, a molecular similarity comparison method that is widely used in two-dimensional molecular comparisons [177]. It is important to note that the similarity calculation was used as a part of the regression equation, and the model should be judged based on final statistical performances rather than individual fragment similarity scores.

2.4.4 Comparisons with different approaches

Traditional Free-Wilson QSAR and 1-nearest-neighbor-QSAR (1NN-QSAR) analyses were performed on the testing sets used in the FS-QSAR model testing for comparison. For all the methods, predictions were done for all the compounds in each testing set. Free-Wilson QSAR failed to make successful predictions on many testing sets and achieved an average correlation coefficient of 0.46 for COX2 inhibitor dataset (Table 2-9, “Free-Wilson QSAR” column) and 0.42 for bis-sulfone dataset (Table 2-10, “Free-Wilson QSAR” column). This is because of the major limitation of the Free-Wilson method where if there is a new fragment which is not in the training set, the activity prediction for a compound with such fragment becomes less accurate. However, FS-QSAR method achieved average r^2 values of 0.62 for COX2 inhibitor dataset and 0.68 for bis-sulfone dataset and performs better than the traditional Free-Wilson QSAR method because of the addition of fragment-similarity concept to the regression equation.

In order to better compare the FS-QSAR model, 1-nearest-neighbor QSAR (1NN-QSAR) analysis were performed. In this analysis, whole-molecule fingerprints were used to calculate similarities among testing and training molecules and the activity of the most similar training molecule was predicted as that of the testing molecule. Tripos Selector program [77] was used to generate fingerprints and calculate similarity scores. Using whole-molecule fingerprints allowed us not to split molecules into scaffolds and substituents as in the Free-Wilson and FS-QSAR methods and to better compare our method. It was observed that 1NN-QSAR failed to make any successful predictions on both COX2 and bis-sulfone datasets resulting in average r^2 values of 0.37 and 0.21 respectively (Table 2-9 and Table 2-10 “1NN-QSAR” column). The results implied that

non-linear whole-molecule-fingerprint 1-nearest-neighbor QSAR approach is not suitable for datasets containing highly similar molecules varied only to a small extent on R-groups and the FS-QSAR approach is more appropriate for such studies. In addition, whole-molecule-fingerprint model may not be suitable for mechanistic interpretations whereas the FS-QSAR model improves the predictability of the original method while retaining the ease of mechanistic interpretation. In particular, for testing molecules having training fragments, one would interpret the model as one would in the Free-Wilson model while for testing molecules having fragments other than training fragments, the FS-QSAR model provides most similar training fragments which are used in activity prediction of the testing molecule. These fragments may also be used in mechanistic interpretation since, for example, they may help a chemist in analog selection with which kind of fragments to further synthesize.

Moreover, the BCUT similarity function as defined by Equation 2.2 had some intrinsic drawbacks and, in order to justify the function, the performance of the BCUT-similarity function was compared to the traditional Tanimoto coefficient (T_c) similarity function (Equation 2.3). The BCUT-similarity function was simply replaced with the T_c similarity function and the quantitative activity prediction on the same testing sets was performed. The Tripos *Selector* program [77] was used for T_c value calculations. The FS-QSAR with Tanimoto similarity function (FS-QSAR-Tanimoto) was tested on the same five testing sets used for testing the FS-QSAR with BCUT similarity function (FS-QSAR-BCUT). Table 2-9 and Table 2-10 contain the prediction summaries for COX2 inhibitor and bis-sulfone data sets. The average r^2 on COX2 inhibitor testing datasets was 0.62 from both FS-QSAR-BCUT and FS-QSAR-Tanimoto whereas the average r^2 on bis-

sulfone testing datasets were 0.68 and 0.62 respectively. It can be inferred that the BCUT-similarity function performs similarly as the Tc similarity function for COX2 dataset and performs better for bis-sulfone dataset in terms of prediction accuracy (r^2).

In order to better compare the two similarity methods and further address an intrinsic drawback present in each similarity function, random fragment pairs from bis-sulfone analogs were selected and the corresponding Tc and BCUT similarity scores were calculated. The scores were then plotted in Figure 2-1. The curve fitting line followed the functional form of $BCUT = \log(T_c)$ with $r^2 = 0.52$ and it can be observed that for small Tc scores, BCUT scores did not agree well and there were significant score differences between those fragment pairs. However, they agreed with higher Tc similarity values. This illustrates the presence of inherent weakness in each method and can be better explained by looking at some of example fragment pairs from Figure 2-1. 7 fragment pairs (labeled circles) were selected and they are listed in Table 2-12. Pair 1 and 5 illustrates that Tc similarity function is better than BCUT method because in pair 1, fragments are not similar and Tc scores reflect such dissimilarity but BCUT failed to do so. In pair 5, two fragments are identical except an extra ring in one of fragments and Tc score indicates they have approximately 60% similarity which is reasonable. On the other hand, pairs 2 and 7 shows BCUT method is superior to Tc method since in pair 2, fragments have similar pharmacophore features such as hydrogen bond donor/acceptor atoms and hydrophobic rings, and BCUT score reflects such similarities to some extent. In pair 7, two fragments are structurally identical, but they are not identical in terms of their electron environments and partial charges since they are attached to different substituent places on the bis-sulfone analogs. Such subtle differences can be captured by

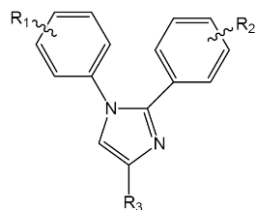
BCUT method while Tc function fails to produce any difference between them. Pairs 3 and 4 explain that both methods have weakness since both Tc and BCUT similarity scores do not reflect all the structural and feature similarities between each pair of fragments. For these two pairs, Tc scores were too low while BCUT scores were too high. However, in pair 6 the only difference between two fragments is an extra methyl group and both methods produce reasonable scores which mirror the similarity between fragments. Hence these particular fragment pairs show intrinsic drawbacks of each similarity function which reflect similarity scores and choices of similar fragments. However, it should be noted that the similarity function is used as a part of the FS-QSAR equation, and both similarity functions performed similarly and improved the traditional Free-Wilson QSAR method.

2.5 CONCLUSION

In this study, a novel and effective QSAR method has been developed to compensate the major limitation of the traditional QSAR method. As a result, the FS-QSAR method has an improved predictive power. The increased performance was attributed to the newly-introduced fragment-similarity concept which tackles the major limitation of the original method. Two types of similarity functions were implemented in the FS-QSAR method for fragment similarity calculations. It was observed that the BCUT similarity function performed similarly as the Tc similarity function for COX2 dataset, and performed better for bis-sulfone dataset. The FS-QSAR method is for the first time presented as the 2D-QSAR method using fragment similarity concept to model cannabinoid (bis-sulfone)

dataset and COX2 inhibitor dataset, showing the superior predictive power when compared to the traditional and 1NN-QSAR methods. To conclude, the FS-QSAR is robust and easier to implement with good prediction accuracy and this approach provides a novel fragment-based QSAR method which can potentially be a useful tool in fragment-based drug discovery for ligand biological activity prediction and analog selection.

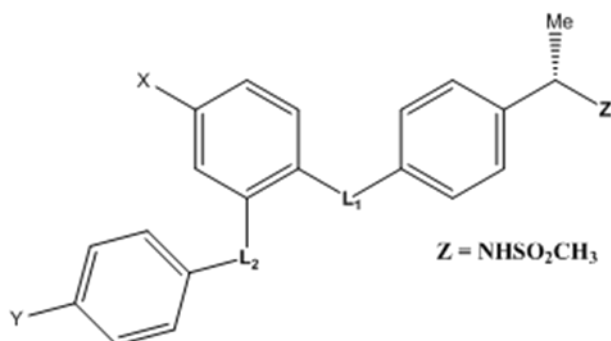
Table 2-1. COX2 analogs [168] used for FS-QSAR modeling



Compound	R ₁	R ₂	R ₃	IC ₅₀ (μM)
3-5	-4-SO ₂ Me	-4-Cl	-Me	0.24
3-6	-4-SO ₂ Me	-4-Cl	-CF ₃	0.11
3-9	-4-SO ₂ Me	-4-F	-CF ₃	0.1
3-10	-4-SO ₂ Me	-H	-CF ₃	0.12
3-11	-4-SO ₂ Me	-4-Me	-CF ₃	0.16
3-12	-4-SO ₂ Me	-4-OMe	-CF ₃	0.57
3-13	-4-SO ₂ Me	-4-NHMe	-CF ₃	1.47
3-14	-4-SO ₂ Me	-4-NMe ₂	-CF ₃	0.7
3-15	-4-SO ₂ Me	-4-SMe	-CF ₃	0.16
3-17	-4-SO ₂ Me	-4-SO ₂ Me	-CF ₃	5.7
3-18	-4-SO ₂ NH ₂	-4-Cl	-CF ₃	0.01
3-19	-4-SO ₂ NH ₂	-4-F	-CF ₃	0.01
3-20	-4-SO ₂ NH ₂	-H	-CF ₃	0.04
3-21	-4-SO ₂ NH ₂	-4-Me	-CF ₃	0.04
3-22	-4-SO ₂ Me	-3-Cl	-CF ₃	0.06
3-23	-4-SO ₂ Me	-3-F	-CF ₃	0.12
3-24	-4-SO ₂ Me	-3-Br	-CF ₃	0.08
3-25	-4-SO ₂ Me	-3-Me	-CF ₃	0.06
3-26	-4-SO ₂ Me	-3-CF ₃	-CF ₃	0.21
3-27	-4-SO ₂ Me	-3-OMe	-CF ₃	0.35
3-28	-4-SO ₂ Me	-3-SMe	-CF ₃	0.35
3-29	-4-SO ₂ Me	-3-CH ₂ OMe	-CF ₃	68.1
3-30	-4-SO ₂ Me	-3-NMe ₂	-CF ₃	3.2
3-31	-4-SO ₂ Me	-3-NHMe	-CF ₃	0.92
3-32	-4-SO ₂ Me	-3-NH ₂	-CF ₃	5.89
3-33	-4-SO ₂ Me	-3-NO ₂	-CF ₃	0.58
3-34	-4-SO ₂ NH ₂	-3-Cl	-CF ₃	0.008
3-35	-4-SO ₂ NH ₂	-3-F	-CF ₃	0.03
3-36	-4-SO ₂ NH ₂	-3-Br	-CF ₃	0.007
3-37	-4-SO ₂ NH ₂	-3-Me	-CF ₃	0.03
3-38	-4-SO ₂ Me	-2-Cl	-CF ₃	0.9
3-39	-4-SO ₂ Me	-2-F	-CF ₃	0.4
3-40	-4-SO ₂ Me	-2-Me	-CF ₃	0.8
3-42	-4-SO ₂ NH ₂	-2-F	-CF ₃	0.1
3-43	-4-SO ₂ NH ₂	-2-Me	-CF ₃	0.2
3-44	-4-SO ₂ Me	-4-OMe-3-F	-CF ₃	0.15
3-45	-4-SO ₂ Me	-4-OMe-3-Cl	-CF ₃	0.13
3-46	-4-SO ₂ Me	-4-SMe-3-Cl	-CF ₃	0.04
3-47	-4-SO ₂ Me	-4-NMe ₂ -3-Cl	-CF ₃	0.32
3-48	-4-SO ₂ Me	-4-NMe ₂ -3-F	-CF ₃	0.33
3-49	-4-SO ₂ Me	-4-NHMe-3-Cl	-CF ₃	0.66
3-50	-4-SO ₂ Me	-4-Me-3-Cl	-CF ₃	0.03
3-51	-4-SO ₂ Me	-4-Me-3-F	-CF ₃	0.11
3-52	-4-SO ₂ Me	-3-Me-4-F	-CF ₃	0.17
3-53	-4-SO ₂ Me	-3-Me-4-Cl	-CF ₃	0.09
3-54	-4-SO ₂ Me	-3-OMe-4-Cl	-CF ₃	0.25
3-55	-4-SO ₂ Me	-3-NMe ₂ -4-Cl	-CF ₃	1.04
3-56	-4-SO ₂ Me	-3,4-OCH ₂ O	-CF ₃	0.17
3-57	-4-SO ₂ Me	-3,4-F	-CF ₃	0.12
3-58	-4-SO ₂ Me	-3,4-Me	-CF ₃	0.33
3-59	-4-SO ₂ Me	-3-Me-5-Cl	-CF ₃	0.08
3-60	-4-SO ₂ Me	-3-Me-5-F	-CF ₃	0.11
3-61	-4-SO ₂ Me	-3-OMe-5-F	-CF ₃	0.96
3-63	-4-SO ₂ Me	-3,5-Cl	-CF ₃	0.17
3-66	-4-SO ₂ NH ₂	-4-OMe-3-F	-CF ₃	0.03
3-67	-4-SO ₂ NH ₂	-4-OMe-3-Cl	-CF ₃	0.02
3-68	-4-SO ₂ NH ₂	-4-OMe-3-Br	-CF ₃	0.03
3-69	-4-SO ₂ NH ₂	-4-SMe-3-Cl	-CF ₃	0.01
3-70	-4-SO ₂ NH ₂	-4-Me-3-Cl	-CF ₃	0.003
3-71	-4-SO ₂ NH ₂	-3-OMe-4-Cl	-CF ₃	0.02
3-72	-4-SO ₂ NH ₂	-3,4-F	-CF ₃	0.03
3-73	-4-SO ₂ NH ₂	-3-Me-5-Cl	-CF ₃	0.04
3-74	-4-SO ₂ NH ₂	-3-Me-5-F	-CF ₃	0.03

3-75	-4-SO ₂ NH ₂	-3-OMe-5-Cl	-CF ₃	0.46
3-76	-4-SO ₂ Me	-4-OMe-3,5-F	-CF ₃	0.17
3-77	-4-SO ₂ Me	-4-OMe-3,5-Cl	-CF ₃	0.14
3-78	-4-SO ₂ Me	-4-OMe-3,5-Br	-CF ₃	0.09
3-79	-4-SO ₂ Me	-4-OMe-3,5-Me	-CF ₃	0.72
3-80	-4-SO ₂ Me	-4-OMe-2,5-Me	-CF ₃	12.2
3-81	-4-SO ₂ Me	-4-NMe ₂ -3,5-Cl	-CF ₃	0.14
3-82	-4-SO ₂ NH ₂	-4-OMe-3,5-F	-CF ₃	0.03
3-87	-4-SO ₂ Me	4-Cl	-CHF ₂	0.61
3-88	-4-SO ₂ Me	-4-Cl	-CH ₂ F	0.41
3-89	-4-SO ₂ Me	-4-Cl	-CHO	1.6
3-90	-4-SO ₂ Me	-4-Cl	-CN	0.23
3-91	-4-SO ₂ Me	-4-Cl	-CO ₂ Et	5.7
3-96	-4-SO ₂ Me	-4-Cl	-Ph	0.24
3-97	-4-SO ₂ Me	-4-Cl	-CH ₂ OC ₆ H ₄ -4-Cl	0.03
3-98	-4-SO ₂ Me	-4-Cl	-CH ₂ SC ₆ H ₄ -4-Cl	0.05
3-99	-4-SO ₂ Me	-4-Cl	-CH ₂ OMe	3.72
3-100	-4-SO ₂ Me	-4-Cl	-CH ₂ OH	8.35
3-101	-4-SO ₂ Me	-4-Cl	-CH ₂ SMe	0.32
3-103	-4-SO ₂ Me	-4-Cl	-CH ₂ CN	1.54

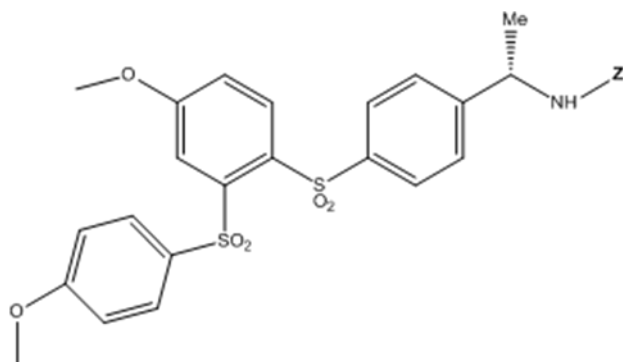
Table 2-2. Bis-sulfone analogs from Shankar *et al.* [155] used for FS-QSAR modeling



Compound	L1	L2	X	Y	K _i
1,23	CH ₂	SO ₂	4-OCH ₃	4-OCH ₃	0.6
1,8	CH ₂	SO ₂	4-Cl	4-Cl	6.7
1,24	CH ₂	SO ₂	4-OCF ₃	2-F	0.4
1,25	CH ₂	SO ₂	4-CF ₃	2-F	1
1,26	CH ₂	SO ₂	4-CF ₃	2,6-DiF	1.8
1,27	CH ₂	SO ₂	4-OCF ₃	2-OCF ₃	16
1,28	CH ₂	SO ₂	4-CF ₃	3-F	3.5
1,29	CH ₂	SO ₂	4-OCF ₃	3-CF ₃	35
1,30	CO	SO ₂	4-OCH ₃	4-OCH ₃	44
1,31	CO	SO ₂	4-OCF ₃	2-F	179
1,32	CO	SO ₂	4-Cl	4-Cl	410
1,33	C(CH ₃) ₂	SO ₂	4-Cl	4-Cl	76
1,9	C:CH ₂	SO ₂	4-OCF ₃	2-F	86
1,34	C:NOCH ₃	SO ₂	4-Cl	4-Cl	406
1,35	SO ₂	CH ₂	4-Cl	4-Cl	164
1,36	SO ₂	CO	4-Cl	4-Cl	192
1,15	SO ₂	C:CH ₂	4-Cl	2-F	247
1,37	SO ₂	C(CH ₃)(OH)	4-Cl	2-F	77
1,38	SO ₂	C(CH ₃)(OH)	4-Cl	4-Cl	230
1,18	SO ₂	NH	4-Cl	4-Cl	278
1,19	SO ₂	O	4-Cl	4-Cl	983
1,1	SO ₂	SO ₂	4-OCH ₃	4-OMe	0.4
1,39	SO ₂	SO ₂	4-OCH ₃	H	0.6
1,40	SO ₂	SO ₂	4-OCH ₃	4-Cl	0.9
1,41	SO ₂	SO ₂	4-OH	4-Cl	13
1,42	SO ₂	SO ₂	4-O- <i>o</i> -C ₆ H ₅	4-Cl	232
1,43	SO ₂	SO ₂	4-CF ₃	4-Cl	8
1,44	SO ₂	SO ₂	H	4-Cl	58
1,45	SO ₂	SO ₂	4-CH ₃	2-F	0.5
1,46	SO ₂	SO ₂	4-CF ₃	2-F	0.9
1,47	SO ₂	SO ₂	H	2-F	9
1,48	SO ₂	SO ₂	4-Cl	H	2
1,49	SO ₂	SO ₂	4-Cl	2-Cl	6
1,50	SO ₂	SO ₂	4-Cl	3-Cl	23
1,51	SO ₂	SO ₂	4-Cl	4-Cl	10
1,52	SO ₂	SO ₂	4-Cl	2-F	1
1,53	SO ₂	SO ₂	4-Cl	2-OCF ₃	128

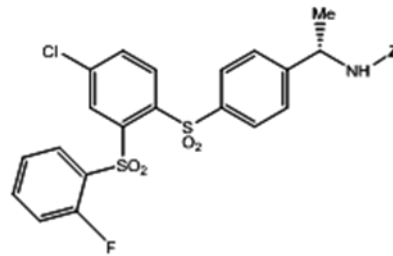
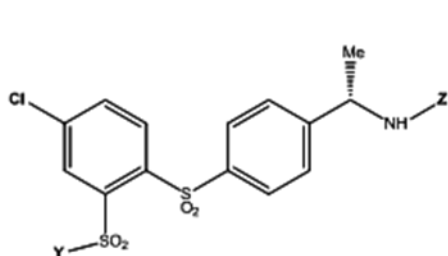
Note: A combination of Table 2-2, Table 2-3 and Table 2-4 represents all 85 bis-sulfone compounds used in this work. The unit for K_i values is nM.

Table 2-3. Bis-sulfone analogs from Lavey *et al.* [153] used for FS-QSAR modeling



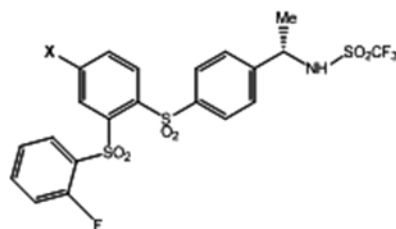
Compound	Z	Ki
2,2	(C=O)-Trifluoromethyl	0.3
2,4a	(C=O)-Methyl	7.7
2,4b	(C=O)-Ethyl	29
2,4c	(C=O)-Propyl	6.2
2,4d	(C=O)-Cyclopropyl	15.7
2,4e	(C=O)- <i>t</i> -Butyl	38.6
2,4f	(C=O)-Benzyl	412
2,4g	(C=O)- <i>p</i> -Toluy	289
2,4i	(C=O)-3,4-Dichlorophenyl	167
2,4j	(SO ₂)-Methyl	0.4
2,4k	(SO ₂)-Ethyl	1
2,4l	(SO ₂)-Butyl	41
2,4m	(SO ₂)-Phenyl	239
2,4n	(SO ₂)- <i>p</i> -Toluy	495
2,4p	(SO ₂)-Benzyl	160
2,4q	(CONH)-Propyl	495
2,4r	(CONH)- <i>p</i> -Toluy	802

Table 2-4. Bis-sulfone analogs from Lavey *et al.* [154] used for FS-QSAR modeling



Compound	Y	Z	Ki
3,15	CH ₂ CH ₂ CH ₃	NHSO ₂ CH ₃	308
3,16		NHSO ₂ CH ₃	2.9
3,18		NHSO ₂ CH ₃	129
3,19		NHSO ₂ CH ₃	204
3,20		NHSO ₂ CH ₃	8.9
3,42		NHSO ₂ CF ₃	3.6
3,11	2,6-Difluorophenyl	NHSO ₂ CF ₃	5.8
3,43		NHSO ₂ CF ₃	21
3,44		NHSO ₂ CF ₃	4.1
3,45		NHSO ₂ CF ₃	22
3,46		NHSO ₂ CF ₃	14.4
3,47		NHSO ₂ CF ₃	115
3,48		NHSO ₂ CF ₃	140
3,49	NH- <i>t</i> -Bu	NHSO ₂ CF ₃	136
3,50	<i>N</i> -Piperidine	NHSO ₂ CF ₃	68
3,51	<i>N</i> -Morpholine	NHSO ₂ CF ₃	192

Compound	Z	Ki
3,24	NHSO ₂ CF ₃	2
3,25	NHCOCH ₃	60.7
3,26	NHCOCF ₃	3.7
3,29	NCH ₂ CF ₃	7
3,30	NCH ₂ CH ₂ CF ₃	17



Compound	X	Ki
3,31	H	25
3,32	OH	1.3
3,33	OCF ₃	5
3,34	OCH ₂ CH ₃	0.9
3,35	OCH ₂ CH ₂ OCH ₃	2.5
3,36	CF ₃	2.5
3,37	CH ₃	1.1
3,38	Cyclopropyl	1.4
3,39	NH ₂	10
3,40	NH-cyclopropyl	49

Table 2-5. A list of COX2 analogs for 5 different testing sets

Testing Set	Compound	Testing Set	Compound
1	3,6	3	3,37
1	3,12	3	3,47
1	3,20	3	3,51
1	3,34	3	3,69
1	3,40	3	3,74
1	3,57	3	3,75
1	3,60	3	3,78
1	3,68		
1	3,69	4	3,9
1	3,70	4	3,10
1	3,76	4	3,19
1	3,90	4	3,32
1	3,97	4	3,34
1	3,99	4	3,36
		4	3,44
2	3,10	4	3,45
2	3,12	4	3,46
2	3,14	4	3,57
2	3,23	4	3,63
2	3,27	4	3,66
2	3,28	4	3,67
2	3,32	4	3,78
2	3,38		
2	3,40	5	3,5
2	3,42	5	3,14
2	3,59	5	3,19
2	3,66	5	3,36
2	3,74	5	3,40
2	3,91	5	3,42
		5	3,48
3	3,9	5	3,58
3	3,10	5	3,59
3	3,11	5	3,61
3	3,12	5	3,70
3	3,15	5	3,77
3	3,24	5	3,80
3	3,32	5	3,90

Table 2-6. A list of bis-sulfone analogs for 5 different testing sets

Testing Set	Compound	Testing Set	Compound
1	1,29	3	2,4p
1	1,15	3	3,18
1	2,4b	3	3,20
1	2,4k	3	3,36
1	2,4p		
1	3,26	4	1,33
1	3,29	4	1,15
1	3,42	4	1,46
		4	1,52
2	1,24	4	2,4b
2	1,31	4	3,24
2	1,19	4	3,33
2	1,39	4	3,43
2	2,4i		
2	3,29	5	1,32
2	3,40	5	1,18
2	3,50	5	1,41
		5	1,43
3	1,1	5	1,47
3	1,44	5	2,4L
3	2,2	5	2,4p
3	2,4n	5	3,39

Table 2-7. Results of leave-one-out cross-validation (LOOCV) on COX2 analogs

Training Set	Q ²	Scaling Factor	<i>L</i> or <i>H</i>
1	0.60	0.05	<i>L</i>
2	0.63	0.66	<i>L</i>
3	0.63	0.69	<i>L</i>
4	0.58	0.71	<i>L</i>
5	0.60	0.06	<i>L</i>

q² = leave-one-out cross-validated correlation coefficient; *L* or *H* = *L*owest or *H*ighest eigen values used.

Table 2-8. Results of leave-one-out cross-validation (LOOCV) on bis-sulfone analogs

Training Set	Q ²	Scaling Factor	<i>L</i> or <i>H</i>
1	0.64	0.36	<i>H</i>
2	0.49	0.36	<i>H</i>
3	0.63	0.38	<i>L</i>
4	0.58	0.36	<i>H</i>
5	0.65	0.06	<i>L</i>

Table 2-9. Result summary on five COX2 testing sets

Testing Set	Correlation coefficients (R ²)			
	Free-Wilson QSAR	FS-QSAR (BCUT)	FS-QSAR (Tanimoto)	1NN-QSAR
1	0.70	0.69	0.69	0.38
2	0.39	0.56	0.53	0.35
3	0.07	0.53	0.46	0.18
4	0.45	0.69	0.76	0.41
5	0.69	0.65	0.65	0.52
average R ²	0.46	0.62	0.62	0.37
standard deviation	0.26	0.08	0.12	0.12
variance	0.07	0.01	0.01	0.02

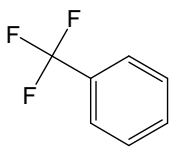
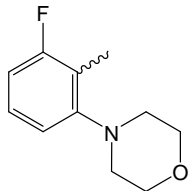
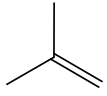

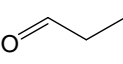
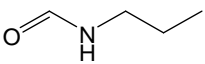
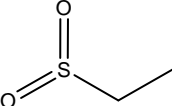
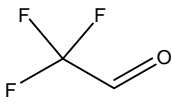
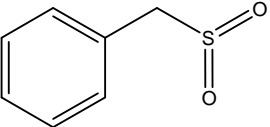
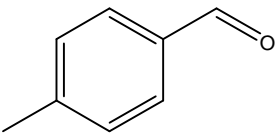
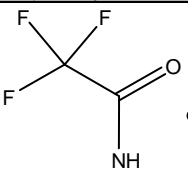
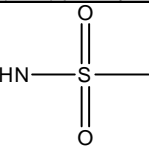
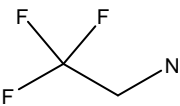
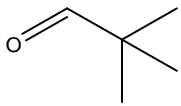
FS-QSAR (BCUT) = FS-QSAR using BCUT similarity function; FS-QSAR (Tanimoto) = FS-QSAR using Tanimoto similarity function; 1NN-QSAR = 1-nearest-neighbor QSAR

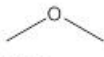

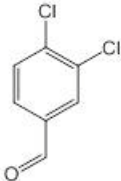
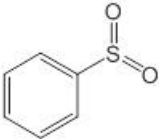
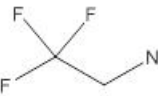
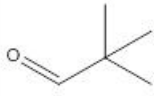
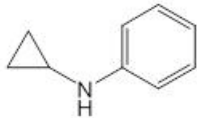
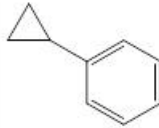


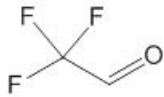
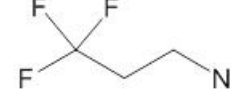
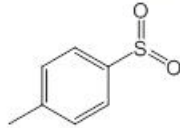
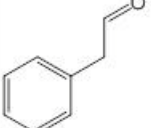
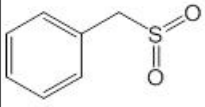
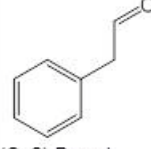
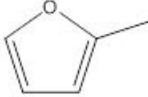
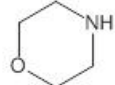


Table 2-10. Result summary on five bis-sulfone testing sets

Testing Set	Correlation coefficients (R^2)			
	Free-Wilson QSAR	FS-QSAR (BCUT)	FS-QSAR (Tanimoto)	1NN-QSAR
1	0.00	0.55	0.14	0.02
2	0.82	0.77	0.87	0.39
3	0.02	0.78	0.64	0.14
4	0.67	0.56	0.63	0.38
5	0.60	0.74	0.82	0.14
average R^2	0.42	0.68	0.62	0.21
standard deviation	0.39	0.12	0.29	0.16
variance	0.15	0.01	0.08	0.03

Table 2-11. Fragment pairs and their BCUT-similarity scores for each bis-sulfone testing set.

BCUT scores were calculated using eigen values that were derived based on the parameters listed in the table. SF = scaling factor; *H* or *L* = Highest or Lowest eigen values used.

Testing compound	Set	Testing fragment not found in training set	Chosen training fragment	Substituent position	Parameters (H or L)_SF	BCUT Sim Score
1,29	1			Y	H_0.36	0.9962
1,15	1			L2	H_0.36	0.9286
2,4b	1			Z	H_0.36	0.9964
2,4k	1			Z	H_0.36	0.9921
2,4p	1			Z	H_0.36	0.9994
3,26	1			Z	H_0.36	0.9931
3,29	1			Z	H_0.36	0.9862

1,19	2			L2	H_0.36	0.8754
2,4i	2			Z	H_0.36	0.9823
3,29	2			Z	H_0.36	0.9862
3,40	2			X	H_0.36	0.9975
3,50	2			Y	H_0.36	0.9833
2,2	3			Z	L_0.375	0.9839
2,4n	3			Z	L_0.375	0.9895
2,4p	3			Z	L_0.375	0.9953
3,18	3			Y	L_0.375	0.9685
3,20	3			Y	L_0.375	0.9951

1.33	4			L1	H_0.36	0.8919
1.15	4			L2	H_0.36	0.9286
2.4b	4			Z	H_0.36	0.9964
3.43	4			Y	H_0.36	0.9872
1.18	5			L2	H_0.36	0.8754
2.4L	5			Z	H_0.36	0.9914
2.4p	5			Z	H_0.36	0.9994
3.39	5			X	H_0.36	0.9975

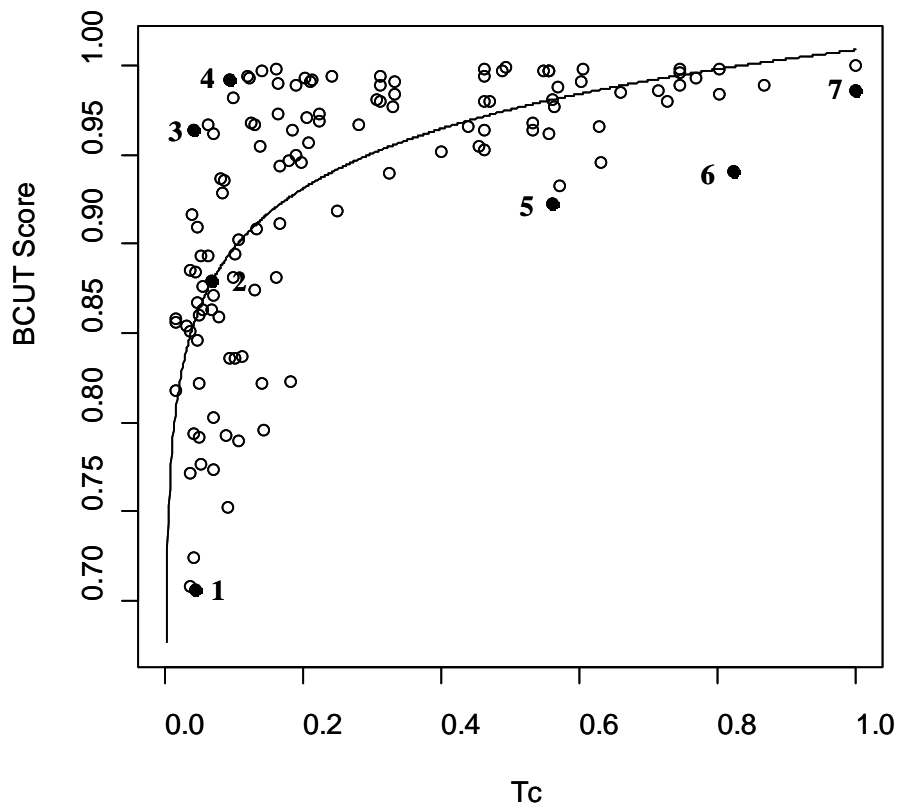
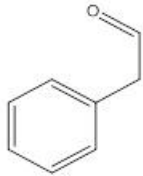
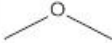
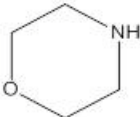
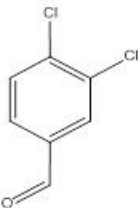
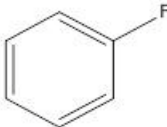
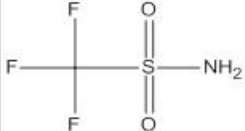
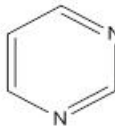
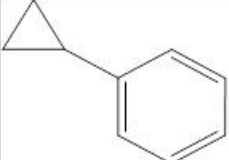
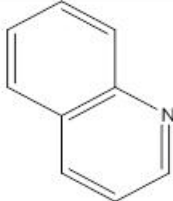
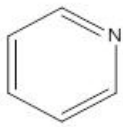


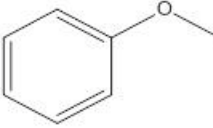
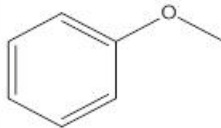


Figure 2-1. A plot of Tanimoto (Tc) vs. BCUT similarity scores of random fragment pairs from bis-sulfone analogs. The labeled points are the selected fragment pairs listed in Table 2-12.

Table 2-12. Fragment pairs chosen in Figure 2-1 and their corresponding similarity scores

ID	Fragment 1	Fragment 2	Tc Score	BCUT Score
1			0.044	0.706
2			0.068	0.879
3			0.041	0.964
4			0.092	0.992
5			0.562	0.922
6			0.824	0.941
7			1	0.986

3.0 NEW QSAR PREDICTION MODELS DERIVED FROM GPCR CB2- ANTAGONISTIC TRIARYL BIS-SULFONE ANALOGS BY A COMBINED MOLECULAR MORPHOLOGICAL AND PHARMACOPHORIC APPROACH

3.1 INTRODUCTION

Among several ligand-based virtual screening technologies as discussed in Section 1.2.2, pharmacophore modeling has been used extensively in drug discovery research to search for novel lead compounds in both academic and pharmaceutical industry environments [166, 182]. Common pharmacophore features derived from known ligands are often used to search compound databases such as PubChem database [183, 184] in order to identify molecules with similar molecular features as potential bioactive ligands. On the other hand, such pharmacophore-based methods do not often consider molecular morphological or shape characteristics of a bioactive ligand which is also an important property for a ligand to bind to an active site of a receptor. In fact, both pharmacophoric feature compatibility and molecular surface complementarity between a ligand and its receptor would influence its binding to the receptor. In other words, in-silico screened hits based solely on a pharmacophoric model may or may not be bioactive since recognition of small molecules by a protein receptor is also mediated by molecular surface complementarity [82]. In fact, molecular shape similarity search approach is emerging as a valuable ligand-based virtual screening strategy. However, there is still not a

good score function that has been developed by combining both pharmacophoric features and molecular surface similarity to evaluate and rank virtual screening hits from ligand-based drug design procedures.

In this work, a novel QSAR ranking score function was developed to exploit the combination of both pharmacophoric features and molecular shape similarity in order to predict bioactivities of hits from a ligand-based virtual screening process. A set of triaryl bis-sulfonal derivatives regarded as CB2 antagonists was used to generate the QSAR model employing pharmacophore and shape-based descriptors as well as other calculated molecular properties. In our studies, four bioactive triaryl bis-sulfones were first applied to produce a 3D pharmacophore model using Sybyl/GALAHAD program [82]. According to the CB2 antagonistic SAR results [153-155] of bioactive triaryl bis-sulfonal derivatives, the simulated pharmacophore model was then modified to eliminate unnecessary pharmacophore features. The refined pharmacophore query was then used for virtual screening of a compound database containing 45 triaryl bis-sulfonal derivatives to determine their pharmacophoric similarity using the Sybyl/UNITY module [96]. Furthermore, the same four triaryl bis-sulfones were employed to build up hypermolecular superimposition hypothesis by using a shape-based molecular alignment algorithm. The generated hypothesis was regarded as a structural template to measure the molecular shape similarity of 45 triaryl bis-sulfonal compounds using the Surfex-Sim program [185]. The whole database of 45 triaryl bis-sulfonal compounds was then randomly split into two datasets, a training set of 25 compounds and a test set of 20 compounds. By combining pharmacophore-based molecular similarity, shape-based molecular similarity, and calculated molecular properties (e.g., molecular weight, ClogP, molecular volume, molecular polar area as descriptors), a QSAR model was generated with partial least squares (*PLS*) regression analysis

using the training set of compounds. The developed pharmacophore and shape-based (or Pharmshape) QSAR score function allowed us to predict the K_i values of 20 triaryl bis-sulfones in the test dataset. Our results suggested that the established Pharmshape computational algorithm which generates QSAR prediction models by combining pharmacophore-based molecular similarity and shape-based molecular similarity can be used to predict the biological properties of untested compounds or virtual hits obtained from ligand-based virtual screening studies.

3.2 METHODS

3.2.1 Pharmacophore-based molecular similarity calculation

In our studies, four highly CB2 selective triaryl bis-sulfones, namely compounds 2, 5, 8, and 11 in Table 3-1, were first selected to develop pharmacophoric queries for this class of CB2 antagonist using Sybyl/GALAHAD program [82]. They were aligned based on their pharmacophoric features to produce pharmacophore hypotheses. The default definitions included six types of features: H-bond donor and acceptor atoms, positive nitrogen, negative and hydrophobic centers, and steric features. The alignment process was divided into two stages. The selected ligands were first aligned to each other in an internal coordinate space with an advanced genetic algorithm (GA) to allow ligand flexibility and the generation of a molecular alignment was then based on the superimposition of their pharmacophoric features in the GA-generated conformations of ligands in the Cartesian space. In order to get rational hypermolecular alignments, the population size and max generations for the GA simulation were

set up to 200 and 15000, respectively. Other default parameters for genetic operators were *Chromosome-Torsion* options and *Chromosome-Conformer* options such as Mutation Rate (0.4, 0.2), Mutation Decay (1.0, 1.0), and Crossover Rate (1.0, 1.0) in the Tripos/GALAHAD program [82]. The GALAHAD scoring terms include H-bonding, Sterics, and Energy. After GALAHAD simulations, the generated pharmacophore model was modified to delete the unnecessary pharmacophoric features and to retain the pharmacophoric features required for their bioactivities based on the SAR studies of triaryl bis-sulfones as CB2 antagonists [153-155]. Following the simulation and modification of the pharmacophore model for triaryl bis-sulfones, the model was used as a query to search a compound training dataset composed of 45 bioactive triaryl bis-sulfones listed in Table 3-1 and Table 3-2 via the Tripos/UNITY program. A QFIT value was computed for each screened compound hit to rank the matching rate of its required structural features on the pharmacophoric query.

3.2.2 Morphology-based molecular similarity calculation

In our studies, the Surfex-Sim program [185] was utilized to conduct the shape-based molecular similarity comparison among the triaryl bis-sulfones as CB2 antagonists. Surfex-Sim [185] is one of the shape-based molecular similarity computational methods. It utilizes a morphological similarity function and a fast pose generation technique to make putative alignments of molecules or molecular fragments to other molecules. The same four triaryl bis-sulfones (compounds 2, 5, 8, and 11 in Table 3-1) used in the pharmacophore-based modeling were applied to develop the hypermolecular superimposition/alignment hypotheses based on their morphological similarity. This was achieved by the *Surfex-Sim Mutual Alignment* mode in the Surfex-Sim program. The default parameters were selected with a max of 20 conformations per

fragment, additional starting conformations of 0 per molecule, and a maximum number of 100 rotatable bonds per molecule to ensure valid geometries of four triaryl bis-sulfones to the hypothesis generated during the calculation. After simulation, the generated superimposition hypotheses were sorted by scores ranging from 0 to 1 with a higher score indicating a better alignment. The best-scored hypothesis was chosen to serve as a template in the subsequent calculation for the measurement of the shape-based molecular similarity. After generating the shape alignment hypothesis as a template, it was then applied to perform a molecular shape similarity comparison to all 45 triaryl bis-sulfones (Table 3-1 and Table 3-2) using the *Surflex-Sim Flexible Superposition* mode in the program with the same default parameters as stated before. Such shape-based alignment simulations resulted in molecular poses and corresponding similarity scores of all bis-sulfone analogs to the alignment hypothesis generated based on four highly bioactive bis-sulfones.

3.2.3 QSAR model generation

QSAR analyses are often used to derive relationships between the molecular properties of compounds and their biological activities by statistical methods [186]. In order to develop a reliable QSAR model for triaryl bis-sulfones as CB2 antagonists, additional molecular properties, including molecular weight, ClogP, molecular volume, molecular polar volume, molecular polar surface area, and molecular area, were calculated using Tripos/Sybyl [82]. Besides molecular weight and ClogP, other molecular properties were defined as follows. The molecular polar surface area [187] was computed as a solvent accessible surface to all O, N, and S as well as protons covalently bonded to these atoms, essentially a van der Waals (vdWs) surface with inflated radii in a compound. Molecular area is defined as the total surface area of a

compound. Similarly, molecular polar volume was computed as the polar volume corresponding to all O, N, and S atoms as well as hydrogen atoms covalently bonded to these atoms in a compound. Molecular volume is defined as the total volume of a compound. These molecular properties as well as the dipole moment were calculated for all 45 triaryl bis-sulfones based on their conformations that resulted from the previous shape-based simulation. After the molecular properties of 45 triaryl bis-sulfones were calculated, the PLS regression analyses were performed to derive QSAR regression models between the CB2 receptor binding affinities (K_i) and molecular descriptors including query fit (QFIT) values (from pharmacophore-based modeling), the molecular morphology similarity scores (from shape-based modeling), and the calculated molecular properties using a training database of 25 triaryl bis-sulfones (Table 3-1). In particular, with different combinations of molecular properties, QFIT, and morphological similarity score, the *PLS* analyses were performed with a “Leave-One-Out” cross-validation to get an optimal number of components. Then, the calculated component value was used for the *PLS* analysis with no-validation to get the regression equation representing the relationship between pK_i ($-\log K_i$), and molecular descriptors including molecular properties, QFIT, and similarity score. In the mean time, r^2 and standard error of the estimates were determined to characterize the predictive ability of the generated Pharmshape-QSAR model. The generated QSAR model was further validated using a test set of 20 triaryl bis-sulfones (Table 3-2).

3.3 RESULTS AND DISCUSSION

3.3.1 Generation of the pharmacophore model and score of the pharmacophoric match

Figure 3-1 (Left) shows four representative computer-generated hypermolecular alignments (templates) and their corresponding pharmacophore queries generated from four triaryl bis-sulfones, including compounds 2, 5, 8, and 11. A pharmacophore is an ensemble of steric and electronic features that are necessary for a ligand to have for optimal molecular interactions with a specific biological target and to trigger (or block) its biological response [188]. In principle, a pharmacophore model could be simulated by deriving common essential structure characteristics responsible for their bioactivities based on molecular alignment of a set of known bioactive ligands. The generated pharmacophore model could then be applied to virtually screen a compound database for chemically-diverse molecules that share similar structural features and their relative spatial arrangement defined in the pharmacophore model.

Other important results are also given in Figure 3-1 (Left) for each GALAHAD-generated pharmacophore model. For example, the ENERGY term of computer-generated pharmacophore model A is 6.78 kcal/mol, which designates the total energy (using the Tripos force field) of all four molecules in their respective conformations encoded in the torsional chromosome. Meanwhile, the values of STERICS, HBOND, and MOL_QRY were computed to be 5052.30, 277.10 and 200.25 respectively in the model A. In the GALAHAD algorithm, STERICS is defined as the overall steric similarity among ligand conformers, HBOND as the overall pharmacophoric similarity among ligand conformers, and MOL_QRY as the agreement between the query tuplet and the pharmacophoric tuplets of target ligands as a group [96]. In general, a good pharmacophore model should have a maximized steric consensus, maximized

pharmacophore consensus, and minimized energy. As shown in Figure 3-1 (Left), the pharmacophore model A had the highest values of STERICS and MOL_QRY, the second highest value of HBOND, and the lowest value of ENERGY in the comparison with the other three models. Thus, the model A was selected as the final pharmacophore model for triaryl bis-sulfones.

As illustrated in Figure 3-1 (A), the GALAHAD-generated pharmacophore model A consists of five H-bond acceptor features (AA_1, AA_2, AA_5, AA_6, AA_7), an H-bond donor feature (DA_1), and three hydrophobic centers (HY_8, HY_9, HY_10). Obviously, an active pharmacophore model of a typical small molecule does not require so many hypothetical pharmacophore features, which is reflected by a low *Specificity* value of 4.32 for the model A. The GALAHAD-calculated term *Specificity* [82] is a logarithmic indicator of an expected discrimination for each query based on its numbers of pharmacophore features, their allotment across any partial match constraints, and the degree to which the features are separated in space. Since the *Specificity* value of the model A is below 5, the query could not be directly used for a UNITY flexible search and it was needed to be manually modified. It was also clear that not every pharmacophoric feature in the model A was necessary or should be weighted equally for their bioactivities according to the published SAR studies [153-155] of triaryl bis-sulfones as CB2 antagonists. For example, some compounds, such as compounds 1 to 5 in Table 3-1, with a methylene-linkage but no sulfone-linkage between rings A and B show high CB2 binding activity. Thus there is no H-bond structural feature necessary in the corresponding position. Another thing to consider is that each oxygen atom in the sulfone group could be treated as an H-bond acceptor feature in the GALAHAD results. As a result, there are two closed H-bond acceptor features in the L1 and L2 (AA_5 and AA_6) positions or in the Z (AA_1 and AA_2)

position (Figure 3-1 (A) and Table 3-1). There are several bioactive compounds in Table 3-1 and Table 3-2 with only a carbonyl group in either L1 or Z position. Therefore, only one H-bond acceptor feature is enough in either L1 or Z site for this class of CB2 antagonist ligands. There is no biological data to indicate that the AA_7 feature on the 4- substituent position of B-ring is required for the bioactivity of ligands binding to the CB2 receptor. Previous pharmacological data show that the aromatic ring is necessary only at the ring B position but not for the rings A and C positions to be bioactive as CB2 antagonists. As a result, the hydrophobic center HY_9 was re-defined as an aromatic center but HY_8 and HY_10 were kept as hydrophobic centers. Based on the SAR discussion above, the pharmacophore model was modified accordingly to a query consisting of two hydrophobic centers, an aromatic center, two H-bond acceptors, and an H-bond donor as shown in Figure 3-2 (Left). This refined pharmacophore model was used as a hypothetical query in the next UNITY database screening.

After the 3D flexible UNITY database search using the pharmacophore query defined in Figure 3-2 (Left), all 45 ligands in both training set and test set were returned as “hits”. UNITY uses a query fit (QFIT) as a score to measure the feature match value of a hit to the pharmacophore model. The QFIT value for each ligand was summarized in Table 3-3 and Table 3-4. The QFIT score is a value between 0 and 100, where 100 is the best and represents how close the ligand’s features match the query’s. The QFIT value could be regarded as a measure of the degree of closeness that the hit pharmacophore matches the corresponding query features coordinates within the certain range of a spatial constraint tolerance. The closer the fit is, the higher the value is. However, the later statistical results indicated that the QFIT values of ligands did not have a strong correlation with their K_i values. In fact, this is one of the reasons to

develop a QSAR model by not only considering pharmacophore features but also other morphological properties.

3.3.2 Generation of hypermolecular alignment and scoring of shape-based molecular similarity

In the current study, the Surfex-Sim program [185] was applied to perform molecular shape-based virtual screening simulation using triaryl bis-sulfone analogs. By using molecular fragmentation to address molecular flexibility and Gaussian-like function to score molecular shape similarity, Surfex-Sim generates putative alignments of molecules or molecular fragments to other molecules. Figure 3-1 (Right) illustrates four hypermolecular alignments generated from 4 highly bioactive bis-sulfones by the Surfex-Sim program. One of them will be used as a template for the calculation of shape-based molecular similarity.

Pharmacophore-based design methods seek to identify important molecular feature points, assumed to directly interact with proteins, within a range of tolerances among them. However, molecular surface complementarity, which is not generally considered in the pharmacophore-based or fingerprint-based methods, plays a critical role in the recognition of small molecules by their receptors. As described above, a pharmacophore-based virtual screening procedure is based on a structural template with a spatial arrangement of important molecular interaction features. Such a method does not consider the molecular volume of a bioactive ligand, which is also a critical property of a ligand binding to its receptor. Therefore, pharmacophore consistency alone does not mean that the screened hits will have a similar binding capability at a protein as a true bioactive ligand. This is because molecular structure (or morphological) inconsistency could block a small molecule fitting into a binding pocket of a

protein. Thus, shape consistency should also be considered as a necessary condition in ligand-based virtual screening. In fact, such molecular morphological features have become increasingly important factors in order to improve virtual screening effectiveness during the identification of high-quality leads [189-191].

In our shape-based simulation of four triaryl bis-sulfones, 100 hypotheses, with their shape similarity scores, were generated. As shown in Figure 3-1 (Right), molecular alignments were first performed for four representative superimposition triaryl bis-sulfones models with varied surface similarity scores. The data shows that the generated four hypermolecular alignments have the Surfex-Sim scores in a range of 0.468 to 0.482. The highest score alignment (Hypothesis_30, model III) in Figure 3-1 (Right) was then chosen as a morphological template to score shape-based similarity for all 45 triaryl bis-sulfones. The result also indicated that the shape hypothesis was quite similar to the molecular alignment models produced by pharmacophore based method when the pharmacophore model A (Figure 3-1 (A) or Figure 3-2 (Left)) and the shape-based model III (Figure 3-1 (III)) were superimposed. Figure 3-2 (Right) illustrates the MOLCAD-generated graphical representation of a combination of Surfex-Sim shape hypothesis and the GALAHAD pharmacophoric query. Such a congruent result provided us a foundation for further QSAR model calculation by the combined pharmacophoric and morphological approaches.

Unlike pharmacophoric feature modeling using a pharmacophore-based method, Gaussian-like morphological similarity function [185] makes the molecular alignment procedure much more complicated and slower than the pharmacophore-based alignment since it is dependent on surface shape and atomic charge characteristics of the ligands. Our Surfex-Sim simulation studies indicated that Surfex-Sim had a slow screening rate due to the intrinsic

properties of the Surfex-Sim search and match algorithm [185]. Such a slow screening speed would not be efficient enough for computer-aided virtual screening of a huge compound database like PubChem compound database [192] containing more than 15 million compounds. Therefore, the Surfex-Sim method was used to screen a focused library that has a limited number of compounds from a pharmacophore-based virtual screening. Such a secondary search filtered out molecules with incompatible shapes.

Therefore, based on the selected optimal hypothetical alignment that has the highest morphological similarity score as the template, flexible superimposition searches were then carried out to calculate shape similarity score for each compound in Table 3-1 and Table 3-2. Calculated shape-based molecular similarity scores (Surflex-Sim Search score, SSS) were summarized in Table 3-3 and Table 3-4. The later statistical analyses indicated that the shape-based molecular similarity has a much better relationship with molecular biological activity than QFIT. Our results also revealed that when the shape-based molecular similarity was applied to rank the screened hits from the pharmacophore-based virtual screening, it improved the hit rate of ligand-based virtual screening in the identification of potential bioactive leads as discussed below.

3.3.3 Development of the PharmShape algorithm based on the QSAR prediction model

In addition to the QFIT value from a pharmacophore-based query search and the SSS (morphological similarity score) values from Surfex-Sim search, other calculated molecular properties, including molecular weight, ClogP, molecular volume, molecular polar volume, molecular area, molecular polar surface area, and dipole moment for all of 45 triaryl bis-sulfones are summarized in Table 3-3 and Table 3-4. By incorporating these 2D/3D molecular

descriptors, the QSAR prediction models were developed to predict biological activity (K_i) values. This was done by partial least squares (*PLS*) regression using a training set of 25 triaryl bis-sulfones listed in Table 3-3. The *PLS* analyses were performed with the leave-one-out cross validation (LOOCV) to get an optimal number of components with different combination of molecular properties, QFIT and morphological similarity score. Then, with the calculated component value, non-cross-validated *PLS* analysis with “No Validation” was performed to get a regression equation representing the relationship between K_i and the defined molecular descriptors. Cross-validated r_{cv}^2 , non cross-validated r^2 , and standard error of estimates were determined to characterize the predictive ability of the generated QSAR model. They are defined [82] as follows:

$$\text{Cross-validated } r_{cv}^2 = \frac{SD - PRESS}{SD} \quad (3.1)$$

$$\text{Non cross-validated } r^2 = 1.0 - \frac{\sum_i (pK_{i(pred)} - pK_{i(exp)})^2}{\sum_i (pK_{i(exp)} - pK_{i(mean)})^2} \quad (3.2)$$

$$\text{Standard error of estimate } s = \sqrt{\frac{PRESS}{N - n - 1}} \quad (3.3)$$

where SD is the sum of the squared deviations of each biological property value from their mean; PRESS (PRedictive Error Sum of Squares) is the sum, over all compounds, of the squared differences between the actual and “predicted” biological property values; N is number of

descriptors; n is number of components; $pK_{i(\text{pred})}$ is the calculated $-\log K_i$; $pK_{i(\text{expt})}$ is experimental $-\log K_i$; $pK_{i(\text{mean})}$ is the best estimate of the mean of all values that might be predicted.

Table 3-5 summarizes the *PLS* analysis results of the generated quantitative relationships between the bioactivity pK_i and different combinations of pharmacophore QFIT value, morphological similarity SSS score, and calculated molecular properties. The low r^2 value (0.05) of the model 1 indicates the QFIT value could not be directly applied to judge whether the hit is a good lead or if it should be selected for further experimental bioassay testing. A better r^2 value (0.77) of model 2 suggests that morphological similarity score (SSS) has much better relationship with pK_i than QFIT. In addition, statistical analysis with the combination of SSS and QFIT values (the model 3 in Table 3-5) did not significantly improve the relationship between pK_i and two score values, showing r^2 value of 0.78. Furthermore, as shown from models 3 to 13 listed in Table 3-5, incorporating calculated molecular properties into statistical analyses significantly improved the r_{cv}^2 value up to 0.84, in particular the model 4 showing a good relationship between $-\log K_i$ and the combination of pharmacophore-based and shape-based molecular similarity, as well as molecular area and molecular polar area descriptors.

QSAR models 14 to 27 represent additional different combinations of the calculated descriptors for *PLS* analyses, showing no improvement of r_{cv}^2 value in comparison with that of model 4. *PLS* analysis without QFIT value produced QSAR models 28 and 30 with the low r^2 values of 0.77 and 0.78, respectively, suggesting a decreased relationship between pK_i and the combination of Surflex-Sim score and calculated molecular properties. In addition, no good QSAR relationships were achieved by using only the calculated molecular properties descriptors without shape-based and pharmacophore-based descriptors. For example, r_{cv}^2 was 0.36 when *PLS* analysis was performed to generate the relationship between $-\log K_i$ and molecular properties of

molecular weight, ClogP, molecular area, molecular volume, molecular polar area, molecular polar volume, and Dipole_Moment. Our results demonstrated that both pharmacophore-based molecular similarity (QFit) and shape-based molecular similarity (SSS) played important roles in generating QSAR models.

Thus, it was concluded that the highest r_{cv}^2 and r^2 values (Model 4 in Table 3-5) were obtained when *PLS* analysis was used to generate the relationship between bioactivity $-\log K_i$ and the combination of pharmacophore similarity, shape similarity, molecular surface area, and molecular polar surface area. Such an analysis gave the following regression equation for the generated QSAR prediction model:

$$-\log K_i = -10.675 + 22.595 * SSS + 0.016 * QFIT - 0.005 * MolSA + 0.014 * MolPSA \quad (3.4)$$

where SSS is molecular surface similarity score obtained from Surfex-Sim similarity; QFIT is pharmacophore feature fit obtained from UNITY based on GALAHAD-generated pharmacophore query; MolSA is molecular surface area; and MolPSA is molecular polar surface area. The equation do not enclose other molecular properties, including molecular polar volume (MolPV), molecular volume (MolVol), ClogP, molecular weight (MW), and molecular Dipole_Moment (DM) descriptors as shown in Table 3-3, because these molecular descriptors decreased the value of r_{cv}^2 in *PLS* analyses as indicated in Table 3-5.

In order to evaluate the robustness of our generated Pharmshape QSAR model validated by the leave-one-out cross validation protocol, progressive scrambling method, which was developed by Clark [193, 194], was further carried out to examine our generated QSAR model that correlates molecular bioactivities of pKi with their SSS, QFIT, MolSA, MolPSA. With the

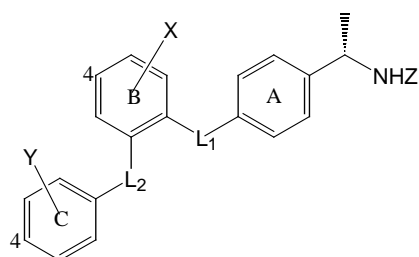
number of components set to 3, triple screen mode and other default parameters defined in the Scrambling Stability Test of QSAR module in Tripos Sybyl [82], the progressive scrambling produced the values of r_{yy}^2 , cSDEP (cross-validated standard error of prediction), and dq^2/dr_{yy}^2 as 0.61, 0.81, and 1.03, respectively. According to the literature and Sybyl QSAR manual [82, 193, 194], dq^2/dr_{yy}^2 is regarded as the critical statics in progressive scrambling. In general, QSAR model will be accepted as stable when dq^2/dr_{yy}^2 is less than 1.20, and stable models have the dq^2/dr_{yy}^2 values of near unity. Therefore, the progressive scrambling test demonstrated that our generated Pharmshape QSAR has reliable predictive capability.

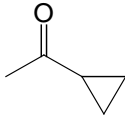
After the Pharmshape QSAR model was derived from the training set of compounds shown in Table 3-1, compounds listed in Table 3-2 were then used to test the model. The predicted pK_i values for the test set were summarized in Table 3-4. As shown in Figure 3-3, the linear relationship plotting between the calculated and measured K_i values of all of 45 compounds in both training set ($r_{cv}^2 = 0.84$, Table 3-5) and test set ($r^2=0.89$, Table 3-5) reveals a good correlation for the QSAR score function model for the binding affinities of triaryl bis-sulfones as CB2 antagonists. Our results confirmed that the QSAR model generated from the training dataset has a relatively good prediction capability of the receptor binding affinity for the testing dataset. Again, our studies concluded that both pharmacophore feature similarity and molecular shape similarity have positive influence to the prediction of the biological activity of a ligand.

3.4 CONCLUSION

In this chapter, we presented a successful work of generating QSAR score function model by combining pharmacophoric and morphological approaches. The developed QSAR prediction model for triaryl bis-sulfones as CB2 antagonist indicated that there was a good structure-activity relationship between their bioactivities and the combination of pharmacophore-based molecular similarity, shape-based molecular similarity, and calculated molecular properties. The *PLS* cross-validation (training set) r_{cv}^2 value of 0.84 and no-validation (test set) r^2 value of 0.89 conclude that the developed Pharmshape QSAR model has a relatively good capability to predict binding activities of hits obtained from ligand-based virtual screening based on triaryl bis-sulfone analogs. Our studies are congruent with the current research reports [189-191] demonstrating that the combination of the pharmacophore-based molecular similarity and the shape-based molecular similarity might be useful for ranking the virtual hits based on the predicted bioactivity values. The established Pharmshape QSAR score function algorithm, combining both the shape-based simulation and the pharmacophore matching comparison, provides a practical route to predict the CB2 bioactivity of hits from CB2 ligand-based virtual screening before bioassay testing.

Table 3-1. The structures and K_i values of the Triaryl bis-sulfones used as a training set to generate QSAR prediction model



Cmpd.	L1	L2	X	Y	Z	K_i (CB2) (nM)	CB1/CB2
1	CH ₂	SO ₂	4-OCH ₃	4-OCH ₃	-SO ₂ CH ₃	0.6	1300
2	CH ₂	SO ₂	4-Cl	4-Cl	-SO ₂ CH ₃	6.7	188
3	CH ₂	SO ₂	4-OCF ₃	2-F	-SO ₂ CH ₃	0.4	1178
4	CH ₂	SO ₂	4-CF ₃	2-F	-SO ₂ CH ₃	1.0	674
5	CH ₂	SO ₂	4-CF ₃	2,6-Di F	-SO ₂ CH ₃	1.8	584
6	SO ₂	CO	4-Cl	4-Cl	-SO ₂ CH ₃	192.0	284
7	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃	-SO ₂ CH ₃	0.4	2262
8	SO ₂	SO ₂	4-OCH ₃	-H	-SO ₂ CH ₃	0.6	2482
9	SO ₂	SO ₂	4-OCH ₃	4-Cl	-SO ₂ CH ₃	0.9	1146
10	SO ₂	SO ₂	4-OH	4-Cl	-SO ₂ CH ₃	0.3	783
11	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃	-COCF ₃	0.3	783
12	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃	-COCH ₃	7.7	231
13	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃	-COC ₂ H ₅	29.0	122
14	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃	-COC ₃ H ₇	6.2	189
15	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃		15.7	676

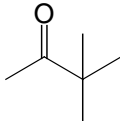
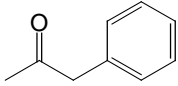
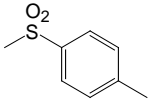
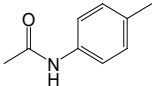
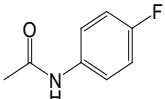
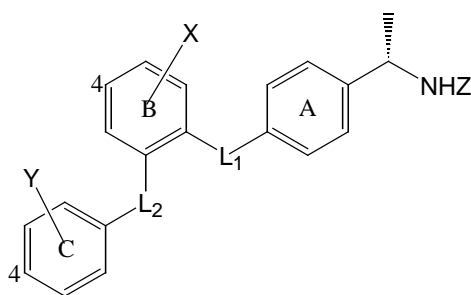
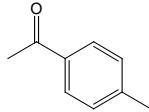
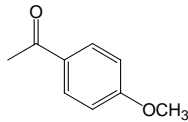
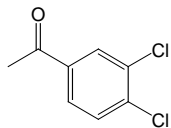
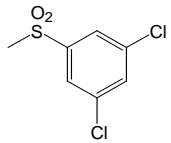
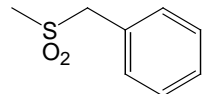
16	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃		38.6	7.41
17	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃		412.0	1.2
18	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃	-SO ₂ CH ₃	0.4	9.40
19	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃	-SO ₂ C ₂ H ₅	1.0	9.00
20	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃	-SO ₂ C ₄ H ₉	41.0	27
21	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃	-SO ₂ C ₆ H ₅	239.0	3.6
22	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃		495.0	3.9
23	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃	-CONHC ₃ H ₇	495.0	145
24	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃		802.0	125
25	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃		1136.0	15

Table 3-2. The structures and K_i values of the Triaryl bis-sulfones used to test the generated QSAR prediction model



Cmpd.	L1	L2	X	Y	Z	K_i (CB2) (nM)	CB1/CB2
26	CH2	SO ₂	4-OCF ₃	4-CF ₃	-SO ₂ CH ₃	35.0	27
27	CO	SO ₂	4-OCH ₃	4-OCH ₃	-SO ₂ CH ₃	44.0	108
28	CO	SO ₂	4-OCF ₃	2-F	-SO ₂ CH ₃	179.0	28
29	C(CH ₃) ₂	SO ₂	4-Cl	4-Cl	-SO ₂ CH ₃	76.0	48
30	SO ₂	SO ₂	4-CF ₃	4-Cl	-SO ₂ CH ₃	8.0	558
31	SO ₂	SO ₂	H	4-Cl	-SO ₂ CH ₃	58.0	127
32	SO ₂	SO ₂	4-CH ₃	2-F	-SO ₂ CH ₃	0.5	1741
33	SO ₂	SO ₂	4-CF ₃	2-F	-SO ₂ CH ₃	0.9	3552
34	SO ₂	SO ₂	H	2-F	-SO ₂ CH ₃	9.0	1449
35	SO ₂	SO ₂	4-Cl	H	-SO ₂ CH ₃	2.0	1941
36	SO ₂	SO ₂	4-Cl	2-Cl	-SO ₂ CH ₃	6.0	1778
37	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃		289.0	884
38	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃		1258.0	0.9
39	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃		167.0	0.5
40	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃		2476.0	0.05
41	SO ₂	SO ₂	4-OCH ₃	4-OCH ₃		160.0	0.5

42	SO ₂	SO ₂	4-Cl	3-Cl	-SO ₂ CH ₃	23.0	238
43	SO ₂	SO ₂	4-Cl	4-Cl	-SO ₂ CH ₃	10.0	687
44	SO ₂	SO ₂	4-Cl	2-F	-SO ₂ CH ₃	1.0	4387
45	SO ₂	SO ₂	4-Cl	2-OCF ₃	-SO ₂ CH ₃	128.0	134

Table 3-3. Pharmacophore-based molecular similarity score, shape-based molecular similarity score and the calculated molecular properties for the training set of triaryl bis-sulfones listed in Table 3-1.

Compd.	$-\log K_i$	MW	MolVol	SSS	QFIT	MolSA	ClogP	DM	MolPV	MolPSA	Pred. $-\log K_i$
1	9.22	489.61	1328.87	0.89	62.29	786.16	3.87	5.81	262.47	159.78	9.01
2	8.17	498.45	1232.71	0.88	80.50	736.43	5.07	4.71	241.64	136.82	8.91
3	9.40	531.54	1294.72	0.88	80.55	759.27	5.15	5.09	252.74	152.33	8.95
4	9.00	515.54	1252.52	0.88	80.45	736.56	4.76	5.59	227.30	137.68	9.01
5	8.74	533.53	1266.55	0.88	75.88	741.87	4.92	6.53	226.13	136.29	8.82
6	6.72	512.43	1214.70	0.82	27.85	722.65	4.36	6.44	220.11	148.93	6.94
7	9.40	539.65	1380.27	0.87	73.84	797.10	2.28	8.89	349.27	206.08	9.31
8	9.22	509.62	1277.83	0.88	79.36	745.86	2.31	9.44	322.73	197.61	9.79
9	9.05	544.06	1334.95	0.88	73.43	772.74	3.06	9.19	328.56	193.77	9.41
10	7.89	530.04	1266.15	0.76	75.42	738.84	2.98	8.40	337.57	237.27	7.56
11	9.52	557.56	1376.24	0.86	83.03	808.49	3.53	8.12	248.52	159.48	8.56
12	8.11	503.59	1297.88	0.86	79.98	783.38	2.42	5.56	234.15	161.89	8.50

13	7.54	517.62	1358.23	0.84	81.89	810.06	2.95	5.48	280.22	153.55	7.98
14	8.21	531.64	1399.92	0.83	80.37	847.23	3.48	5.69	243.94	150.94	7.38
15	7.80	529.63	1396.75	0.83	58.88	824.67	3.00	5.38	260.81	158.01	7.41
16	7.41	545.67	1469.80	0.83	79.30	853.53	3.65	6.31	266.06	139.41	7.28
17	6.39	579.69	1537.59	0.79	78.46	899.26	4.19	5.28	243.11	151.42	6.27
18	9.40	539.65	1374.74	0.87	74.51	797.91	2.28	6.84	347.91	207.09	9.33
19	9.00	553.67	1434.67	0.86	74.51	825.38	2.81	8.90	368.69	159.78	8.87
20	7.39	581.73	1535.77	0.82	77.04	889.46	3.86	8.77	396.90	136.82	7.62
21	6.62	601.72	1529.91	0.79	64.55	870.29	4.16	8.69	375.07	152.33	6.96
22	6.31	615.74	1574.43	0.78	64.54	900.86	4.66	8.79	371.28	137.68	6.52
23	6.31	546.66	1415.69	0.79	78.63	871.78	3.75	5.60	281.90	136.29	6.64
24	6.10	594.70	1539.06	0.77	78.22	921.42	4.84	5.60	236.48	148.93	6.02
25	5.94	598.67	1495.68	0.77	78.28	899.16	4.78	6.19	264.55	206.08	6.15

MW is molecular weight, MolVol is molecular volume, SSS is molecular surface similarity score obtained from Surflex-Sim similarity, QFIT is pharmacophore feature fit obtained from UNITY based on GALAHAD-generated pharmacophore query, CLogP is molecular octanol-water partition coefficient computed with Tripos/Sybyl8.0, MolPSA is molecular polar surface area, MolPV molecular polar volume MolSA is molecular surface area, and DM is molecular Dipole_Moment.

Table 3-4. Pharmacophore-based molecular similarity score, shape-based molecular similarity score and the calculated molecular properties for the testing set of triaryl bis-sulfones listed in Table 3-2.

Compd.	$-\log K_i$	MW	MolVol	SSS	QFIT	MolSA	ClogP	DM	MolPV	MolPSA	Pred. $-\log K_i$
26	7.46	581.55	1384.38	0.84	80.57	817.97	5.98	10.90	224.76	147.20	7.64
27	7.36	503.59	1336.87	0.88	57.58	791.13	2.95	3.48	277.68	181.80	8.65
28	6.75	545.53	1296.39	0.85	57.63	773.89	4.24	14.08	255.77	167.37	7.94
29	7.12	526.50	1285.34	0.87	80.94	751.14	5.87	10.62	226.99	120.56	8.16
30	8.10	582.04	1352.13	0.86	80.83	775.77	4.04	24.71	326.08	182.34	8.79
31	7.24	514.04	1231.31	0.84	80.11	724.95	3.08	24.5	305.28	182.11	8.56
32	9.30	511.61	1244.56	0.86	80.12	737.15	3.01	24.49	304.53	180.87	8.98
33	9.05	565.58	1302.28	0.87	79.92	756.22	3.47	24.44	318.10	180.63	8.98
34	8.05	497.58	1201.77	0.85	79.69	705.99	2.51	23.77	306.95	180.63	8.74
35	8.70	514.04	1226.46	0.87	80.19	724.28	3.10	24.89	302.26	182.17	9.17
36	8.22	548.48	1253.48	0.84	79.9	737.27	3.82	24.82	298.68	178.21	8.46

37	6.54	579.69	1526.59	0.77	55.89	899.82	4.80	9.91	247.10	156.66	5.37
38	5.90	595.69	1566.49	0.75	53.99	910.66	4.50	9.76	259.55	173.03	5.10
39	6.78	634.55	1565.53	0.76	65.05	910.12	5.87	10.00	229.73	154.94	5.25
40	5.61	670.61	1634.02	0.74	60.62	914.54	5.99	8.76	369.10	206.89	5.28
41	6.80	615.74	1591.99	0.81	78.10	906.91	3.97	6.78	338.32	192.19	7.01
42	7.64	548.48	1278.64	0.85	79.82	747.54	3.82	24.69	300.33	182.12	8.62
43	8.00	548.48	1274.41	0.86	80.55	747.97	3.82	25.52	309.26	182.12	8.91
44	9.00	532.03	1238.06	0.87	79.82	729.02	3.25	24.80	309.24	180.59	9.09
45	6.89	598.04	1362.59	0.82	80.23	782.14	3.84	24.80	309.90	183.12	7.76

MW is molecular weight, MolVol is molecular volume, SSS is molecular surface similarity score obtained from Surfex-Sim similarity, QFIT is pharmacophore feature fit obtained from UNITY based on GALAHAD-generated pharmacophore query, CLogP is molecular octanol-water partition coefficient computed with Tripos/Sybyl8.0, MolPSA is molecular polar surface area, MolPV molecular polar volume MolSA is molecular surface area, and DM is molecular Dipole_Moment.

Table 3-5. Partial Least Squares Analysis to generate relationships between bioactive data and combinations of their pharmacophore-based molecular similarity, shape-based molecular similarity and calculated molecular property descriptors

Model	Relationship	Partial Least Squares Analysis			
		Leave-One-Out		No Validation	
		r_{cv}^2	Optimum # of components	r^2	Standard Error of Estimate
1	$-\log(K_i)$ vs QFIT	N/A	N/A	0.05	1.27
2	$-\log(K_i)$ vs SSS	N/A	N/A	0.77	0.59
3	$-\log(K_i)$ vs SSS, QFIT	0.72	1	0.78	0.58
4	$-\log(K_i)$ vs SSS, QFIT, MolSA, MolPSA	0.84	3	0.89	0.43
5	$-\log(K_i)$ vs SSS, QFIT, MolPSA	0.81	2	0.87	0.45
6	$-\log(K_i)$ vs SSS, QFIT, MolSA	0.73	1	0.79	0.57
7	$-\log(K_i)$ vs SSS, QFIT, MolPV, MolVol	0.73	3	0.85	0.51
8	$-\log(K_i)$ vs SSS, QFIT, MolPV	0.75	3	0.82	0.56
9	$-\log(K_i)$ vs SSS, QFIT, MolVol	0.68	1	0.75	0.62
10	$-\log(K_i)$ vs SSS, QFIT, MW, ClogP	0.68	1	0.75	0.62

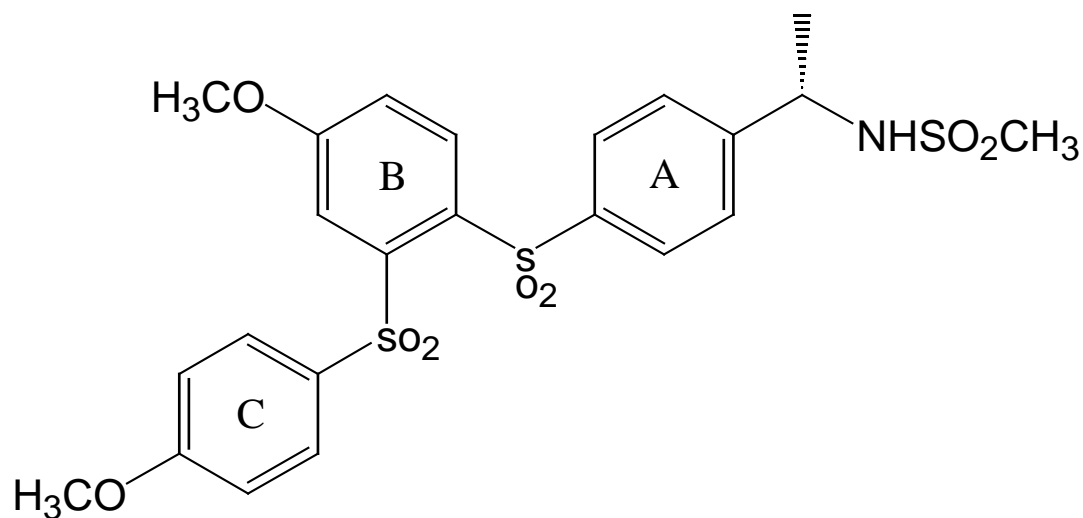
11	$-\log(K_i)$ vs SSS, QFIT, MW	0.66	1	0.72	0.66
12	$-\log(K_i)$ vs SSS, QFIT, ClogP	0.75	2	0.83	0.52
13	$-\log(K_i)$ vs SSS, QFIT, DM	0.78	1	0.84	0.51
14	$-\log(K_i)$ vs SSS, QFIT, MW, MolSA, ClogP, MolVol.	0.72	3	0.87	0.46
15	$-\log(K_i)$ vs SSS, QFIT, MW, MolSA, ClogP, MolVol, DM	0.76	3	0.87	0.46
16	$-\log(K_i)$ vs SSS, QFIT, MW, MolSA, ClogP, MolVol, DM, MolPSA	0.76	4	0.88	0.46
17	$-\log(K_i)$ vs SSS, QFIT, MW, MolSA, ClogP, MolVol, MolPSA	0.81	3	0.89	0.43
18	$-\log(K_i)$ vs SSS, QFIT, MW, MolSA, ClogP, MolVol, DM, MolPSA, MolPV	0.74	3	0.86	0.48
19	$-\log(K_i)$ vs SSS, QFIT, MW, ClogP, DM, MolPV	0.75	2	0.85	0.50
20	$-\log(K_i)$ vs SSS, QFIT, MW, ClogP, DM, MolPV, MolVol	0.73	3	0.86	0.50
21	$-\log(K_i)$ vs SSS, QFIT, MW, ClogP, MolPV, MolVol	0.74	3	0.86	0.50
22	$\log(K_i)$ vs SSS, QFIT, MW, MolSA,	0.82	4	0.90	0.43

	ClogP, MolPSA				
23	log(K_i) vs SSS, QFIT, MW, MolSA,	0.73	3	0.88	0.46
	ClogP				
24	log(K_i) vs SSS, QFIT, MW, MolPSA,	0.80	3	0.88	0.45
	ClogP				
25	log(K_i) vs SSS, QFIT, MW, MolPSA,	0.76	3	0.87	0.47
	ClogP, DM				
26	log(K_i) vs SSS, QFIT, MW, MolPSA,	0.77	4	0.87	0.46
	ClogP, MolSA, DM				
27	log(K_i) vs SSS, QFIT, MolSA, ClogP,	0.82	4	0.89	0.44
	MolPSA				
28	log(K_i) vs SSS, MolSA, MolPSA	0.77	2	0.84	0.51
29	log(K_i) vs QFIT, MolSA, MolPSA	0.39	1	0.55	0.83
30	-log(K_i) vs SSS, MW, MolSA, ClogP,	0.74	3	0.87	0.47
	MolVol, MolPSA, MolPV				
31	-log(K_i) vs QFIT, MW, MolSA, ClogP,	0.40	1	0.52	0.86
	MolVol, MolPSA, MolPV				
32	-log(K_i) vs SSS, MW, MolSA, ClogP,				
	MolVol, DM, MolPSA, MolPV.	0.67	2	0.77	0.61
33	-log(K_i) vs MW, MolSA, ClogP, MolVol,	0.37	2	0.54	0.86

DM, MolPSA, MolPV.

34 $-\log(K_i)$ vs MW, MolSA, ClogP, MolVol, 0.36 1 0.43 0.90
MolPSA

MW is molecular weight, MolVol is molecular volume, SSS is molecular surface similarity score obtained from Surfex-Sim similarity, QFIT is pharmacophore feature fit obtained from UNITY based on GALAHAD-generated pharmacophore query, CLogP is molecular octanol-water partition coefficient computed with Tripos/Sybyl8.0, MolPSA is molecular polar surface area, MolPV molecular polar volume MolSA is molecular surface area, and DM is molecular Dipole_Moment.



Scheme 1, Triaryl bis-sulfones

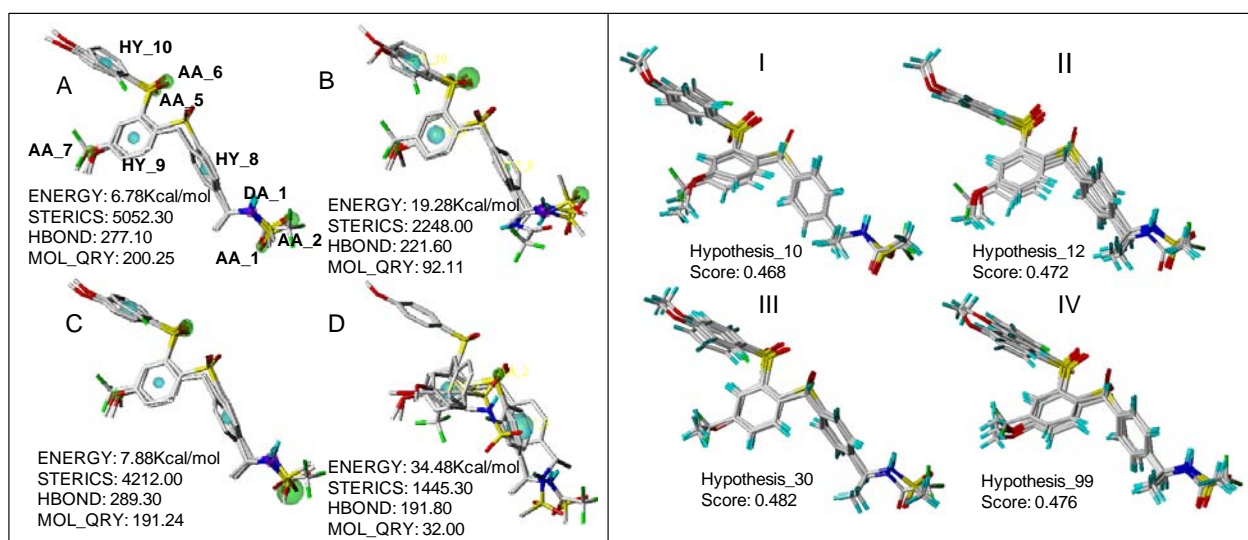


Figure 3-1. (Left) Four representative GALAHAD-generated hypermolecular superimpositions and correspondent pharmacophore models with four triaryl bis-sulfones 2, 5, 8, and 11. (Right) Four representative Surflex-Sim generated hypermolecular superimposition with four triaryl bis-sulfones 2, 5, 8, and 11.

As shown, the annotation label AA (green color) means the site of H-bond acceptor atom, DA (magenta color) is the site of H-bond donor atom, and HY (cyan color) means the site of hydrophobic center.

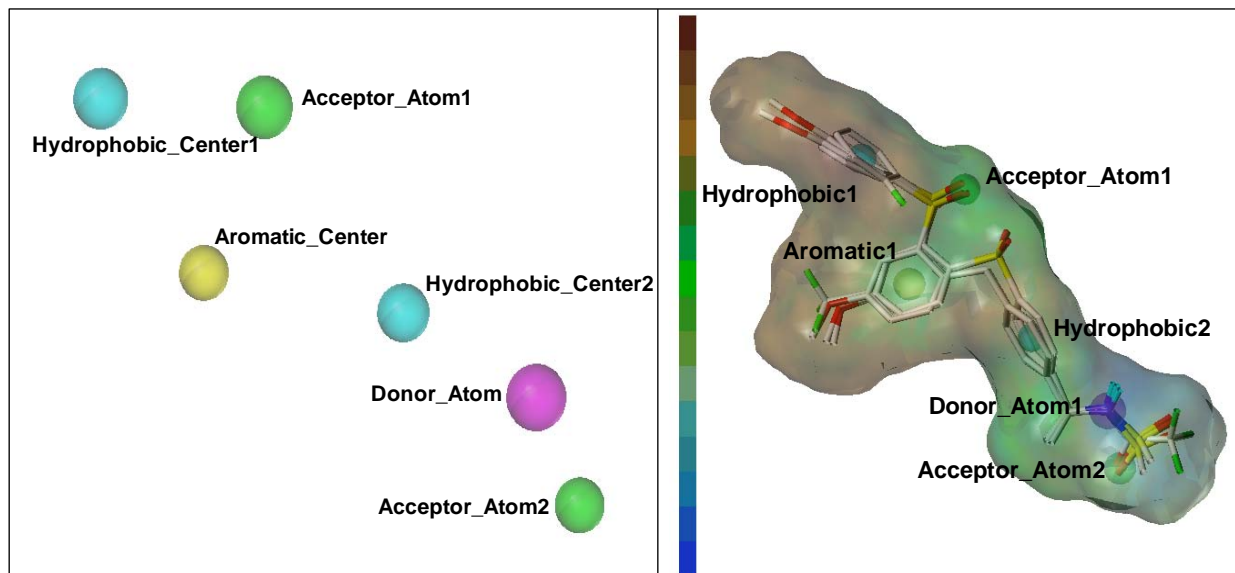


Figure 3-2. (Left) The modified pharmacophore query defined according to the SAR studies. (Right) Combination of the MOLCAD-generated molecular surface based on molecular shape hypothesis and the modified pharmacophore query based on GALAHAD-generated pharmacophore model.

The color bar indicates the color ramps for lipophilic properties on the molecular surface, and the top (or brown) is the most lipophilic.

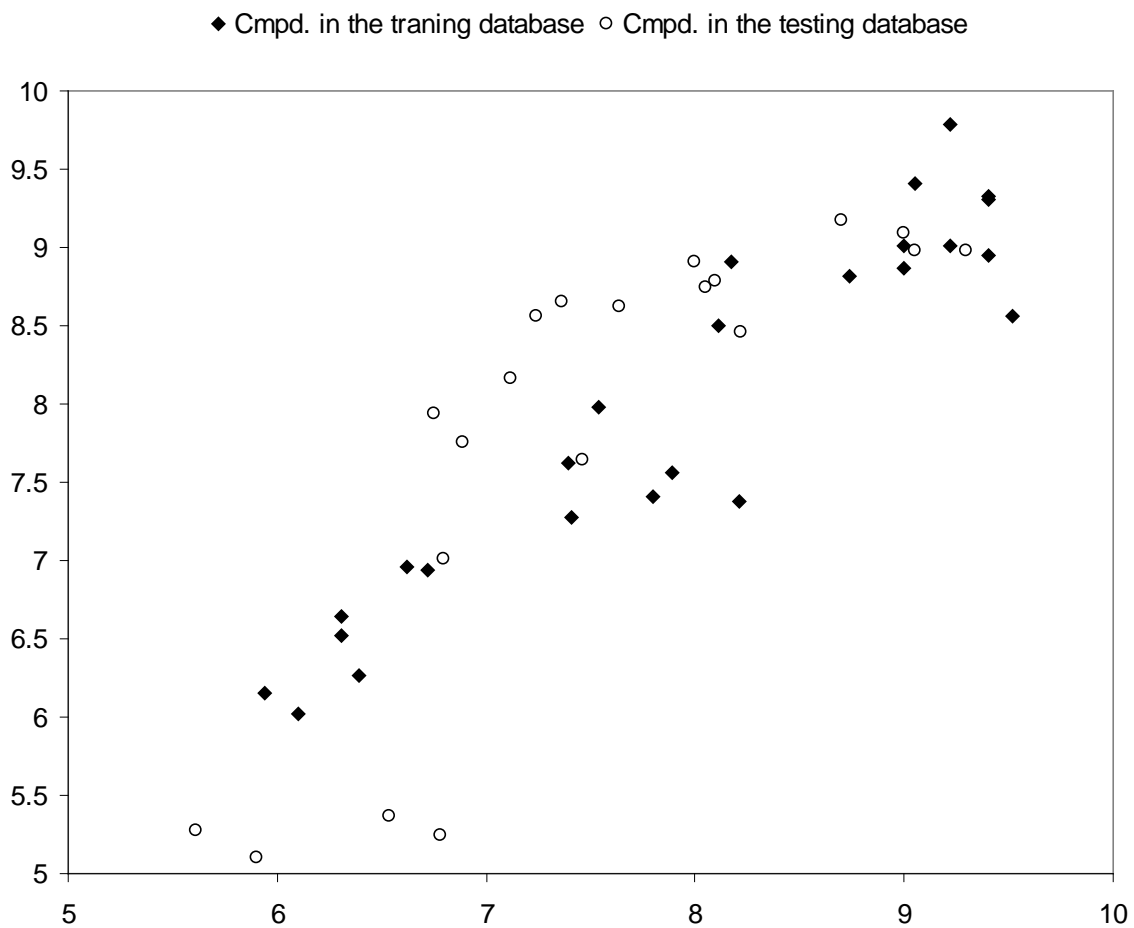


Figure 3-3. Plot of the corresponding QSAR calculated and experimental values of binding affinity (given as pK_i) of triaryl bis-sulfones in the training and testing sets at the CB2 receptor.

Filled squares indicate the training compounds, and unfilled circles indicate the testing compounds.

4.0 FINGERPRINT-BASED ARTIFICIAL NEURAL NETWORKS QSAR (FANN-QSAR) FOR LIGAND BIOLOGICAL ACTIVITY PREDICTIONSINTRODUCTION

Artificial neural networks (ANN) have been shown to be an effective tool in solving non-linear problems in several case studies ranging from engineering to biological applications [195-201]. ANN have several unique attributes that make them robust for non-linear generalization problems with multidimensional inputs. For instance, the networks have adaptive learning behaviors in which they learn from previous examples and adapt to changes in input parameters. In addition, they possess good generalization and pattern recognition property for unseen data. Several studies have used ANN to predict physicochemical and biological properties of chemical analogs. 2D and 3D molecular descriptors of molecular physical properties were used as neural network inputs to predict molecular properties or biological endpoints in several case studies such as anti-diabetes, anti-cancer and anti-HIV research [202-204]. However, to the best of our knowledge, there are no studies which use molecular fingerprints as descriptors in developing ANN-QSAR models to predict biological activities (such as pIC_{50} or pK_i) of chemical ligands although there are a few studies reported to predict ligand classes [205, 206]. In this work, three types of molecular fingerprints were used to train ANN-QSAR models, namely fingerprint-based ANN-QSAR (FANN-QSAR), and the results were compared to known 2D and 3D QSAR methods using five data sets. As a case study, the FANN-QSAR approach was used to predict cannabinoid receptor binding activities using a large and structurally diverse cannabinoid ligand data set. In fact, cannabinoid drug research is experiencing a great challenge as the first CB_1

antagonist drug, Rimonabant, launched in 2006 as an anorectic/anti-obesity drug, was recently withdrawn from the European market due to the complications of suicide and depression side effects [157]. As is known, structure-based design of novel CB₂ ligands that do not confer psychotropic side effects is hindered because of a lack of information about experimental 3D receptors structures, which is true, in general, for all drug discovery research involving G-protein coupled receptors (GPCRs). Thus, developing ligand-based QSAR approaches has its advantage for new CB₂ ligand design and discovery. To prove one of useful applications of the FANN-QSAR model, it was applied as a virtual screening tool to find new cannabinoid ligands from a large NCI database containing over 200,000 compounds, and four compounds with good cannabinoid receptor binding affinities were found through this exercise. This study demonstrated that combination of molecular fingerprints and ANN can lead to a reliable and robust high-throughput virtual screening method that can be a useful tool in chemogenomics and computer-aided drug discovery research.

4.2 METHODS

4.2.1 Data sets

A total of six data sets were used in this study. Five of them were compiled by Sutherland *et al.* [207] and were downloaded from their supplemental data. The sixth data set was curated by our lab. Data sets are: (1) A set of 114 angiotensin-converting enzyme (ACE) inhibitors [208] with pIC₅₀ values ranging from 2.1 to 9.9. (2) A set of 111 acetylcholinesterase (AChE) inhibitors [209, 210] with pIC₅₀ values ranging from 4.3 to 9.5. (3) A set of 147 ligands for the

benzodiazepine receptor (BZR) [211] with pIC_{50} values ranging from 5.5 to 8.9 after removal of 16 inactive compounds with a single pIC_{50} value of 5.0. (4) A total of 282 cyclooxygenase-2 (COX2) inhibitors [212-221] with pIC_{50} values ranging from 4.1 to 9.0 after removal of 40 inactive COX2 compounds with a single pIC_{50} value of 4.0. (5) A total of 361 dihydrofolate reductase inhibitors (DHFR) from the work of Queener *et al.* [222-226] with pIC_{50} ranging from 3.3 to 9.8 after removal of 36 ligands with a single pIC_{50} value of 3.3. Figure 4-1 contains representative structures from the above 5 data sets. (6) A set of cannabinoid receptor subtype 2 (CB₂) ligands [227] with pK_i values ranging from 3.9 to 10.8. For the cannabinoid ligand (CBID) dataset, ligand structures and their bioactivities (K_i) were curated by our lab. If there were more than one reported CB₂ activity for a ligand, an average activity was used. Figure 4-2 contains representative CB₂ ligands displaying the structural diversity of the data set.

For the ACE, AchE, BZR, COX2 and DHFR datasets, the same training and testing data sets provided by Sutherland *et al.* were used for direct comparisons of FANN-QSAR models to 3D and 2D QSAR models reported by Sutherland *et al.* For each dataset, 10% of randomly selected compounds from the training set were used as a validation set. For the cannabinoid data set, the training and test sets were randomly divided. The training set contained 80% of compounds while the test set contains 10%. The other 10% were used as a validation set. Numbers of compounds found in each training, validation and test sets for each data set are summarized in Table 4-1. The training set was used to train the model while the validation set was used to prevent overfitting of the model. The test set was used as an external set to evaluate the generalization ability of the trained FANN-QSAR models. For statistical modeling, the process was repeated five times resulting in five different pairs of randomly divided training and test sets.

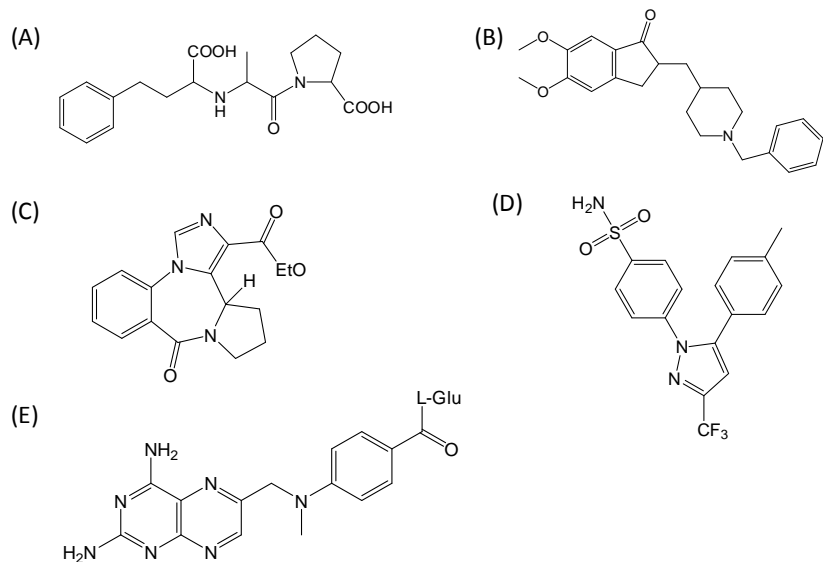


Figure 4-1. Representative compounds from five QSAR data sets: (A) enalaprat (ACE); (B) E2020 (AchE); (C) Ro14-5974 (BZR); (D) celecoxib (COX2); (E) methotrexate (DHFR).

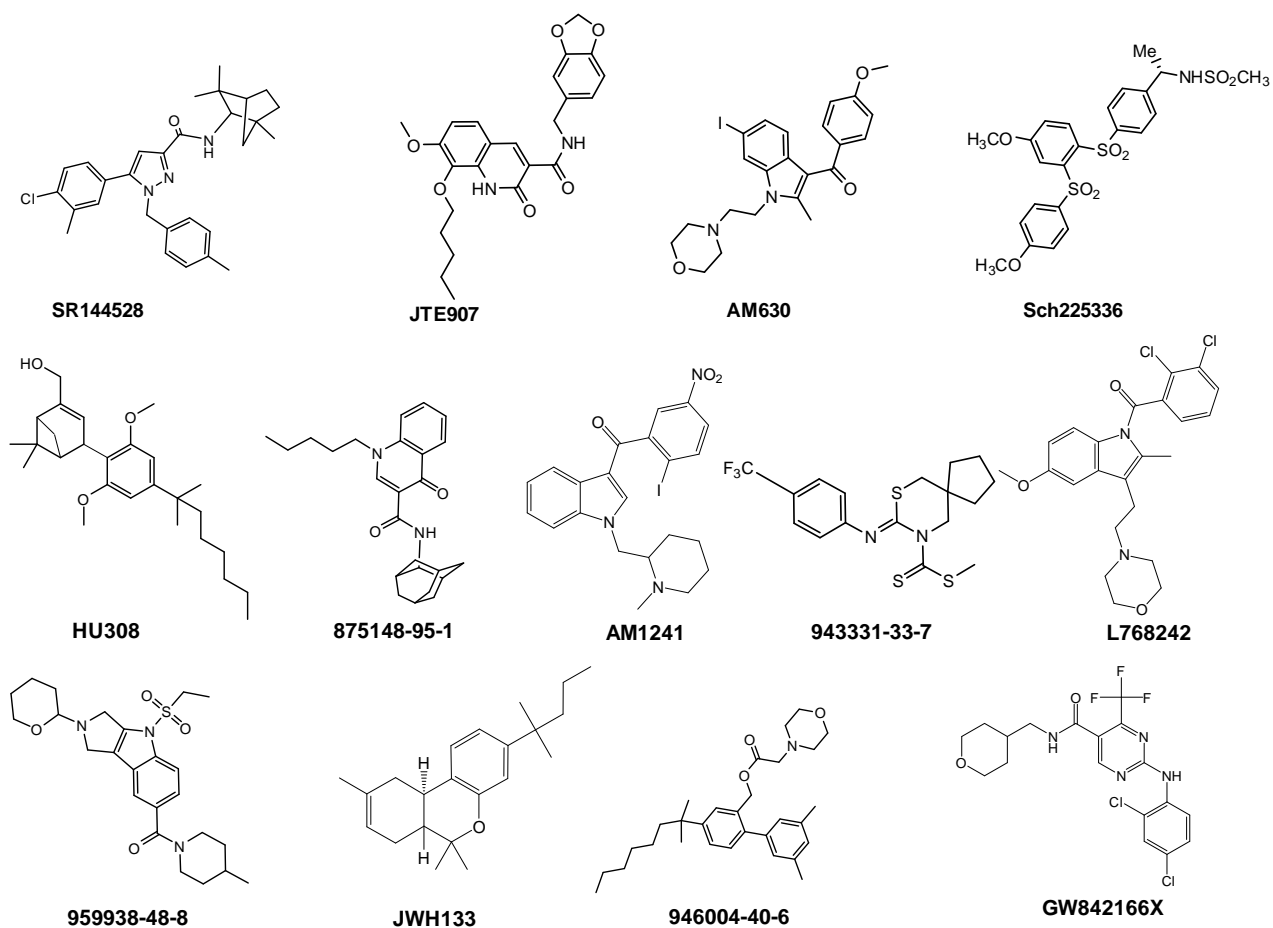


Figure 4-2. Representative CB₂ compounds from CBID data set, reflecting the structural diversity of the data set.

Table 4-1. Numbers of training, validation and test set compounds in each data set.

	ACE	AchE	BZR	COX2	DHFR	CB ₂
Training size	69	67	89	170	214	1361
Validation size	7	7	9	18	23	169
Test size	38	37	49	94	124	169
Total	114	111	147	282	361	1699

4.2.2 Fingerprint generation

Three different types of molecular fingerprints, namely FP2 [228, 229], MACCS [230] and Extended-Connectivity Fingerprint (ECFP6) [231], were used in this study. FP2 is a path-based fingerprint which indexes molecular fragments and MACCS is a key-based fingerprint which uses 166 predefined keys whereas ECFP6 is a circular topological fingerprint which is derived using a variant of the *Morgan* algorithm [232]. FP2 and MACCS fingerprints were generated using the *'babel'* command from the OpenBabel program [228, 229] while ECFP6 fingerprints were generated using the *'generatemd'* command from the ChemAxon program [233]. Ligand chemical structures stored in SDF format were used as inputs to generate fingerprints. For each ligand, polar hydrogens were added using the OpenBabel program [228] before fingerprint generations. All fingerprints are fixed-length binary representations with 1024 bits for both ECFP6 and FP2, and 256 bits for MACCS fingerprint. Fingerprints were generated for each ligand in the datasets and used as inputs to train the FANN-QSAR models.

4.2.3 Fingerprint-based Artificial Neural Network QSAR

A feed-forward back-propagation neural network method was implemented using MATLAB® R2007b Neural Network Toolbox [234]. As shown in Figure 4-3, there are three layers in the network: an input layer, a hidden layer and an output layer. The number of input layer neurons is equal to the size of fingerprint. For example, FP2 and ECFP6 fingerprints have 1024 bits and therefore, the number of input neurons is equal to 1024. Similarly, there are 256 input neurons for MACCS fingerprint. The number of hidden layer neurons was varied between 100 and 1000. The networks were trained using the gradient descent with momentum training function

(*traingdm*) to update weights and biases, the tangent sigmoid transfer function (*tansig*) for the hidden layer and the linear transfer function (*purelin*) for the output layer. 10% randomly selected compounds from the training data was used as a validation set to decide when to stop training. The model training was stopped after 4000 epochs (iterations) or if the mean-square-error (MSE) of prediction on the training set had reached the minimum value of 0.1. In addition, an early stopping was enabled when the prediction error on the validation set kept increasing for 300 epochs and the weights and biases at the minimum of the validation error were returned. The optimal number of hidden neurons was selected via cross-validation experiments in which the model was trained using different number of hidden neurons and an average of training set and validation set mean squared errors (MSE) was calculated. The number of hidden neurons which gave the lowest average MSE was used as the optimal number for subsequent model testing on the test set. The mean square error (MSE) is defined as:

$$MSE = \frac{1}{T} \sum_{i=1}^T (t(i) - p(i))^2 \quad (4.1)$$

where T is the total number of training samples; $t(i)$ is the target value of i^{th} sample and $p(i)$ is the predicted value of i^{th} sample.

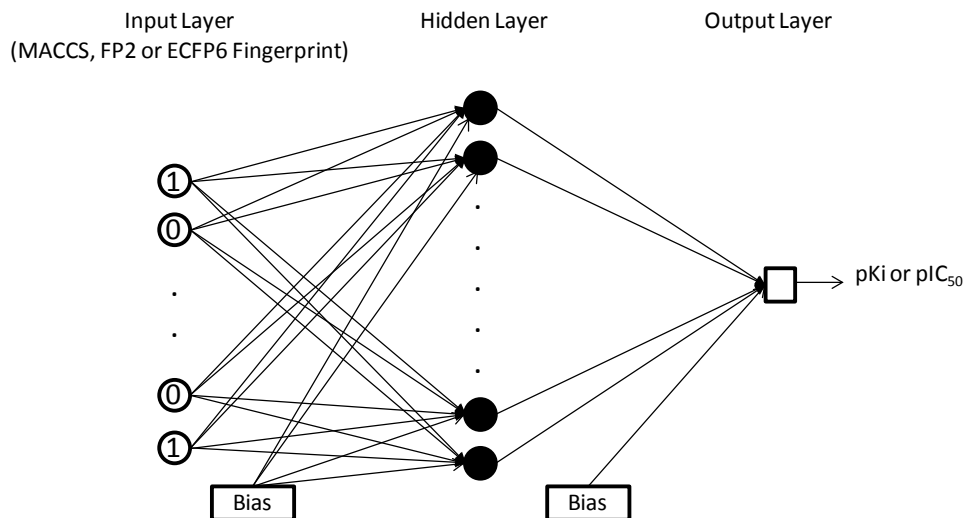


Figure 4-3. The architecture of fingerprint-based ANN-QSAR (FANN-QSAR) model.

4.2.4 Radioligand competition binding assay

In order to evaluate CB₂ binding activity of virtually screened ligands, competition binding assays were performed to determine each ligand's ability to displace radioactive [³H]CP-55940 radioligand. The experimental protocol has been established based on previously reported procedures [235-237] and is described briefly below.

A Perkin Elmer 96-well TopCounter was used in our laboratory to measure the CB receptor binding affinity (K_i) of the *in silico* screened ligands by displacing [³H]CP-55940. In competition binding experiments, ligands were diluted in dilution buffer (50 mM Tris, pH 7.4, containing 5 mM MgCl₂, 2.5 mM EGTA, 0.1% (w/v) fatty acid free bovine serum albumin (BSA), 10% dimethylsulfoxide and 0.4% methyl cellulose). Various concentrations of ligands/samples were added in the same volume to 2.5 nM [³H]CP-55940. Incubation buffer (50 mM Tris, pH 7.4, containing 2.5 mM EGTA, 5 mM MgCl₂, and 0.1% (w/v) fatty acid free BSA)

and cell membrane preparations from CHO cells that expressing CB₂ receptors (5 µg per well) were added to a final volume of 200 µL. For the saturation binding experiments, varying concentrations of [³H]CP-55940 (0.05-4 nM) with or without 5 µM of an unlabeled known ligand (CP-55940) were incubated with the receptor membrane preparations to determine K_d and nonspecific binding. After the binding suspensions were incubated at 30 °C for 1 hr, the reaction was terminated by rapid filtration through microfiltration plates (Unifilter GF/B filterplate, Perkin Elmer) followed by 5 washes with ice cold TME buffer containing 0.1% BSA on a Packard Filtermate Harvester (Perkin Elmer). The plates were then dried overnight and 30 µl MicroScint 0 scintillation liquid was added to each well of the dried filter plates. Then the bound radioactivity was then counted using a Perkin Elmer 96-well TopCounter. The K_i was calculated by using nonlinear regression analysis (Prism 5; GraphPad Software Inc., La Jolla, CA), with the K_d values for [³H]CP-55940 determined from saturation binding experiments. This assay was used for determining binding affinity parameters (K_i) of ligand-receptor interactions between the CB₂ receptor and ligands.

4.3 RESULTS AND DISCUSSION

The FANN-QSAR models were first compared to known QSAR methods and then it was used to predict cannabinoid receptor binding activity on structurally diverse data set. Moreover, the generalization ability of the FANN-QSAR model was also examined by predicting activities of newly reported cannabinoid ligands. In addition, it was used as a virtual screening tool to screen

large NCI compound database for potential cannabinoid lead ligands and virtual hits were also validated by radioligand competition binding assays.

4.3.1 Comparisons with other 3D and 2D-QSAR methods

The performances of the derived FANN-QSAR models were evaluated and compared to the reported 3D and 2D QSAR methods, including CoMFA [30], CoMSIA [61], Hologram QSAR (HQSAR) [10, 185], QSAR by eigenvalue analysis (EVA) [238], back-propagation feed-forward neural network implemented in Cerius2 using 2.5D descriptors (NN 2.5D) and ensemble neural network [239] (NN-ens) using 2.5D descriptors which were implemented and tested by Sutherland *et al.* [207] For an objective comparison, the FANN-QSAR models were trained and tested on the same training and test data sets provided by Sutherland *et al.* Three different fingerprints were used as inputs for FANN-QSAR models and each model was trained separately for each fingerprint type. During each training process, a cross-validation experiment was performed to decide the optimal number of hidden neurons which was used subsequently on the test set prediction. Cross-validation results can be found in Appendix A (Figures A1-A5).

Final correlation coefficient (r^2 train and r^2 test) values of each dataset are listed in Table 4-2. Comparisons of r^2 (test) values across all data sets show that ECFP6 fingerprint-based ANN-QSAR model (ECFP6-ANN-QSAR) performed better than FP2 and MACCS fingerprint-based models for all data sets. For ACE, AchE and COX2 data sets, the CoMFA model performed better than ECFP6-ANN-QSAR model but by a small margin. The ECFP6-ANN-QSAR model performed better for the DHFR and BZR datasets. The CoMSIA model performed similarly as the ECFP6-ANN-QSAR model. It is important to note that CoMFA and CoMSIA are field-based 3D QSAR methods which require similar scaffolds and high quality molecular

alignments to make effective predictions [5]. On the other hand, ECFP6-ANN-QSAR is a fingerprint-based method which works on structurally diverse data sets and requires no alignment during the model training process which makes it more robust and high-throughput in virtual screening. However, different fingerprints can produce different results and, in our work, ECFP6 produced an overall better result across different data sets compared to FP2 and MACCS fingerprints. In addition to 3D QSAR methods, the FANN-QSAR models were compared to another 2D QSAR method known as Hologram QSAR (HQSAR), which is based on molecular holograms containing counts of molecular fragments similar to fingerprints. It can be observed that ECFP6-ANN-QSAR performed consistently better than HQSAR in all datasets except for DHFR dataset resulting in the same r^2 test value (0.63). Moreover, the FANN-QSAR models were also compared to other neural network approaches that used 2.5D descriptors as reported by Sutherland *et al.* ECFP6-ANN-QSAR model performed better than NN (2.5D) method in 3 out of 5 datasets and an ensemble of 10 neural networks (NN-ens) approach using 2.5D descriptors performed slightly better than ECFP6-ANN-QSAR model in 3 out of 5 data sets. It is important to note that all QSAR models failed for COX2 and BZR datasets (r^2 test < 0.34) and had moderate performances (r^2 test < 0.64) for the other three datasets. Overall, ECFP6-ANN-QSAR model performed consistently across all datasets and its performance was comparable to other 3D, 2D, and neural networks QSAR methods previously reported.

Table 4-2. FANN-QSAR performance comparisons with other QSAR methods*

ECFP6-	FP2-	MACCS-	CoMFA	CoMSIA	HQSAR	EVA	NN	NN-
ANN-	ANN-	ANN-		basic			(2.5D)	ens
QSAR	QSAR	QSAR						(2.5D)

ACE									
r^2 train	0.75	0.93	0.23	0.80	0.76	0.84	0.84	0.78	0.84
r^2 test	0.41	0.20	0.08	0.49	0.52	0.30	0.36	0.39	0.51
AchE									
r^2 train	0.94	0.57	0.62	0.88	0.86	0.72	0.96	0.68	0.63
r^2 test	0.43	0.13	0.04	0.47	0.44	0.37	0.28	-0.04	0.21
BZR									
r^2 train	0.76	0.78	0.78	0.61	0.62	0.64	0.51	0.62	0.66
r^2 test	0.31	0.08	0.06	0.00	0.08	0.17	0.16	0.39	0.34
COX2									
r^2 train	0.73	0.76	0.89	0.70	0.69	0.70	0.68	0.65	0.65
r^2 test	0.28	0.22	0.23	0.29	0.03	0.27	0.17	0.31	0.32
DHFR									
r^2 train	0.94	0.72	0.84	0.79	0.76	0.81	0.81	0.78	0.79
r^2 test	0.63	0.43	0.48	0.59	0.52	0.63	0.57	0.42	0.54

*Other QSAR methods including CoMFA, CoMSIA basic, HQSAR, EVA, NN (2.5D) and NN-ens (2.5D) performance indicators were taken from the work of Sutherland et al. [207] FANN-QSAR models were trained and tested on the identical training and test sets provided by Sutherland et al. for comparison purposes.

4.3.2 Prediction of Cannabinoid receptor binding activity using FANN-QSAR method

A total of 1699 structurally diverse cannabinoid ligands with reported CB₂ binding affinities were used. The ligands were randomly divided into training and test sets. FANN-QSAR models using different fingerprints were trained on training sets and the optimal numbers of hidden neurons were selected via cross-validation. Figure 4-4 contains a summary of cross-validation results for all three FANN-QSAR models. It can be observed that different training and test sets as well as different types of fingerprints resulted in different optimal number of hidden neuron which suggested that cross-validation experiments are necessary to train neural networks for the best results.

After such training and parameters tuning, the predictive accuracy of the final model on the test set was evaluated. The process was repeated 5 times and a summary of r^2 values from each round of experiment can be seen in Table 4-3. Within each round, the same training and test compounds were used across all three FANN-QSAR models. For example, the same training and test compounds in Round 1 of ECFP6-ANN-QSAR model were used in the Round 1 of FP2-ANN-QSAR and MACCS-ANN-QSAR model. As shown in the table, ECFP6-ANN-QSAR model consistently outperformed FP2- and MACCS-ANN-QSAR models in all five rounds of experiments. The ECFP6-ANN-QSAR model achieved an average r^2 test value of 0.56 ($r = 0.75$) across all repeat experiments while 0.48 ($r = 0.69$) and 0.45 ($r = 0.67$) for FP2- and MACCS-ANN-QSAR models respectively. Results showed that ECFP6 fingerprint was better than FP2 and MACCS fingerprints for the cannabinoid data set as well as other five data sets. In fact, it has been also reported that circular fingerprints such as ECFP6 fingerprints are found to be more useful in virtual screening and ADMET properties prediction studies [240, 241]. Our results suggested that ECFP6 fingerprint-based ANN-QSAR model can be used in virtual screening of

chemical ligands in high throughput manner since it only requires 2D fingerprints as inputs instead of 3D molecular alignments and bioactive conformations as in other 3D QSAR methods.

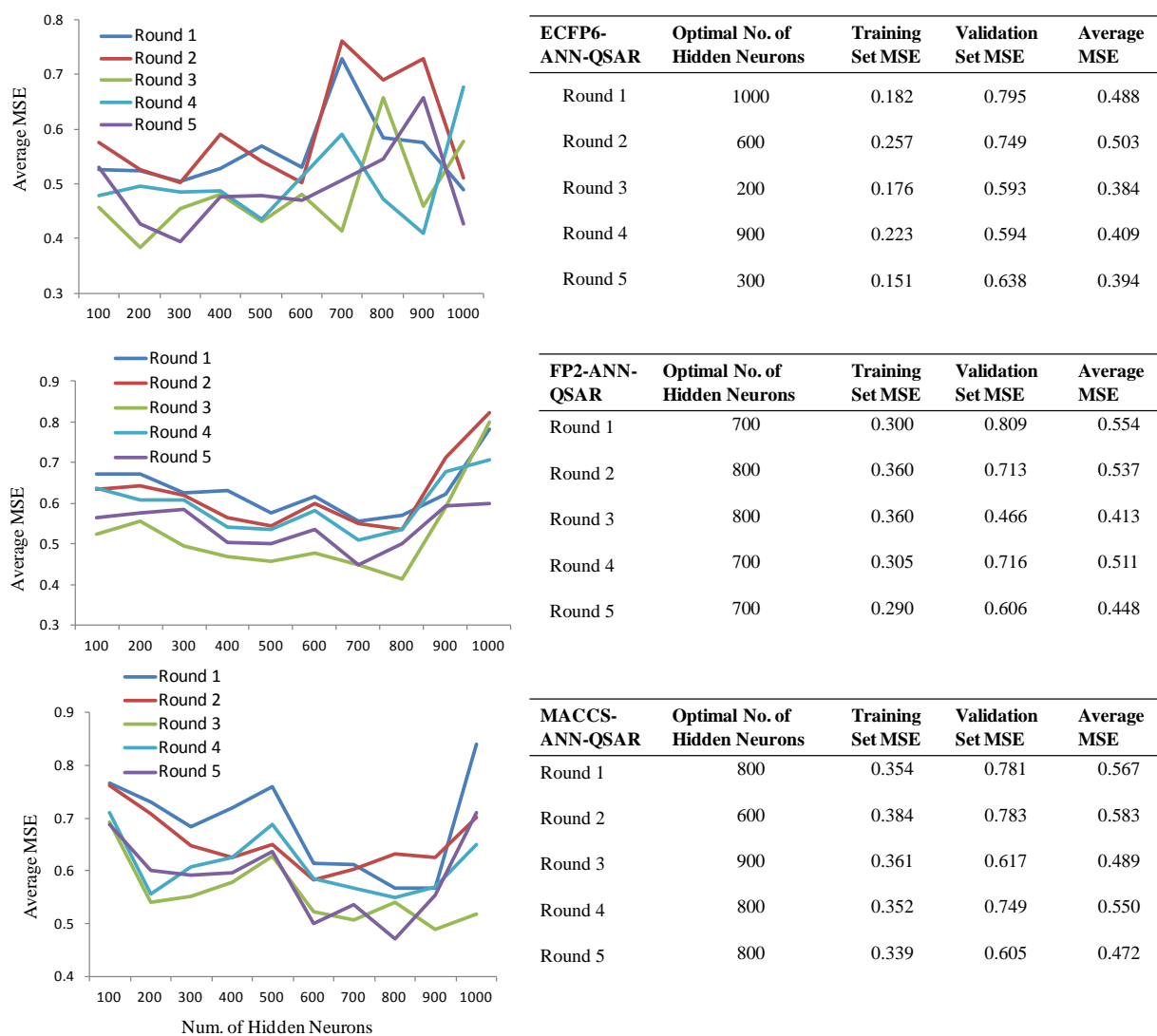


Figure 4-4. Cross-validation results of each FANN-QSAR method on CB2 ligand data set.

Table 4-3. A summary of the performance of each FANN-QSAR model on CB₂ ligand data set

Round	r^2 training	r^2 test
ECFP6-ANN-QSAR		
1	0.86	0.55
2	0.81	0.63
3	0.87	0.53
4	0.84	0.56
5	0.89	0.54
FP2-ANN-QSAR		
1	0.78	0.55
2	0.74	0.60
3	0.74	0.38
4	0.77	0.46
5	0.79	0.40
MACCS-ANN-QSAR		
1	0.74	0.48
2	0.72	0.53
3	0.74	0.37
4	0.74	0.47
5	0.75	0.41

4.3.3 Generalization ability of FANN-QSAR method on newly reported cannabinoid ligands

To test the predictive ability of FANN-QSAR method on new cannabinoid compounds which are not in the Xie group's cannabinoid ligand training data set, the most recently reported cannabinoid ligands and associated CB₂ binding affinity data were downloaded from ChEMBL database [242]. These compounds were not found in the training (CBID) data set and were collected to be used as a new test set in order to evaluate the FANN-QSAR performance. The new test data set consists of 295 compounds with reported CB₂ Ki values which were then converted to pKi values. 41.55% of new CB₂ ligands were less than 80% similar (2D Tanimoto similarity) and 25.34% were less than 70% similar to the training compounds. This similarity analysis indicated that the newly reported CB₂ compounds contained a good mixture of similar and dissimilar compounds to the training database. As discussed in Section 4.2.3, an ECFP6-ANN-QSAR model was trained accordingly using the 1699 CB₂ ligand (CBID) data set. Twenty independent rounds of training and testing were performed. For each round, randomly selected 90% of the database was used for training and 10% was used for validation. During this exercise, more training rounds, compared to 5 rounds in the previous section (Section 4.3.2), were applied in order to have a better coverage on the diversity of training molecules since the models would be tested on molecules that were not found in the training CBID data set. A summary of training and cross-validation results can be seen in the supplemental table in Appendix A (Table A1). After 20 rounds of predictions, an average predicted value for each test compound was calculated. Probability density function and cumulative distribution function plots of residual values of test compounds can be seen in Figure 4-5 indicating the values fall under the Gaussian distribution with the average residual value of 0.046 and standard deviation of 1.03. Seventeen

outlier compounds with residuals more than two standard deviations away from the average residual were removed. Figure 4-6 shows a scatter plot of experimental and predicted pKi values of 278 test compounds after such outlier removal. The linear regression of these 278 data points provided an r of 0.75, slope of 0.686, and Y intercept of 2.249. A plot of these data shown in Figure 4-7 indicated that there was a good correlation between experimental and predicted values given the fact that many of these test compounds have novel structures and were not included in the model training and validation process. The result suggested that the FANN-QSAR possessed good generalization ability for newly reported cannabinoid ligands.

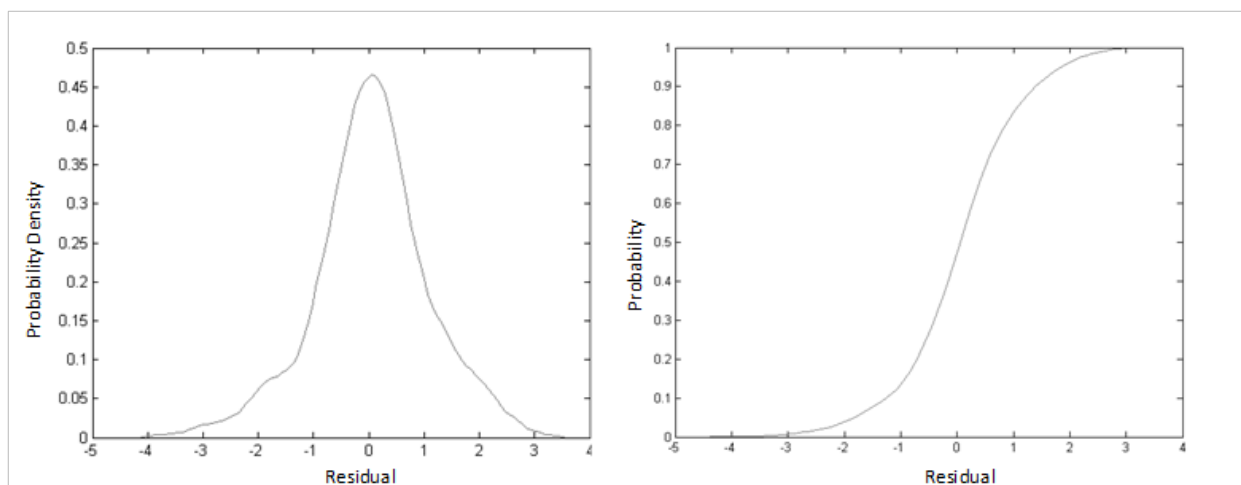


Figure 4-5. Probability density function (left) and cumulative distribution function (right) plots of residual values of 295 newly reported cannabinoid compounds.

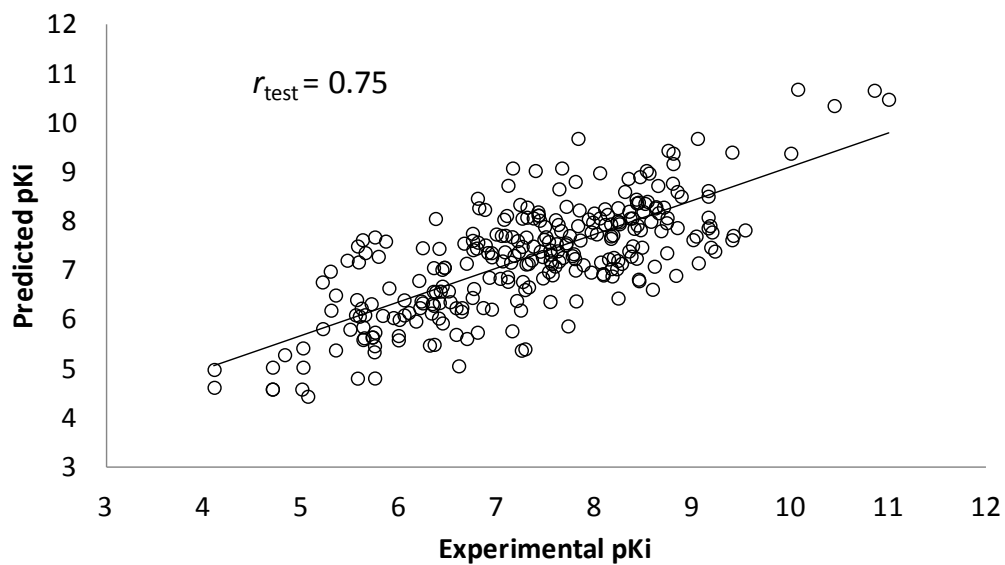


Figure 4-6. Scatter plot between experimental pKi and predicted pKi values of 278 test cannabinoid ligands after the removal of 17 outliers.

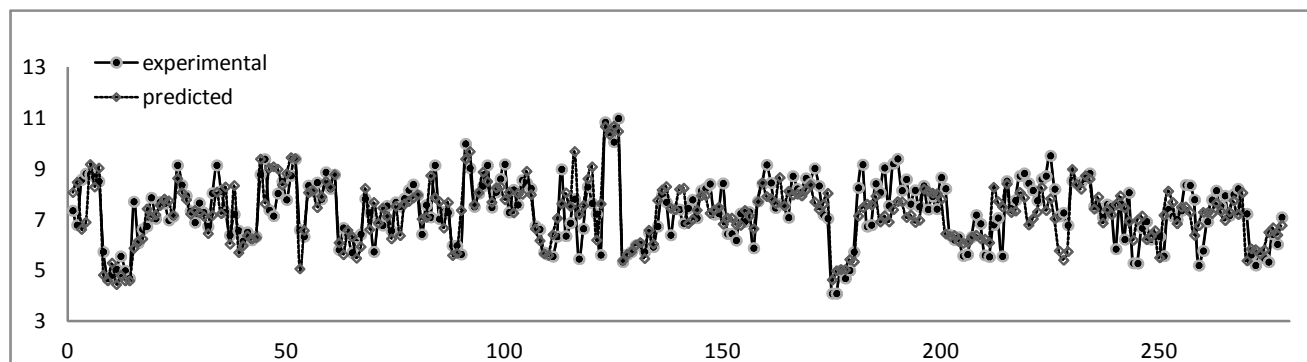


Figure 4-7. Experimental and predicted pKi values of 278 test cannabinoid ligands

4.3.4 An application of FANN-QSAR: virtual screening of the NCI compound database

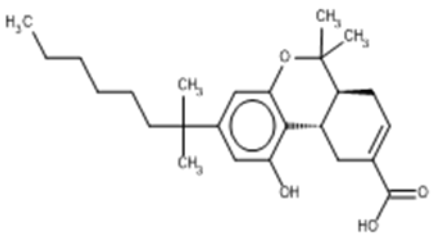
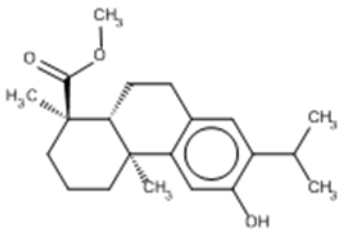
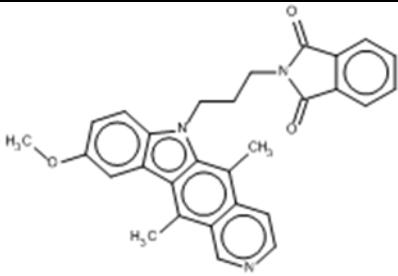
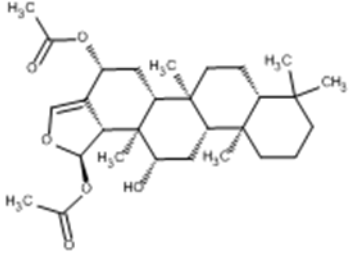
To further illustrate a possible application of FANN-QSAR in drug discovery research, a virtual screening experiment on NCI compound database [243] was performed to search for CB₂ lead ligands. For consistency, the same 20 trained models in the previous section were used. As a test set, compounds from NCI database were used for each round of prediction. Before testing, the database was filtered to remove duplicate compounds, isotopes, metals and mixtures using the Tripos Selector program [185]. The NCI database contained 329,089 compounds, and after such filtrations it was reduced to 211,782 compounds. For each compound, the ECFP6 fingerprint was generated and used as network inputs to predict the CB₂ activity (pKi). After 20 rounds of predictions, an average predicted value for each compound was calculated. The top ranked 50 compounds were selected, but only 10 compounds were available from the NCI via material transfer agreement (MTA). These 10 were for CB₂ activities using [³H]CP-55940 competition binding assay experiments as a validation of the method.

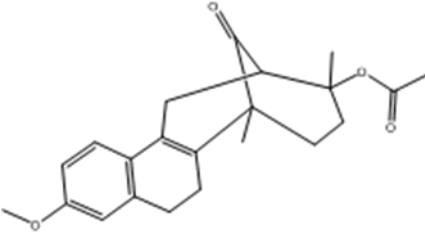
Among the 10 tested NCI compounds, four (NSC49888, NSC174122, NSC369049 and NSC76301) had CB₂ Ki between 6.70 nM (pKi = 8.17) and 3.80 μM (pKi = 5.42). One compound, which has a similar chemical scaffold as the well-known cannabinoid ligand, delta-9-tetrahydrocannabinol, was found to be a high affinity compound with an average CB₂ Ki value of 6.70 nM (pKi = 8.17). These four compounds and other similar compounds (70% 2D Tanimoto similarity threshold was used) [21] were not found in the training database used to train the model. Among top 50 ligands, there was one NCI compound (NSC768843) which was more than 90% similar (Tanimoto coefficient 0.9 using FP2 fingerprint) to a known classical cannabinoid ligand (CAS ID: 112830-95-2 or HU210), an analog of delta-9-tetrahydrocannabinol, reported in the literature [244]. These findings proved that FANN-QSAR

method can find not only novel compounds with good CB₂ binding affinities but also compounds similar to known ligands from a testing database containing thousands of compounds with diverse scaffolds. Hit ligands with novel scaffolds can be used as lead compounds for further medicinal chemistry optimization and SAR studies while hits similar to known ligands provide additional information for scaffold hopping and R-group variations which may be useful for medicinal chemists. Table 4-4 contains the structures of NCI hit compounds and their experimental pK_i as well as predicted values. Except for one compound (NSC746843) that was not available from NCI, the other four compounds were experimentally tested in our lab and competition binding curves are shown in Figure 4-8.

It should be noted that the predicted pK_i correlated well with experimental pK_i for two of five hit ligands but not for the other three ligands. It could be attributed to the experimental variability of the reported CB₂ binding activities of training compounds among different research labs or a possible limitation of 2D fingerprint descriptors which considers individual fragment contributions but sometimes may not be as effective as other 3D descriptors when considering an overall structure of a ligand. Fingerprints such as ECFP6 have, however, been found to be useful in this study as well as other several cheminformatic studies [240, 241], and they are known to be robust and time efficient for high throughput virtual screening applications where hundreds of thousands of chemicals are involved like in this study. To conclude, results from the virtual screening exercise that was validated experimentally demonstrated that the derived FANN-QSAR model is capable of successfully identifying lead CB₂ compounds with good CB₂ binding affinities as well as compounds similar to known cannabinoid ligands, and providing additional insights for R-group and scaffold hopping of known ligands.

Table 4-4. Identified NCI hit compounds with CB₂ binding activities

Structure	NSC ID	MW	ClogP	Experimental pKi	Predicted pKi
	746843	400.55	6.61	8.81*	8.66
	49888	330.46	5.59	8.17**	8.28
	174122	463.52	4.76	5.59***	8.41
	369049	488.66	4.00	5.51***	8.48

	746843	354.44	3.99	5.42***	8.21
---	--------	--------	------	---------	------

*An average literature reported K_i value of a known cannabinoid compound (HU210) which is more than 90% similar to 746843. **An average K_i value of two independent experiments performed in duplicate. ***An experimental K_i value of one experiment performed in duplicate.

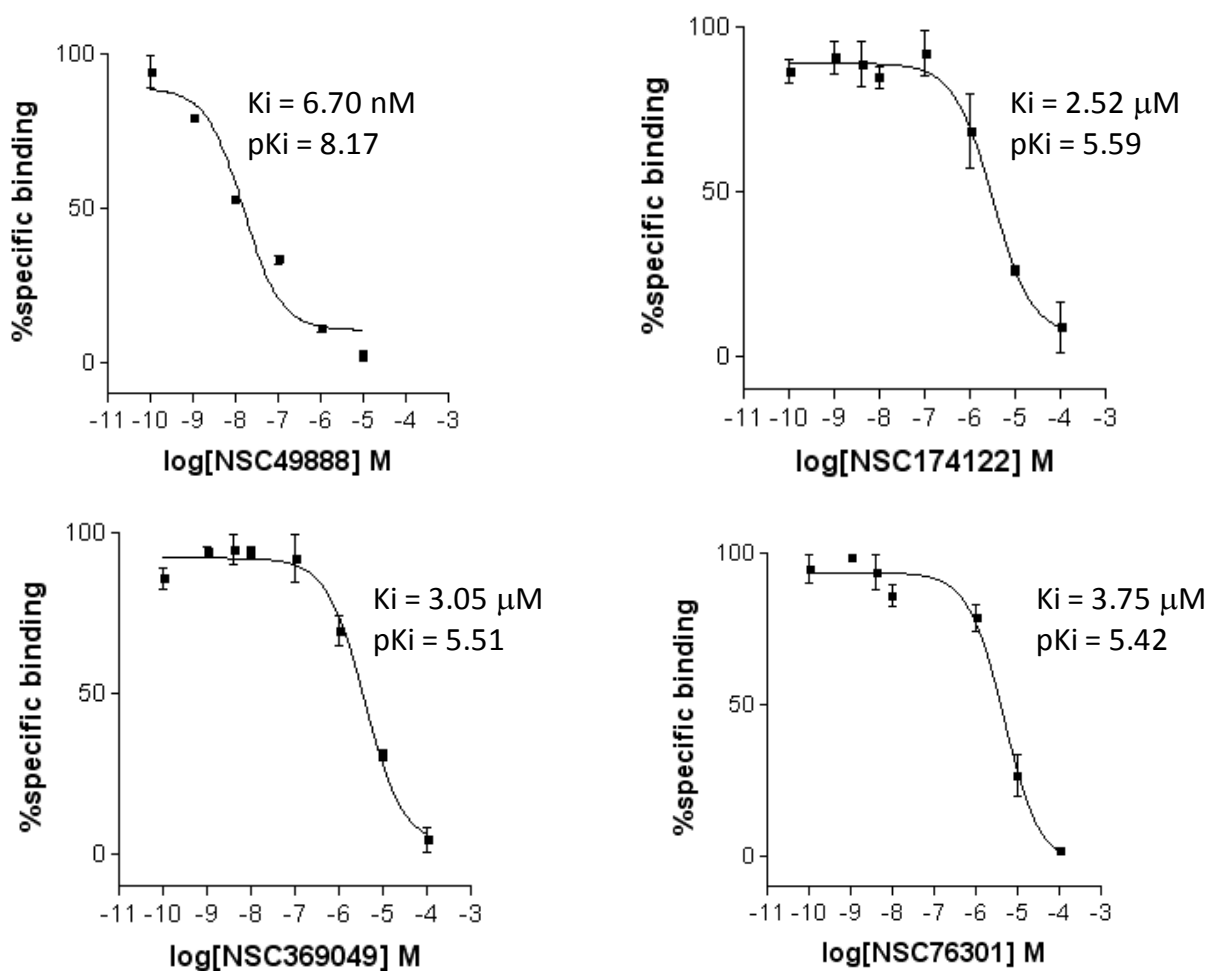


Figure 4-8. CB2 receptor binding affinity K_i values of four NCI hit compounds measured by [3 H]CP-55940 radioligand competition binding assay using human CB2 receptors harvested from transfected CHO-CB2 cells.

4.4 CONCLUSION

In this work, a novel fingerprint-based QSAR algorithm using artificial neural networks approach was introduced. Five data sets were used to compare the developed FANN-QSAR approach to known 3D and 2D QSAR models. The FANN-QSAR model's results were similar to both 3D CoMFA and CoMSIA, and better than HQSAR method for all five data sets. It should be noted that 3D CoMFA and CoMSIA requires knowledge of ligand bioactive conformations and high quality molecule alignments for predictive models whereas the FANN-QSAR model requires only two-dimensional structures for fingerprint generations. The model performance was comparable to other reported QSAR methods such as EVA and neural network approaches using 2.5D descriptors. In addition, the FANN-QSAR model was applied to a large structurally diverse cannabinoid ligand data set to predict CB₂ binding activities and achieved an average *r* (test) value of 0.75. In addition, to evaluate the generalization ability of FANN-QSAR method on unseen cannabinoid ligands, it was tested on a set of 278 newly reported cannabinoid ligands not presented in the training data set and achieved good prediction accuracy. Moreover, to show a useful application of the FANN-QSAR method, it was used to virtually screen NCI compound database and top hits were experimentally validated. Four out of 10 tested compounds were found to have good CB₂ binding affinities. The lead compounds are currently being subjected for further lead optimization and SAR studies. To conclude, the FANN-QSAR method can be a useful application in computer-aided drug discovery research to predict biological activities or

properties of unknown ligands and screen large structurally diverse databases for novel lead discovery.

5.0 MOLECULAR MODELING, DISCOVERY AND QSAR STUDY OF NOVEL CANNABINOID LIGANDS

5.1 INTRODUCTION

As mentioned in Section 1.3.1, the therapeutic potential of CB₂ receptor modulation has prompted the development of CB₂ receptor selective ligands, either as agonists or as antagonists/inverse agonists. Several reviews have already been published to summarize the advances of new CB₂ ligands (Figure 5-1) from literature and patents [130, 157, 245-247]. This chapter discusses efforts in finding novel CB₂ selective inverse agonists.

The first CB₂ inverse agonist discovered is the 1,5-diarylpyrazole, 5-(4-chloro-3-methylphenyl)-1-[(4-methylphenyl)methyl]-N-[(1S,4R,6S)-1,5,5-trimethyl-6-bicyclo[2.2.1]-heptanyl]pyrazole-3-carboxamide (SR144528) [248]. This compound, as well as (6-iodo-2-methyl-1-(2-morpholinoethyl)-1H-indol-3-yl)(4-methoxyphenyl)methanone (AM630) [156], has been extensively used as standards against which to measure the specificity of various cannabinoid agonists at CB₂ receptors in animal models. Recently, the 2-quinolone derivative *N*-(benzo[1,3]-dioxol-5-ylmethyl)-7-methoxy-2-oxo-8-pentyloxy-1,2-dihydro-quinoline-3-carboxamide (JTE-907) [249] and the sulfone derivative *N*-[1(S)-[4-[[4-methoxy-2-[(4-methoxyphenyl)sulfonyl]phenyl]-sulfonyl]phenyl]ethyl]methanesulfonamide (Sch225336) [250] have received much attention for their immunomodulatory properties against inflammatory

disorders in which leukocyte recruitment is involved. Recently, the natural product 4-*O*-methylhonokiol (MH) [251] and several derivatives were shown to selectively target CB₂ receptors and to act as inverse agonists with anti-inflammatory and anti-osteoclastogenic properties. Similarly, JTE-907, Sch225336, MH, the pyrimidine derivative 2-[(2,4-dichlorophenyl)amino]-*N*-[(tetrahydro-2*H*-pyran-4-yl)-methyl]-4-(trifluoromethyl)-5-pyrimidinecarboxamide (GW842166X) were found to be potential promising therapeutic agents for the treatment of inflammatory and neuropathic pain [252]. More recently, it was reported that (6*a*R,10*a*R)-3-(1,1-dimethyl-butyl)-6*a*,7,10,10*a*-tetrahydro-6,6,9-trimethyl-6*H*-dibenzo[*b,d*]pyran (JWH-133) dose-dependently inhibited intravenous cocaine self-administration, cocaine-enhanced locomotion, and cocaine-enhanced accumbens extracellular dopamine in wild-type and CB₁ receptor knockout mice [253]. This result suggests that brain CB₂ receptors may be a drug target for the pharmacotherapy of drug abuse and addiction. More and more novel chemotype CB₂ ligands were discovered, and many of them showed good therapeutic potentials. Such as the natural product (*E*)- β -caryophyllene [(*E*)-BCP] was identified as a functional nonpsychoactive CB₂ receptor ligand and as a macrocyclic anti-inflammatory cannabinoid in *Cannabis* [235]. Taken together, these published studies show that CB₂ receptors are an attractive target for developing potentially therapeutic ligands.

Within a high-throughput screening research program aimed at characterizing new cannabinoid ligands, we identified compound **1** (*N,N'*-((4-(dimethylamino)phenyl)methylene)bis(2-phenylacetamide)) (PAM, Figure 5-2) as a novel chemotype with selective CB₂ activity, which showed significant CB₂ receptor affinity ($K_i = 777$ nM) and good selectivity over the CB₁ receptor (selectivity CB₁/CB₂ > 26). On the basis of this promising result, we chose compound **1** as the prototype for further SAR medicinal chemistry

studies. In this report, we have designed and synthesized a series of novel PAM derivatives. CB₂ binding activities and functional effects of these derivatives have also been investigated to define their structure-activity relationships and functionality. This led to the identification of five new derivatives (Figure 5-2) as potential CB₂ selective ligands. Importantly, some showed promising inhibition activity to osteoclast cells derived from human bone marrow. The toxicity of PAM compounds on normal human mononuclear cells was then investigated. Furthermore, we performed molecular docking studies to assist in our initial ligand design as well as 3D-QSAR CoMFA studies which reveal sterically and electrostatically favored and disfavored regions which play important roles in CB₂ receptor binding activity and selectivity of PAM derivatives.

5.2 EXPERIMENTAL AND COMPUTATIONAL METHODS

5.2.1 Chemistry

All compounds (bolded numbers) mentioned from this point can be found in Table 5-1, Table 5-2, Table 5-3, Table 5-4, and Table 5-5. Syntheses of all the compounds were performed by Dr. Peng Yang and Mr. Abdulrahman Almehezia. All reagents were purchased from commercial sources and used without further purification. Analytical thin-layer chromatography (TLC) was performed on SiO₂ plates on Alumina. Visualization was accomplished by UV irradiation at 254 nm. Preparative TLC was conducted using preparative silica gel TLC plates (1000 μm, 20cm×20cm). Flash column chromatography was performed using the Biotage Isolera flash purification system with SiO₂ 60 (particle size 0.040-0.055 mm, 230-400 mesh). ¹H NMR was

recorded on a Bruker 400 MHz spectrometer. The chemical purity of the target compounds was > 95% as determined using the following conditions: a Shimadzu HPLC-MS with a HAMILTON reversed phase column (HxSil, C18, 3 μ m, 2.1 \times 50 mm (H2)); Eluent A: 5% CH₃CN in H₂O, eluent B: 90% CH₃CN in H₂O; flow rate of 0.2 mL/min; UV detector, 254 nm and 214 nm.

5.2.2 Radioligand competition binding assays

CB ligand competition binding assay was carried out as described previously [235]. Briefly, non-radioactive (or cold) ligands (PAM derivatives and reference ligands) were diluted in binding buffer (50 mM Tris-HCl, pH 7.4, containing 5 mM MgCl₂, 2.5 mM EGTA, 0.1% (w/v) fatty acid free BSA), 10% dimethylsulfoxide, and 0.4% methyl cellulose. Each assay plate well contained a total of 200 μ L of reaction mixture comprised of 5 μ g of CB₁ (or CB₂) membrane protein, labeled [³H]-CP-55,940 at a final concentration of 3 nM and the unlabeled ligand at its varying dilutions as stated above. Plates were incubated at 30 °C for 1 h with gentle shaking. The reaction was terminated by rapid filtration through Unifilter GF/C filter plates using a Unifilter Cell Harvester (PerkinElmer). After the plate was allowed to dry overnight, 30 μ L MicroScint-0 cocktail (PerkinElmer) was added to each well and the radioactivity was counted by using a PerkinElmer TopCounter. Data from these assays were analyzed using GraphPad Prism 5.0 Software. All assays were performed in duplicate and data points represented as mean \pm S.E.M. Bound radioactivity was analyzed for K_i values using non-linear regression analysis by GraphPad Prism 5.0 software.

The saturation binding of [³H]-CP-55,940 to the membrane proteins was performed as described previously [237]. Briefly, the CB₁ (or CB₂) membrane fractions (5 μ g) were incubated

with increasing concentrations of [³H]-CP-55,940 (0.05-4 nM) in 96-well plates at 30 °C with slow shaking for 1 h. The incubation buffer was composed of 50 mM Tris-HCl, pH 7.4, containing 5 mM MgCl₂, 2.5 mM EGTA, and 0.1% (w/v) fatty acid free BSA. Ligand was diluted in incubation buffer supplemented with 10% dimethylsulfoxide and 0.4% methyl cellulose. Non-specific binding was determined in the presence of 1:1000 unlabeled CP-55,940 (5,000 nM) in excess. The reaction was terminated and the radioactivity was counted as stated above. The difference between total and nonspecific binding equals the receptor specific binding. Non-linear regression analysis revealed the receptor density (B_{max}) and the equilibrium dissociation constant (K_d) values of [³H]-CP-55,940 for the CB₂ receptor.

5.2.3 cAMP assays

cAMP levels were measured according to reported method with modifications using LANCE cAMP 384 kits (PerkinElmer) [237]. The assay is based on competition between a Europium-labeled cAMP trace complex and total cAMP for binding sites on cAMP-specific antibodies labeled with a fluorescent dye. The energy emitted from the Eu-chelate is transferred to the dye on the antibodies, which in turn generates a time-resolved fluorescent resonant energy transfer (TR-FRET) signal at 665 nm. The fluorescence intensity (665 nm) decreases in the presence of cAMP from the tested samples and resulting signals are inversely proportional to the cAMP concentration of a sample. CB₂ receptor wild type (WT) transfected CHO cells were seeded in a 384-well white ProxiPlates with a density of 2000 cells per well in 5 µL of RPMI-1640 medium containing 1% dialyzed FBS, 25 mM HEPES, 100 µg/mL penicilin, 100 U/ml streptomycin and 200 µg/mL of G-418. After culture overnight, 2.5 µL of cAMP antibody and RO20-1724 (final concentration 50 µM) in stimulation buffer (DPBS 1x, containing 0.1% BSA) was added to each

well, followed by addition of either 2.5 μ L compound or forskolin (final 5 μ M) for agonist-inhibited adenylate cyclase (AC) activity assay. After incubated at room temperature for 45 min, 10 μ L of detection reagent was added into each well. The plate was then incubated for 1 h at room temperature and measured in Synergy H1 hybrid reader (BioTek) with excitation at 340 nm and emission at 665 nm. Each cAMP determination was made via at least two independent experiments, each in triplicate. EC₅₀ values were determined by nonlinear regression, dose-response curves (GraphPad Prism 5).

5.2.4 Osteoclast formation assay

Human marrow-derived mononuclear cells (2×10^5 cells/well) were seeded in 96-well multiplates at 100 μ L/well in α -MEM containing 20% horse serum, 10 ng/mL M-CSF, and 25 ng/mL RANKL. The tested compounds at the indicated final concentrations were added to the appropriate wells. Half-media changes were carried out twice a week using drug-containing media where appropriate. The culture was incubated for a total of 3 weeks at 37 °C with 5% CO₂ and 95% humidity. Differentiation into OCLs was assessed by staining with monoclonal antibody 23c6 using a Vectastatin-ABC-AP kit (Vector Laboratories, Burlingame, CA). The antibody 23c6, which recognizes CD51/61 dimer constituting the OCL vitronectin receptor, was generously provided by Michael Horton (Rayne Institute, Bone and Mineral Center, London, UK). The 23c6-positive multinucleated OCLs containing 3 or more nuclei per OCL were scored using an inverted microscope [254]. This assay was performed by Dr. Rentian Feng.

5.2.5 Cytotoxicity assay on human mononuclear cells

Peripheral blood was drawn in a heparinized syringe from healthy fasting volunteers who had been without medication for at least 2 weeks. The Peripheral blood mononuclear cell (PBMC) fraction was obtained by gradient centrifugation over Ficoll-Hypaque (Amersham), as described previously [255]. PBMC were washed three times with ice-cold PBS, followed by resuspension at 5×10^5 /mL in the culture medium supplemented with 10% inactivated FBS, 2 mM glutamine, 100 U/mL penicillin and 100 μ g/mL streptomycin (Sigma). The compounds in a stock solution (50 mM in DMSO) was diluted with the culture medium to application conditions and further used for the treatment of PBMC for three days. The final DMSO concentrations were 0.02%. After treatment for 72 h, cell viability was determined using trypan blue exclusion assay. These human cell studies conformed to the guidelines of the Institutional Review Board of the University of Pittsburgh. This assay was performed by Dr. Rentian Feng.

5.2.6 Molecular docking and 3D-QSAR model development

Molecular docking studies were carried out initially to aid in the early design of PAM analogs. The Xie group previously published a CB2 homology model using bovine rhodopsin as a template in 2003 [128]. In order to reflect and include more recent GPCR crystal structure information, our CB2 homology was modified using other known GPCR crystal structures, including antagonist-bound A2A receptor (3EML.pdb) [256], human dopamine D3 receptor in complex with eticlopride (3PBL.pdb) [257], bovine rhodopsin (1L9H.pdb and 1F88.pdb) [116, 258], human beta2-adrenergic receptor (2RH1.pdb) [259], turkey beta1-adrenergic receptor

(2Y00.pdb and 2VT4.pdb) [260, 261] using the I-TASSER program [262, 263]. The best predicted model with the highest C-score was selected and further minimized using a conjugate gradient method with the AMBER force field (AMBER7 FF99) defined in Tripos Sybyl software [77].

Once the final model was obtained, further molecular docking simulations were performed using the Surflex-Dock program [77] to predict putative binding pocket and potential ligand-receptor interactions such as hydrophobic interactions, hydrogen-bonding and pi-pi interactions. To predict a putative binding pocket, MOLCAD analysis [77] was performed on the CB2 model to find a solvent-accessible cavity around important binding residues such as SER90, PHE91, PHE94, PHE106, LEU108, ILE110, SER112, VAL113, MET115, PHE117, PRO168, LEU169, TRP172, LEU182, TYR190, TRP194, PHE197, TRP258, SER268, LEU269 and SER285 reported in the literature [264-268]. During the docking simulation, the receptor structure was fixed while allowing only the ligand to be flexible. The docking mode was *GeomX* and other default docking parameters were used. The binding interaction energy was calculated to include van der Waals, electrostatic, and torsional energy terms defined in the Tripos force field [77].

Out of the 52 compounds from Tables 1-5, 40 compounds were used in the subsequent 3D QSAR CoMFA studies. Twelve compounds that showed no binding, hence no experimental K_i , were ignored in the analysis. Approximately 75% (29 compounds) and 25% (11 compounds) were randomly selected as a training set and a test set respectively. Molecular modeling and CoMFA studies were performed using the Sybyl-X1.2 from Tripos Molecular Modeling package [269]. Using our established protocol [5, 270, 271], molecular dynamic simulations were carried out for the best compound **17**. Briefly, dynamic simulations were simulated at a temperature of

300K with time steps of 1 fs for a total duration of 300 ps, and conformation samples were collected at every 1 ps resulting in 300 conformers of compound **17**. All conformers were then minimized and converged into four families. These four representative conformers derived from MD simulations were compared to the docking pose resulted from the molecular docking experiment using our in-house 3D CB₂ receptor model. The docking experiment was done using the Surflex-Dock module from Tripos modeling software. The conformer with maximum agreement between these two experiments was chosen as a preferred conformer for further CoMFA studies. Structural alignments of all molecules in the training and test sets to the preferred conformer of compound **17** were performed using the MultiFit program in SybylX1.2. The CoMFA study was then carried out using the Sybyl/CoMFA module. The steric and electrostatic field energies (Gasteiger-Hückel charge) were calculated using the default parameters, namely: the Tripos Standard CoMFA field class, distance-dependent dielectric constant, steric and electrostatic field value cutoff set at 30 kcal·mol⁻¹. Leave one-out cross-validation (LOOCV) Partial Least-Square (PLS) analysis was then performed with a minimum σ (column filter) value of 5.0 kcal·mol⁻¹ to improve the signal-to-noise ratio by omitting those lattice points whose energy variation was below this threshold. The final model (non cross-validated analysis) was developed from the LOOCV model with the highest cross-validated r^2 value, using the optimal number of components determined by the LOOCV model.

5.3 RESULTS AND DISCUSSION

5.3.1 Pharmacology and SAR analysis

The binding affinities of these 52 derivatives to CB₂ receptor were determined by performing [³H]CP-55,940 radioligand competition binding assays using membrane proteins of the CHO cells stably expressing human CB₂ receptor. The CB₁ binding assay was also conducted for those compounds with high CB₂ receptor binding potency ($K_i < 1,000$ nM) using membrane proteins harvested from the CHO cells stably transfected with the human CB₁ receptors. CB₂ receptor ligand SR144528 and CB₁ ligand SR141716 were used as positive controls respectively along with the tested compounds. Structures, physiochemical properties and binding activities are shown in Table 5-1, Table 5-2, Table 5-3, Table 5-4, and Table 5-5.

First, the SAR study was focused on the effect of the side chains on aromatic ring C. A set of twenty-one compounds was synthesized (**2-22**, Table 5-1). The aromatic ring C was modified with substituents that varied in their size, electronic character, and position. Removal of the *p*-dimethylamino group (compound **2**, CB₂ $K_i = 9930$ nM) dramatically decreased the CB₂ binding activity. Introducing fluorine atoms to different positions of ring C also lowered the CB₂ receptor affinity (compound **3-5**, CB₂ $K_i = 35330, 12670$ and 10900 nM, respectively). The CB₂ receptor binding affinities of the F-substituted compounds decreased in the order of *o*-F < *m*-F < *p*-F. From these results, we deduced that substitution at the *para*-position of the phenyl ring may play an important role in the CB₂ receptor binding activity. This deduction was also confirmed by compounds **13** and **14**. Compound **14** bearing *p*-trifluoromethyl showed improved activity (**14**, CB₂ $K_i = 596$ nM), while compound **13** bearing *o*-trifluoromethyl showed dramatically decreased activity (**13**, CB₂ $K_i = 11780$ nM). Moreover, replacing the *para*-fluorine with chlorine

(compound **6**) or bromine (compound **7**) relatively increased the activity, but the binding affinities were still weak. While introduction of a methyl group to the *para*-position (compound **8**) improved CB₂ receptor affinity, unfortunately, compound **8** also had high affinity for the CB₁ receptor; the only compound that exhibited significant CB₁ receptor binding activity among the 52 compounds (**8**, CB₁ K_i = 109 nM; CB₂ K_i = 494 nM). Replacement of *p*-dimethylamino group with bioisostere isopropyl (compound **9**) dramatically improved the binding activity and selectivity (CB₂ K_i = 85 nM, selectivity index > 235).

As for the compounds with alkoxy groups (compounds **10-12**), compound bearing methoxyl (**10**) showed similar activity and selectivity compared to the parent compounds, compound bearing ethoxyl (**11**) showed slightly decreased activity, and compound bearing isopropoxyl (**12**) showed slightly increased activity. This result indicated a variety of alkoxy groups was tolerated, but their activity and selectivity for the CB₂ receptor were sensitive to the group size. To explore the electronic and steric effects on CB₂ binding activity, we introduced a nitro group to the benzene ring (**15**), but compound **15** completely lost its binding affinity to CB₂. Reduction of compound **15** to the corresponding amine resulted in compound **16**, which displayed relatively improved activity but was still weak. Replacement of the amine with a diethylamine group, however, resulted in a promising compound (**17**), which showed much improved activity and selectivity compared to the lead compound (CB₂ K_i = 64 nM, selectivity index > 313). When the *p*-diethylamino group was identified as a better chemical group on ring C, additional substituted amino groups were further studied, resulting in several potent compounds **18-22** with *p*-dipropylamino, *p*-dibutylamino, *p*-pyrrolidinyl, *p*-piperidyl and *p*-dibenzylamino respectively. Compared with the lead compound **1**, these five compounds showed

greatly improved activity and selectivity (CB₂ K_i: 22-595 nM, selectivity index: 34-909). When compared to compound **17** bearing a diethylamine group, compound **18** with a *p*-dipropylamino group showed the most potential binding affinity and selectivity (CB₂ K_i = 22 nM, selectivity index > 909). Compound **21** with a *p*-pyrrolidinyl group showed similar activity (CB₂ K_i = 71 nM, selectivity index > 281). The modification result showed that CB₂ binding affinity decreased as the size of the functional group at the *para* position of the benzene ring C increased (compound **19**, **20** and **22**). Hence, we conclude that the substituted amino group at the *para* position plays a significant role in CB₂ receptor binding activity and the *p*-dipropylamino group is optimal.

Subsequently, the SAR was further explored on the variation on aromatic rings A and B by introducing Cl or CF₃, resulting in two series of compounds: **23-31** and **32-38**. Among the first series compounds bearing Cl on ring A and B (**23-31**), five compounds (**26-30**) showed increased CB₂ binding affinity and selectivity. All the compounds with CF₃ on ring A and B in the second series (**32-38**) showed no binding activity to CB₂ receptors. The results indicated that *para*-Cl is a better substituent than CF₃ and H on rings A and B. Comparison of compound **29** with *o*-CF₃ and compound **30** with *p*-CF₃ further indicates that the *para*-position of the phenyl ring C plays an important role in the CB₂ binding activity.

The distance from ring C to the methylene amide group as well as from rings A and B to the amide group was also explored (compound **39-41**, Table 5-2; compound **42-44**, Table 5-3). The data indicated that inserting CH₂ (compound **39**), CH₂CH₂ (compound **40**) or CH=CH double bond (compound **41**) between ring C and methylene amide group resulted in a complete

loss of activity or weak binding affinity. While removing CH₂ from compound **17** or inserting CH₂CH₂ (compound **43**) or CH=CH double bond (compound **44**) between rings A/B and the methylene amide group led to a slight decrease in binding affinity, these compounds still showed good CB₂ binding affinity and selectivity ($167 \leq \text{CB}_2 K_i \leq 688 \text{ nM}$; $29 \leq \text{selectivity index} \leq 119$).

Furthermore, the importance of aromatic ring C in the CB₂ binding activity was explored (compound **45** and **46**, Table 5-4). Replacing ring C with alkyl chain butyl (**40**) or pentyl (**41**) group led to a complete loss of activity or very weak binding affinity. We conclude that the aromatic ring C plays a significant role in CB₂ receptor binding affinity and may be an essential element to retain the activity.

After discovering the importance of the aromatic ring C for CB₂ binding affinity, we then explored the importance of rings A and B by replacing aromatic rings A/B with different alkyl chains (compound **47-52**, Table 5-5). The results indicated that replacing the benzyl group with a branched chain isopropyl (compound **47**) or *tert*-butyl group (compound **48**) dramatically decreased the CB₂ binding affinity, whereas replacing the benzyl group with a long alkyl chain butyl (compound **49**) showed slightly decreased affinity. Interestingly, replacement of benzyl with the straight chain pentyl group led to another promising compound **50**, which showed greatly improved binding affinity and selectivity ($\text{CB}_2 K_i = 25 \text{ nM}$, selectivity index > 800). To further explore the effect of the alkyl chain, we also replaced aromatic rings A/B with longer chains n-C₇H₁₅ (compound **51**) and n-C₉H₁₉ (compound **52**). Compared to compound **50**, however, they both showed slightly decreased binding affinity (**51**, $\text{CB}_2 K_i = 146 \text{ nM}$, selectivity

index > 136; **52**, CB₂ K_i = 160 nM, selectivity index > 125). From these results, we conclude that the aromatic ring A/B may be replaced by alkyl chain and the pentyl group is optimal.

5.3.2 Functional bioactivity at CB₂ receptors in vitro

Cellular bioassay was also carried out by using our published protocol [237] in order to measure the agonistic or antagonistic activities of the CB₂ selective compounds. Briefly, the cell-based LANCE cAMP assays were performed on 384 well plates using CHO cells stably expressing the CB₂ receptors in the presence of phosphodiesterase inhibitor RO20-1724 and adenylyl cyclase activator forskolin. Since CB₂ is a G_{αi}-coupled receptor, an agonist inhibits the forskolin-induced cAMP production, resulting in an increase of the LANCE signal. On the other hand, an antagonist or inverse agonist decreases the LANCE signal toward forskolin-induced cAMP accumulation. Therefore, the detected LANCE signal is inversely proportional to cAMP level. As shown in Figure 5-3, reduction of the LANCE signal occurred with increasing concentrations of compounds **1**, **9**, **17**, **18**, **21**, **50** and SR144528. These ligands acted as inverse agonists, indicated by increasing forskolin-induced cAMP production, with EC₅₀ values of 159.1±8.68 nM, 4.11±3.66 nM, 5.73±6.37 nM, 28.33±2.54 nM, 17.08±2.11 nM, 13.42±2.07 nM and 13.71±2.81 nM, respectively. Such a phenomenon was not observed with agonists CP55940 and HU308, which inhibited cAMP production with EC₅₀ values of 23.29±4.17 nM and 83.81±5.63 nM, respectively. The results clearly indicated that six compounds (**1**, **9**, **17**, **18**, **21** and **50**) indeed behaved as inverse agonists.

5.3.3 Osteoclast Formation Assay

Based on binding affinity, selectivity and drug-likeness, four compounds were selected as candidates for further study. As shown in Figure 5-8, we tested the effect of these most promising CB₂ ligands on osteoclast (OCL) formation using human nonadherent mononuclear bone marrow cells [272]. Each ligand tested induced a concentration-dependent inhibition of osteoclastogenesis. Compared with the known CB₂ inverse agonist SR144528, our compounds exhibited the same or stronger potency in suppressing OCL formation. Especially, compound **17** showed the strongest inhibition activity, with inhibition rates of 72%, 79% and 84% at 0.1 μM, 1 μM and 10 μM, respectively. Importantly, **17** showed a more potent inhibitory effect than the parent ligand Xie95 (PAM), suggesting our medicinal chemistry modification and SAR studies of Xie95 led to overall improved compounds not only for CB₂ activity but also for osteoclastogenesis inhibition.

5.3.4 Cytotoxicity analysis using normal human cells

Our newly discovered compounds showed promising inhibition activity with respect to osteoclastogenesis. To examine whether the impaired osteoclastogenesis in the presence of PAM compounds is due to their toxicity, we investigated the cytotoxicity profile of PAM compounds on normal human cells. First, mononuclear cells were isolated from healthy donors. After treatment of these normal cells with the PAM compounds for 3 days, it was shown that the cell viability was not significantly affected compared with the vehicle control group (Figure 5-9). The best compound **17** did not show any cytotoxic effects at the concentration (1 μM) of 79% inhibition of osteoclastogenesis, and only slight effects on cell viability were observed at high

concentration of 10 μ M. These results indicate that our compounds possess favorable therapeutic indexes and the inhibition of human osteoclastogenesis is not a result of their cytotoxicity.

5.3.5 Molecular docking and 3D-QSAR studies

First of all, a putative binding pocket (Figure 5-4) was predicted via MOLCAD analysis [77] to find a solvent-accessible cavity around known important binding residues such as SER90, PHE91, PHE94, PHE106, LEU108, ILE110, SER112, VAL113, MET115, PHE117, PRO168, LEU169, TRP172, LEU182, TYR190, TRP194, PHE197, TRP258, SER268, LEU269 and SER285 which were previously reported in the literature [264-268]. Then known CB ligands such as CP55940 (agonist) and SR144528 (inverse agonist) as well as our compound **1** and **17** (inverse agonists) were docked to the binding pocket using the Surflex-Dock program [77]. Docking results of each ligand can be seen in Figure 5-5. Several ligand-receptor interactions were observed. For example, CP55940 showed hydrogen-bonding with ASN188 and HIS95 as well as hydrophobic interactions with LEU196, VAL261, LEU201, TRP258, VAL113. Similarly, the bulky group of SR144528 had interactions with other hydrophobic residues such as LEU196, VAL261, LEU201, TRP258 and VAL113. Compound **1** and **17** also had hydrophobic interactions with LEU192, LEU196, VAL261, LEU201, TRP258 and VAL113 in addition to a hydrogen bond with SER285. From such docking results and our previous CoMFA modeling results [5], we concluded that introducing a hydrophobic group to one of aromatic rings of PAM analogs while keeping the hydrogen bond donor/acceptor groups may help improve the binding affinity and selectivity of the analogs.

A series of PAM analogs were then synthesized and 3D QSAR studies were carried out to derive CB₂ CoMFA models by using our published protocol [5, 271]. To search for the

preferred conformations of the most potent CB₂ ligand (compound **17**), molecular dynamic simulations and molecular mechanics (MD/MM) were carried out based on our established computational protocol [270]. As described in the Section 5.2.6, MD simulations were performed with time steps of 1 fs for 300 ps with 1 ps interval recording time, which resulted in 300 conformers sampled after the simulations. All 300 conformations were minimized and converged to four families. Among four representative MD-generated conformers, one conformer had the most similar conformation to the docking pose (Figure 5-5) that resulted from the molecular docking simulation using the CB₂ homology model. The conformer was chosen as one of the preferred active templates and then all compounds from the training and test data sets were aligned to such preferred conformer of compound **17**. The final alignments of each set are depicted in Figure 5-6 (A, B).

After molecular alignment, leave-one-out cross-validation (LOOCV) analysis was performed to determine the optimal number of components and to evaluate the predictive ability of the derived CoMFA model which was measured by a cross-validated r^2 (r_{cv}^2) value. It is defined as:

$$\text{Cross-validated } r^2 = (\text{SD-PRESS})/\text{SD}$$

where SD is the sum of the squared deviations of each biological property value from their mean, and PRESS is the sum, over all compounds, of the squared differences between the actual and predicted biological activity values. The LOOCV analysis showed that the optimal number of components was 4 and the r_{cv}^2 was 0.52, which was within the range of generally accepted criterion for statistical validity.

Subsequently, non-cross-validated PLS analysis was performed and an r^2 value of 0.92 with a standard error of estimate of 0.28 was obtained. Such a result indicates the trained

CoMFA model successfully correlates between PAM analog structures and their CB₂ receptor affinity values. In order to evaluate the derived CoMFA model's generalization ability on unseen compounds, it was used to predict the CB₂ binding activity values of test set compounds which were separated from the training set and hence were not included during the model training. A good correlation coefficient (r^2) value of 0.76 was obtained from such prediction and the result demonstrated that the CoMFA model had a good generalization performance on the test set compounds. As shown in Table 5-6, the predicted pK_i values are close to the experimental pK_i values for molecules in both training and test sets. Figure 5-7 shows the relationship between the calculated and experimental pK_i values for the non-cross-validated training set predictions as well as for the test set predictions. The linearity of the plot indicates a very good correlation and predictive ability of the developed CoMFA model to predict CB₂ receptor binding affinities of PAM derivatives, which is currently used in our new CB₂ ligand design.

To further predict favorable and unfavorable regions of PAM derivatives for CB₂ receptor binding activity, CoMFA contour maps were derived. In particular, CoMFA contour maps depict the color-coded regions around the molecules which associate with ligand biological activities. Green regions indicate favorable steric interactions that enhance binding affinity whereas yellow regions display unfavorable steric interactions. On the other hand, the blue and red regions show preferred and not-preferred electrostatic interactions respectively. As shown in Figure 5-6 (C), there is a sterically preferred region near the *p*-dimethylamino group, which means the hydrophobic pharmacophore feature in this part of the molecule is expected to enhance CB₂ receptor binding affinity. In fact, such a hydrophobic moiety may interact and fit well in the previously suggested hydrophobic pocket within transmembrane regions 3, 5, 6 and 7 [237, 273-275]. Moreover, this finding is consistent with our previous CoMFA studies [5], which

showed that the presence of a steric bulky group enhanced the CB₂ receptor binding activity and selectivity. It also agreed with our initial docking results which suggested introducing a hydrophobic group to PAM analogs. On the other hand, electrostatic interactions are not preferred near the *p*-dimethylamino group as highlighted by a red region (Figure 5-6 (C)). This is congruent with the chemistry modifications of compounds **16** and **31** with *para* -NH₂ and -NO₂ groups respectively, which lost CB₂ binding activity. However, once a hydrophobic feature was reintroduced, the CB₂ affinity and selectivity were restored as demonstrated by compounds **8**, **9**, **12**, **17**, **18**, **19**, **20**, **21**, and **22**. Such CoMFA results confirmed our SAR hypothesis that aromatic ring C plays an important role in CB₂ receptor binding activity and introducing a hydrophobic feature at the *para* position of ring C is expected to enhance CB₂ receptor activity and selectivity.

5.4 CONCLUSION

We reported PAM as a novel chemotype with selective CB₂ receptor binding activity. In continuation of our studies on PAM derivatives, we have synthesized 52 new PAM derivatives designed through variations of the aromatic rings A-C and the substituents of different positions on these three rings. The SAR analysis reveals that: (i) the *para*-substituted amino group on ring C plays a significant role in CB₂ receptor binding activity, a variety of functional groups was tolerated, and the *p*-dipropylamino group is optimal; (ii) *para*-Cl is a much better substituent than CF₃ and H on rings A and B, and aromatic rings A/B may be replaced by alkyl chains with the pentyl group being optimal; and (iii) aromatic ring C is an essential element to retain compound potency to CB₂. Among the derivatives, five compounds **9**, **17**, **18**, **21** and **50** were confirmed as CB₂ inverse agonists with the strongest CB₂ receptor binding affinity and best

selectivity. 3D-QSAR CoMFA modeling studies also confirmed our SAR findings that aromatic ring C is important for CB₂ receptor activity and a hydrophobic feature at the ring C's *para* position is crucial to improve CB₂ activity and selectivity of the PAM analogs. The results were congruent by chemistry, bioassay validation and computer modeling studies. More importantly, osteoclastogenesis assay indicated that PAM compounds have promising inhibition activity to osteoclast cells derived human bone marrow. The most promising compound, **17**, showed 72% inhibition activity even at the low concentration of 0.1 μM. The inhibition of human osteoclastogenesis is not due to cytotoxic effects. Therefore, these PAM derivatives could be used as potential leads for the development of a new type of antiosteoporosis agent. Overall, the data presented here show that PAM is a useful scaffold for the design of new selective CB₂ receptor inverse agonists for further CB₂ signaling and anti-osteoclast studies.

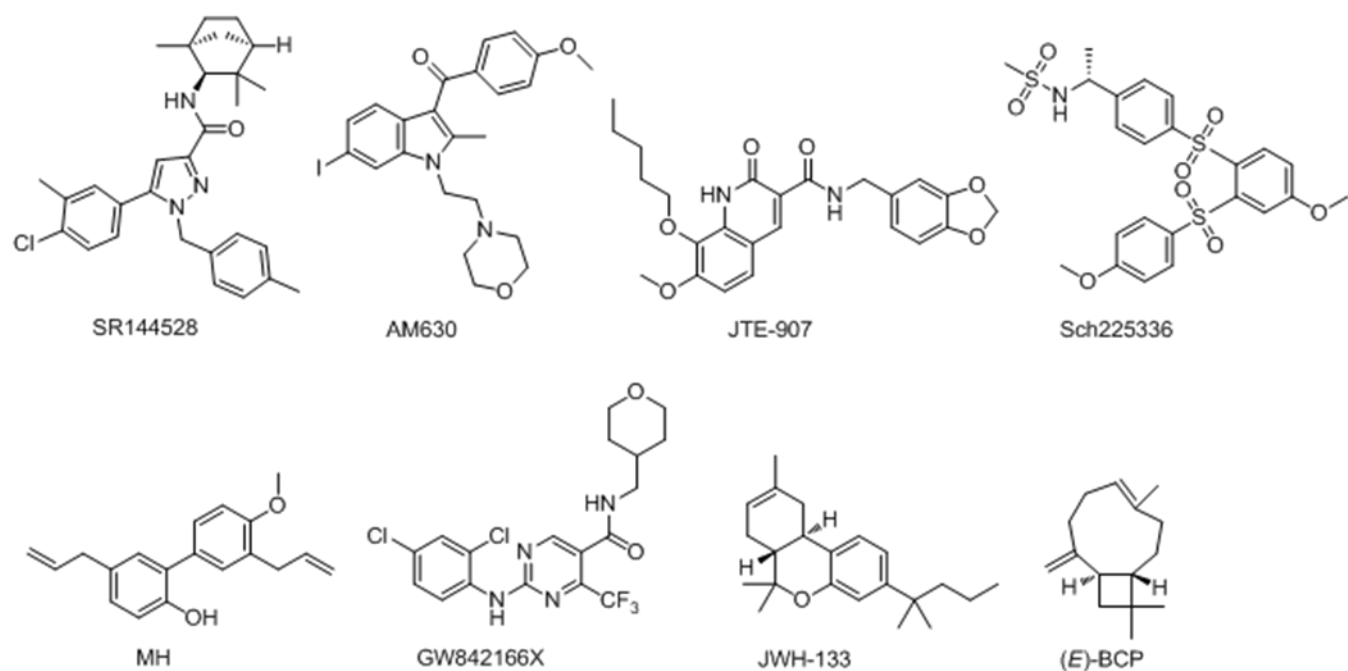
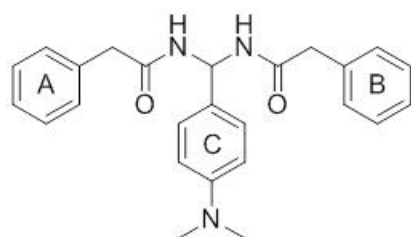
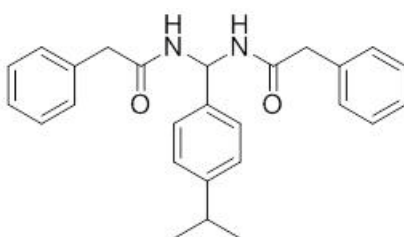


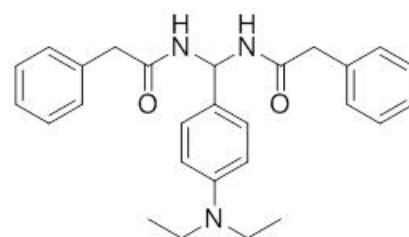
Figure 5-1. Representative CB₂ selective compounds with various chemical scaffolds.



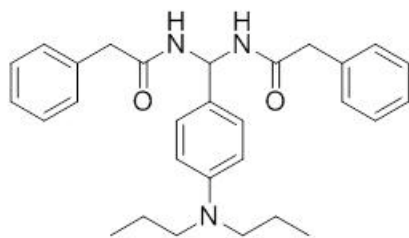
Compound **1**:
 $CB_2 K_i = 777 \text{ nM}$
 $CB_1 K_i > 20,000 \text{ nM}$
 Selectivity $CB_1/CB_2 > 26$



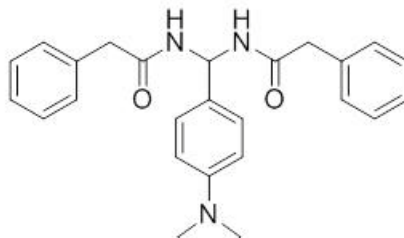
Compound **9**:
 $EC_{50} = 4 \text{ nM}$, $CB_2 K_i = 85 \text{ nM}$
 $CB_1 K_i > 20,000 \text{ nM}$
 Selectivity $CB_1/CB_2 > 235$



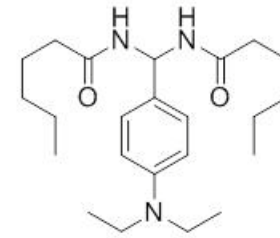
Compound **17**:
 $EC_{50} = 5 \text{ nM}$, $CB_2 K_i = 64 \text{ nM}$
 $CB_1 K_i > 20,000 \text{ nM}$
 Selectivity $CB_1/CB_2 > 313$



Compound **18**:
 $EC_{50} = 28 \text{ nM}$, $CB_2 K_i = 22 \text{ nM}$
 $CB_1 K_i > 20,000 \text{ nM}$
 Selectivity $CB_1/CB_2 > 909$



Compound **21**:
 $EC_{50} = 17 \text{ nM}$, $CB_2 K_i = 71 \text{ nM}$
 $CB_1 K_i > 20,000 \text{ nM}$
 Selectivity $CB_1/CB_2 > 281$



Compound **50**:
 $EC_{50} = 13 \text{ nM}$, $CB_2 K_i = 25 \text{ nM}$
 $CB_1 K_i > 20,000 \text{ nM}$
 Selectivity $CB_1/CB_2 > 800$

Figure 5-2. Structures of the lead compound **1** (or XIE95 or PAM) and the modified target compounds **9**, **17**, **18**, **21** and **50**.

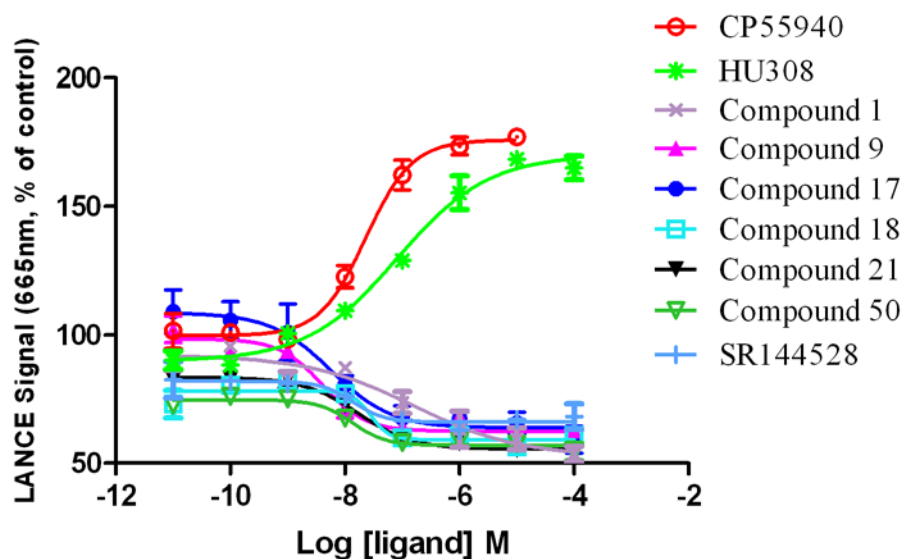


Figure 5-3. Comparisons of LANCE signal of different CB₂ receptor ligands in stably transfected CHO cells expressing human CB₂ receptors in a concentration-dependent fashion.

EC₅₀ values of compounds 1, 9, 17, 18, 21, 50 and SR144528 are 159.1±8.68 nM, 4.11±3.66 nM, 5.73±6.37 nM, 28.33±2.54 nM, 17.08±2.11 nM, 13.42±2.07 nM and 13.7±2.81 nM respectively. EC₅₀ for CP-55,940 and HU308 are 23.29±4.17 nM and 83.81±5.63 nM. Data are mean±S.E.M. of one representative experiment of two or more performed in duplicate or triplicate.

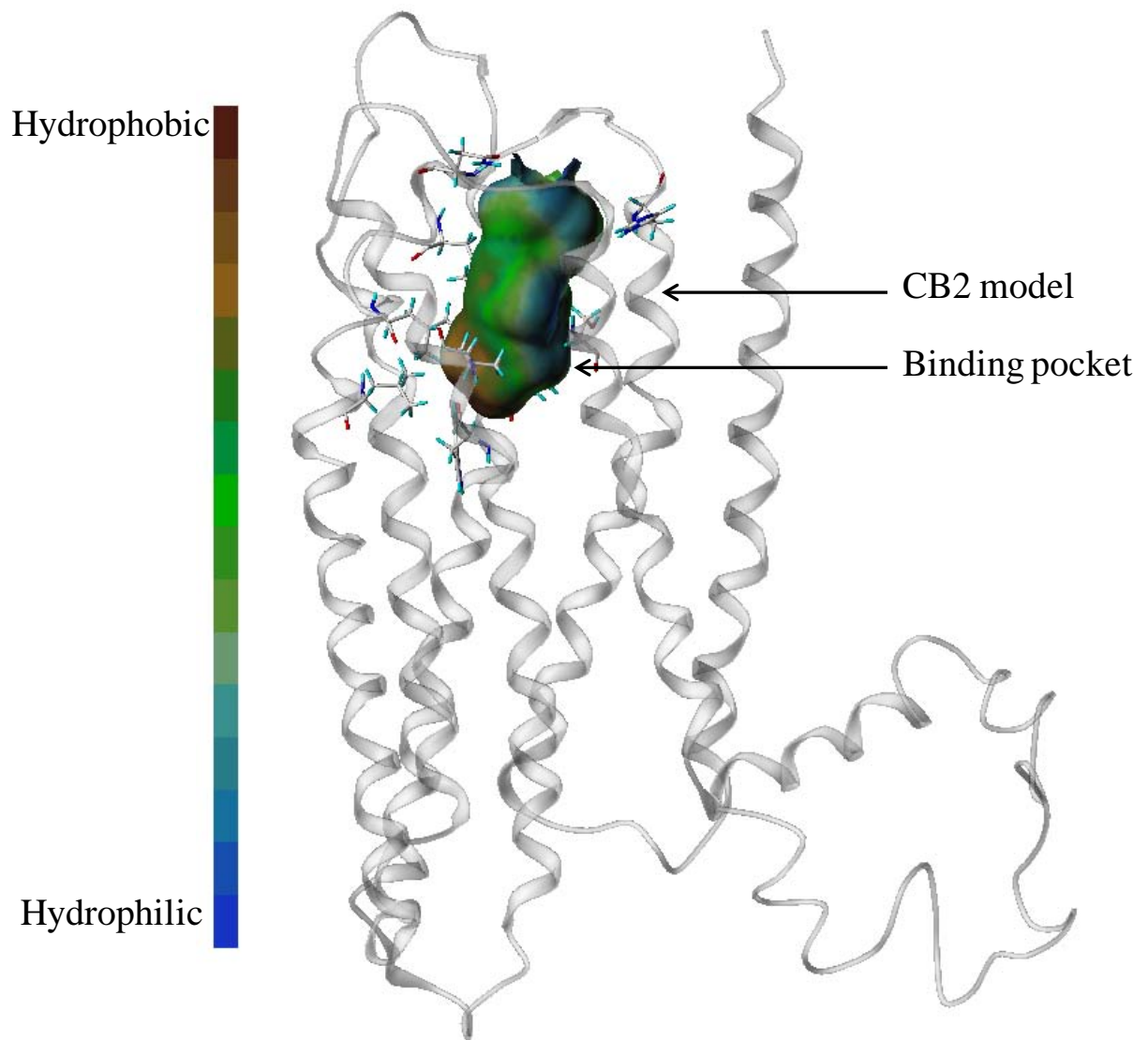


Figure 5-4. A predicted binding pocket around important binding residues of the CB2 receptor

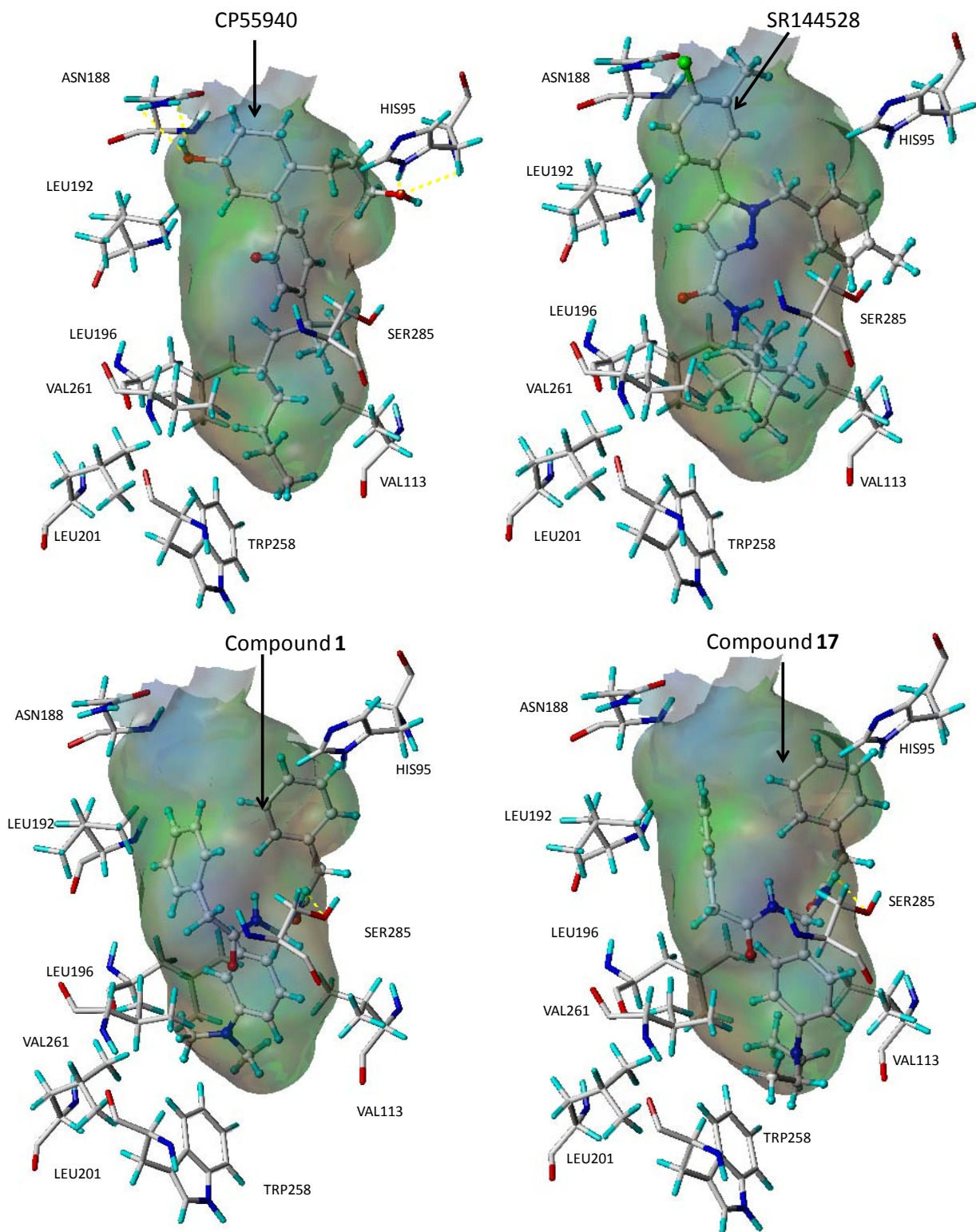


Figure 5-5. CP55940, SR144528, compound **1** and compound **17** (ball and stick models) docked to the CB2 binding pocket displaying hydrogen-bonding and hydrophobic interactions with the receptor

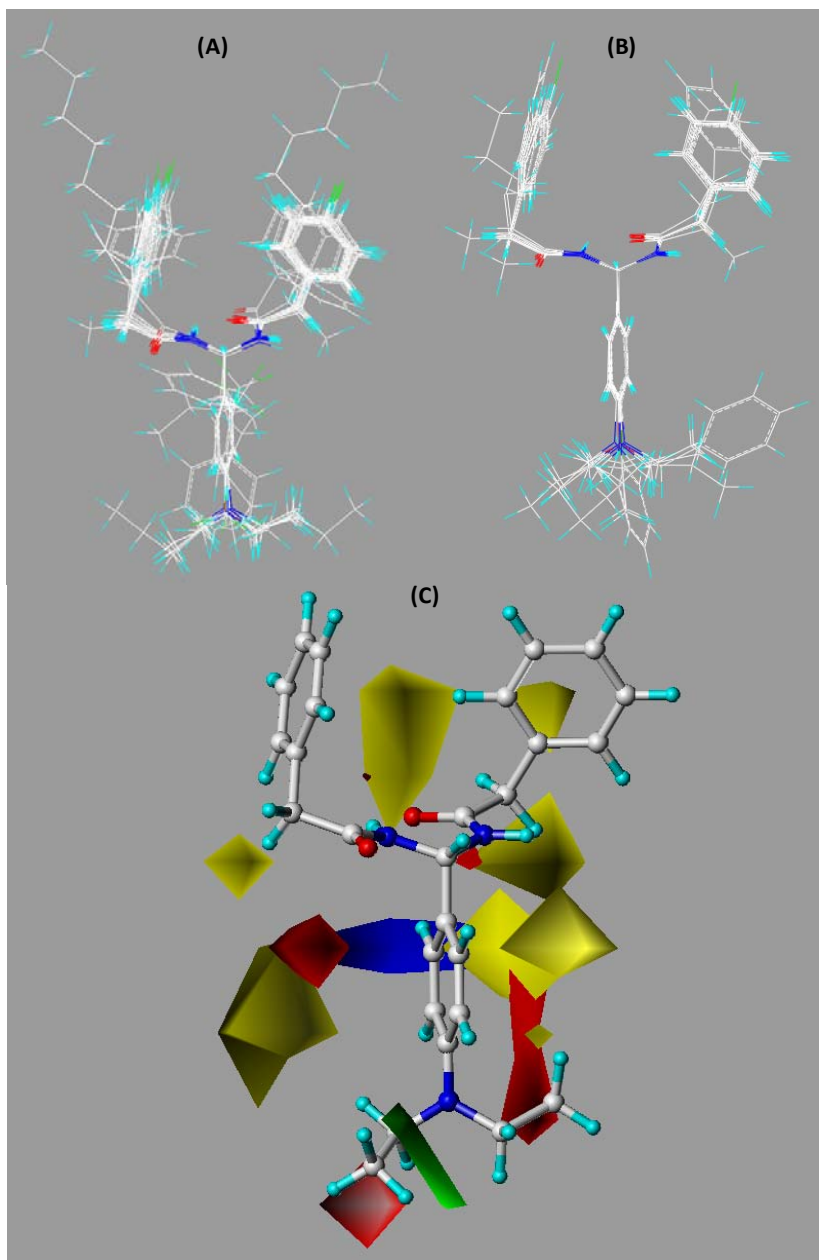


Figure 5-6. Overall alignments of training set molecules (A) and test set molecules (B) to the compound **17** as well as CoMFA contour maps of compound 17 showing steric and electrostatic (C) interactions.

Sterically (bulk) favored areas are color-coded in green and sterically disfavored areas are in yellow. Electrostatically (charge) preferred regions are in blue and red regions are electrostatically disfavored areas.

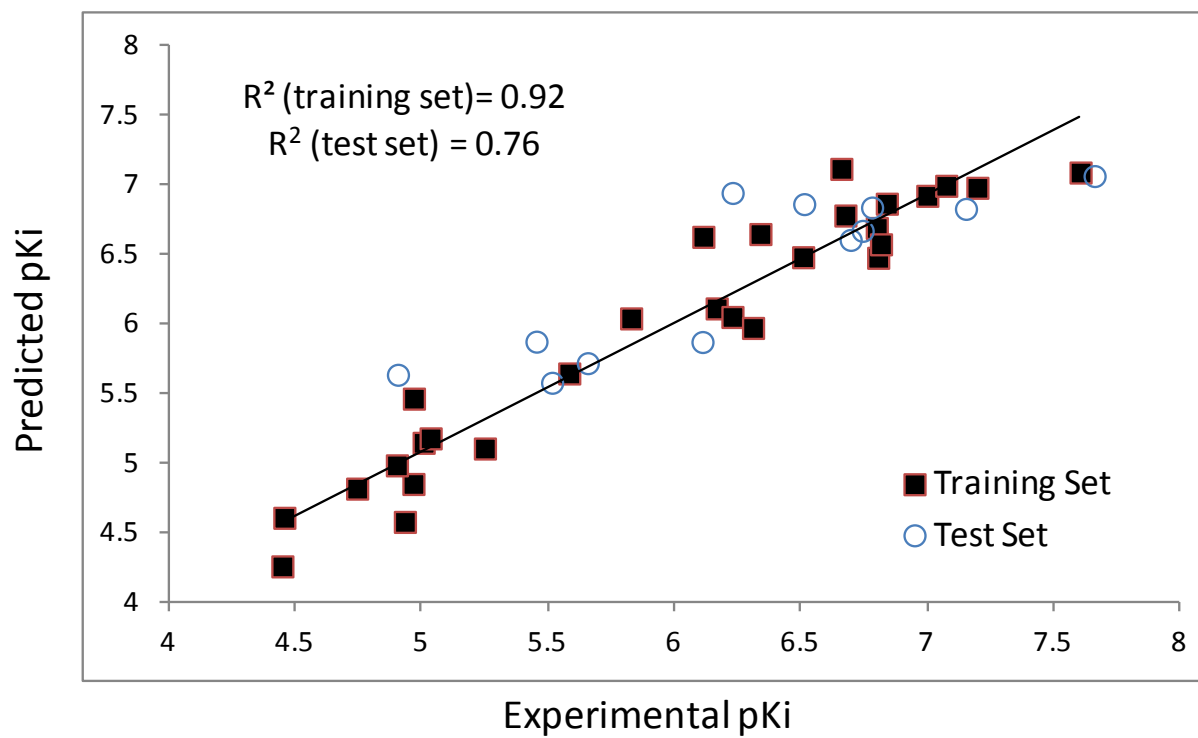


Figure 5-7. Scatter plots of CoMFA-calculated and experimental binding affinity values (pKi) for the training and test sets.

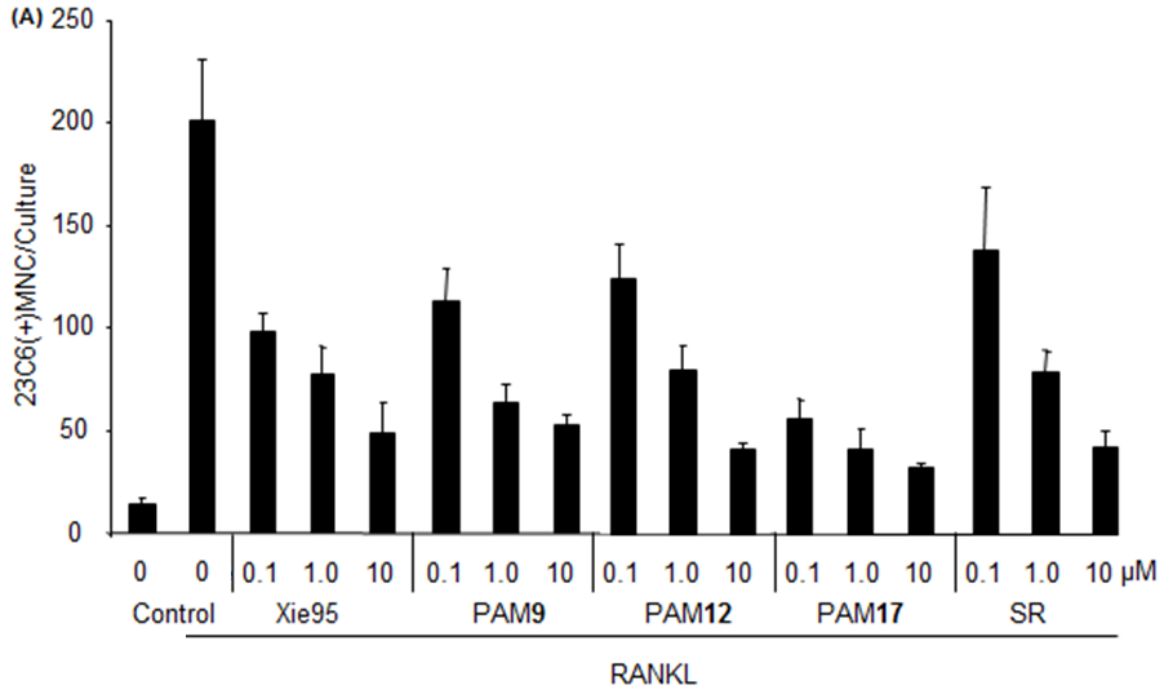


Figure 5-8 Inhibition of human osteoclastogenesis by CB2 ligands.

Human bone marrow-derived mononuclear cells were cultured in a 96-well plate for three weeks in the presence of RANKL (50 ng/mL) to form osteoclast-like cells. After 3 weeks, the cultures were stained with the 23c6 antibody. 23c6-positive OCLs containing 3 or more nuclei were scored microscopically. All experiments were performed in triplicate. Results are shown as mean \pm SD. SR = SR144528. Note: the control on left is vehicle control and the right one is positive control.

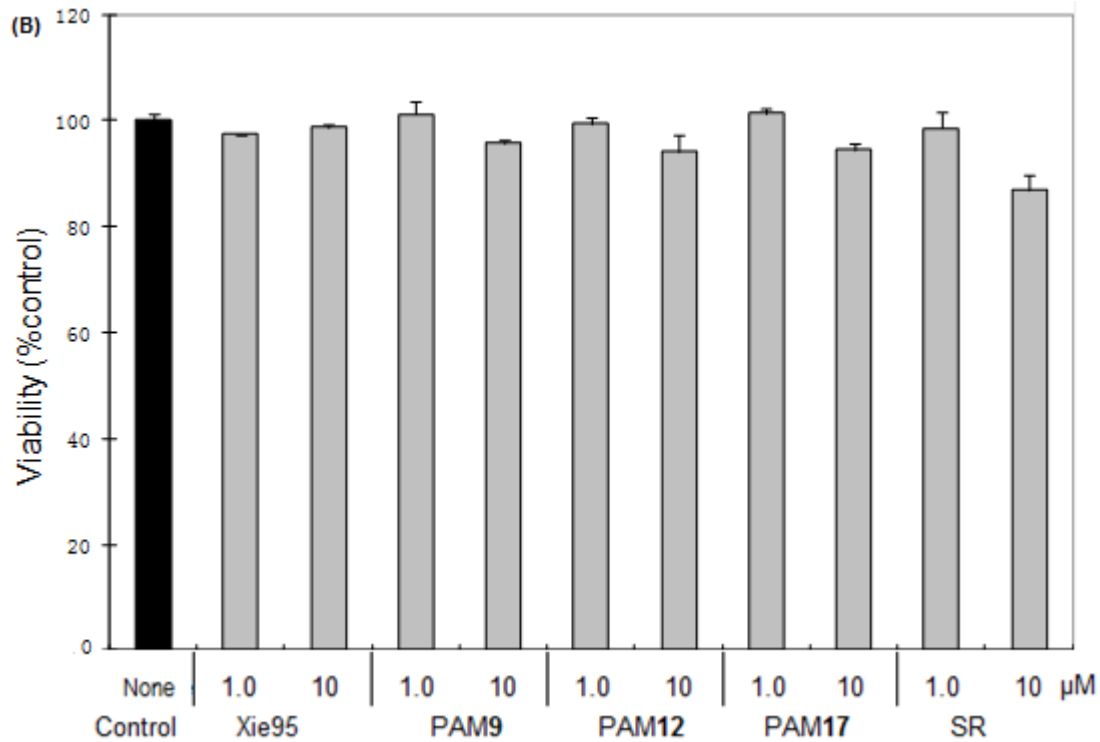
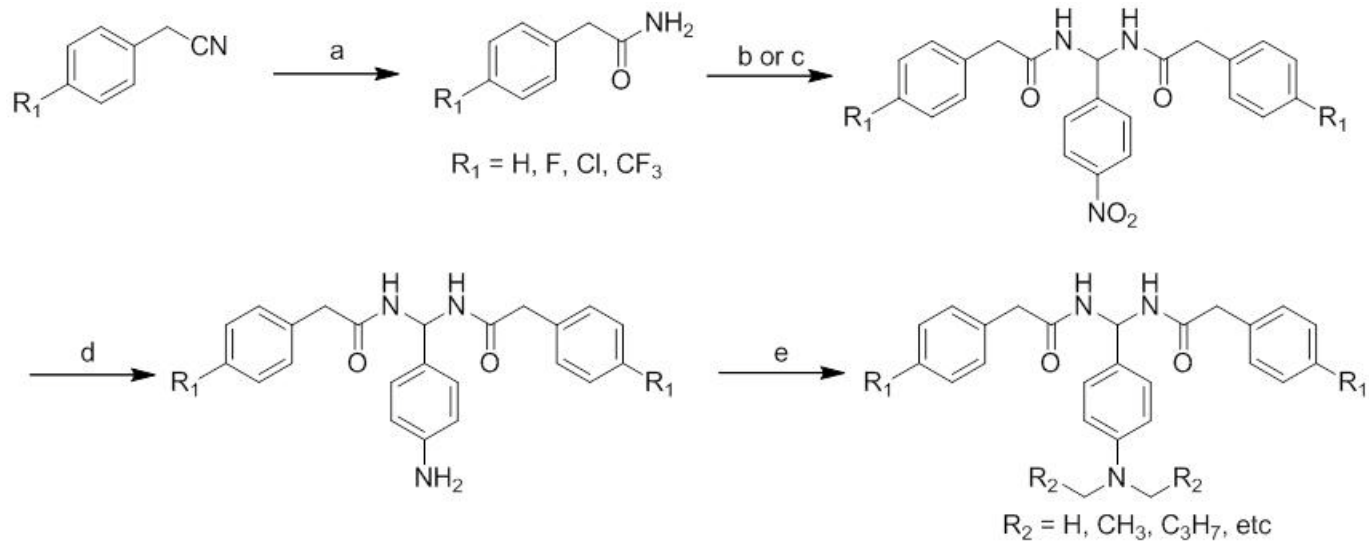


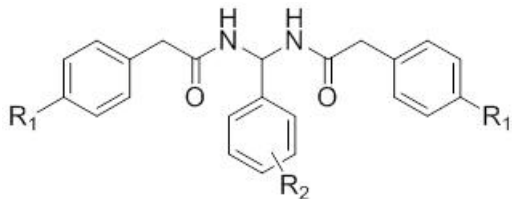
Figure 5-9 Cytotoxic effects of PAM compounds on normal human mononuclear cells.

Samples of primary PBMCs (10^5 cells per well in 96-well plate) from healthy donors were treated in culture for 72 h with the indicated compounds. The viability of cells was determined using trypan blue exclusion assay. The results were presented as mean \pm SD of three assays.

Scheme 1. General Synthesis of PAM Analogues



Reagents and conditions: (a) concentrated H_2SO_4 , $0\text{ }^\circ\text{C}$, 12 h; (b) method 1: aldehyde, anhydrous dichloroethane, TMSCl , $70\text{ }^\circ\text{C}$, 3-12 h; (c) method 2: aldehyde, anhydrous DCM, $\text{F}_3\text{CSO}_3\text{SiMe}_3$, r.t., 12 h; (d) ethanol, palladium (10%), hydrazine, $70\text{ }^\circ\text{C}$, 3 h; (e) DMF, K_2CO_3 , r.t., 12 h.

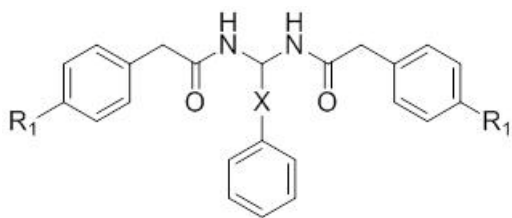
Table 5-1. Radioligand Competition Binding Affinity (K_i) Data of PAM Derivatives

Compd	R ₁	R ₂	MW	cLog P	K _i (CB ₂), nM ^{b, c}	K _i (CB ₁), nM ^{a, d}	SI ^e
1	H	<i>p</i> -(CH ₃) ₂ N-	401.50	4.04	777	> 20,000	> 26
2	H	H-	358.43	3.93	9,930	NT	
3	H	<i>o</i> -F-	376.42	4.08	35,330	NT	
4	H	<i>m</i> -F-	376.42	4.08	12,670	NT	
5	H	<i>p</i> -F-	376.42	4.08	10,900	NT	
6	H	<i>p</i> -Cl-	392.88	4.54	3,081	NT	
7	H	<i>p</i> -Br-	437.33	4.70	2,226	NT	
8	H	<i>p</i> -CH ₃ -	372.46	4.45	494	109	
9	H	<i>p</i> - <i>i</i> -C ₃ H ₇ -	400.51	5.18	85	> 20,000	> 235
10	H	<i>p</i> -CH ₃ O-	388.46	3.78	783	> 20,000	> 26
11	H	<i>p</i> -C ₂ H ₅ O-	402.49	4.13	1,500	NT	
12	H	<i>p</i> - <i>i</i> -C ₃ H ₇ O-	416.51	4.55	313	> 20,000	> 64
13	H	<i>o</i> -CF ₃ -	426.43	4.81	11,780	NT	
14	H	<i>p</i> -CF ₃ -	426.43	4.81	596	> 20,000	> 34
15	H	<i>p</i> -NO ₂ -	403.43	3.87	NB	NT	
16	H	<i>p</i> -H ₂ N-	373.45	2.51	12,550	NT	
17	H	<i>p</i> -(C ₂ H ₅) ₂ N-	429.55	4.76	64	> 20,000	> 313
18	H	<i>p</i> -(C ₃ H ₇) ₂ N-	457.61	5.80	22	> 20,000	> 909
19	H	<i>p</i> -(C ₄ H ₉) ₂ N-	485.66	6.69	221	> 20,000	> 90
20	H	<i>p</i> -(Benzyl) ₂ N-	553.69	7.33	203	> 20,000	> 99
21	H	<i>p</i> -pyrrolidinyl-	427.53	4.45	71	> 20,000	> 281
22	H	<i>p</i> -piperidyl-	441.56	4.89	595	> 20,000	> 34
23	Cl	H-	427.32	5.14	NB	NT	
24	Cl	<i>o</i> -F-	445.31	5.29	10,850	NT	
25	Cl	<i>p</i> -F-	445.31	5.29	NB	NT	
26	Cl	<i>p</i> -Cl-	461.77	5.75	154	> 20,000	> 130
27	Cl	<i>p</i> -CH ₃ -	441.35	5.66	462	> 20,000	> 43
28	Cl	<i>p</i> -CH ₃ O-	457.35	4.98	310	> 20,000	> 65
29	Cl	<i>o</i> -CF ₃ -	495.32	6.02	158	> 20,000	> 127
30	Cl	<i>p</i> -CF ₃ -	495.32	6.02	101	> 20,000	> 198
31	Cl	<i>p</i> -NO ₂ -	472.32	5.08	NB	NT	
32	CF ₃	H-	494.43	5.69	NB	NT	
33	CF ₃	<i>o</i> -F-	512.42	5.83	NB	NT	
34	CF ₃	<i>p</i> -F-	512.42	5.83	NB	NT	
35	CF ₃	<i>p</i> -Cl-	528.87	6.29	NB	NT	

36	CF ₃	<i>p</i> -CH ₃ -	508.46	6.20	NB	NT
37	CF ₃	<i>p</i> -CH ₃ O-	524.45	5.53	NB	NT
38	CF ₃	<i>p</i> -CF ₃ -	562.43	6.57	NB	NT
SR1 ^{f, g}					2.1	NT
SR2 ^{f, h}					NT	10.6

^{a, b} Binding affinities of compounds for CB₁ and CB₂ receptor were evaluated using [³H]CP-55,940 radioligand competition binding assay. ^c NB no binding, $K_i > 20,000$ nM. ^d NT = not tested. ^e SI: selectivity index for CB₂, calculated as $K_i(\text{CB}_1)/K_i(\text{CB}_2)$ ratio. ^f The binding affinities of reference compounds were evaluated in parallel with compounds **1-52** under the same conditions. ^g CB₂ reference compound SR144528 (SR1). ^h CB₁ reference compound SR 141716 (SR2).

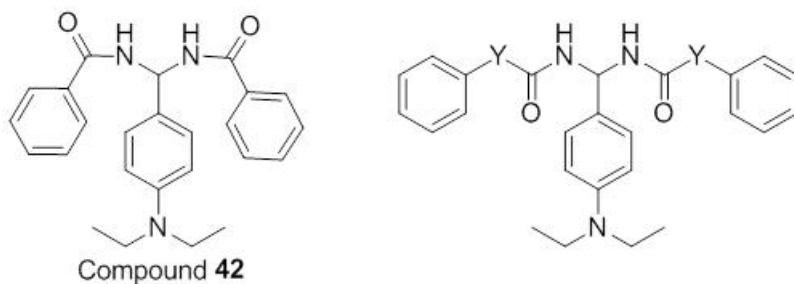
Table 5-2. Radioligand Competition Binding Affinity (K_i) Data of PAM Derivatives



Compd	R ₁	X	MW	cLog P	K_i (CB ₂), nM ^{b, c}	K_i (CB ₁), nM ^{a, d}	SI ^e
39	H	CH ₂	372.46	3.99	NB	NB	
40	H	CH ₂ CH ₂	386.49	4.44	9,319	NB	
41	H	CH=CH	384.47	4.54	5,683	NB	
SR1 ^{f, g}					2.1	NT	
SR2 ^{f, h}					NT	10.6	

^{a- h} Same as described in Table 1.

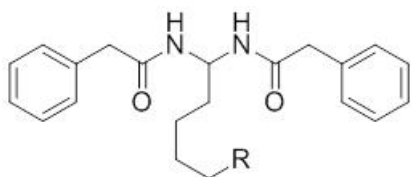
Table 5-3. Radioligand Competition Binding Affinity (K_i) Data of PAM Derivatives



Compd	Y	MW	cLog P	K_i (CB ₂), nM ^{b, c}	K_i (CB ₁), nM ^{a, d}	SI ^e
42		401.50	4.80	688	> 20,000	> 29
43	CH ₂ CH ₂	457.61	5.64	213	> 20,000	> 93
44	CH=CH	453.58	5.80	167	> 20,000	> 119
SR1 ^{f, g}				2.1	NT	
SR2 ^{f, h}				NT	10.6	

^{a-h} Same as described in Table 1.

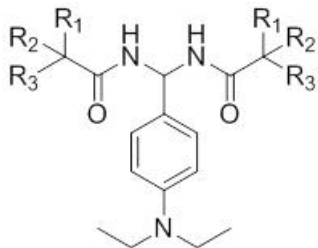
Table 5-4. Radioligand Competition Binding Affinity (K_i) Data of PAM Derivatives



Compd	R	MW	cLog P	K_i (CB ₂), nM ^{b, c}	K_i (CB ₁), nM ^{a, d}
45	H	338.44	3.75	35,970	NT
46	CH ₃	352.47	4.19	18,200	NB
SR1 ^{f, g}				2.1	NT
SR2 ^{f, h}				NT	10.6

^{a-h} Same as described in Table 1.

Table 5-5. Radioligand Competition Binding Affinity (K_i) Data of PAM Derivatives



Compd	R ₁	R ₂	R ₃	MW	cLog P	K _i (CB ₂), nM ^{b, c}	K _i (CB ₁), nM ^{a, d}	SI ^e
47	H	CH ₃	CH ₃	333.46	3.57	2636	NB	
48	CH ₃	CH ₃	CH ₃	361.52	4.69	3553	NB	
49	H	H	C ₃ H ₇	361.52	4.27	182	> 20,000	> 109
50	H	H	C ₄ H ₉	389.57	5.16	25	> 20,000	> 800
51	H	H	C ₆ H ₁₃	445.68	7.9	146	> 20,000	> 136
52	H	H	C ₈ H ₁₇	501.79	10.0	160	> 20,000	> 125
SR1 ^{f, g}						2.1	NT	
SR2 ^{f, h}						NT	10.6	

^{a-h} Same as described in Table 1.

Table 5-6. Experimental (expt) and predicted (pred) pK_i values of PAM derivatives in the training set and test set

Compd	pK _i (expt)	pK _i (pred)	Residual
1	6.11	6.23	-0.12
2	5.00	4.99	0.01
3	4.45	4.287	0.16
4	4.90	4.879	0.02
5	4.96	4.791	0.17
6 ^a	5.51	5.54	-0.03
7 ^a	5.65	5.73	-0.08
8	6.31	6.141	0.17
9	7.07	7.147	-0.08
10 ^a	6.11	5.81	0.30
11	5.82	5.978	-0.16
12	6.50	6.431	0.07
13	4.93	5.083	-0.15
14	6.22	6.336	-0.12
16 ^a	4.90	5.37	-0.47
17	7.19	6.987	0.20
18	6.66	6.711	-0.05
19 ^a	6.23	6.77	-0.54
20 ^a	6.69	6.4	0.29
22	4.96	5.024	-0.06
24	6.81	6.67	0.14

25	6.34	6.45	-0.11
26 ^a	6.51	6.6	-0.09
27	6.80	6.746	0.05
28	7.00	6.971	0.03
38	5.03	5.078	-0.05
39	5.25	5.24	0.01
40	4.44	4.571	-0.13
41	4.74	4.679	0.06

^a Molecules from the test set.

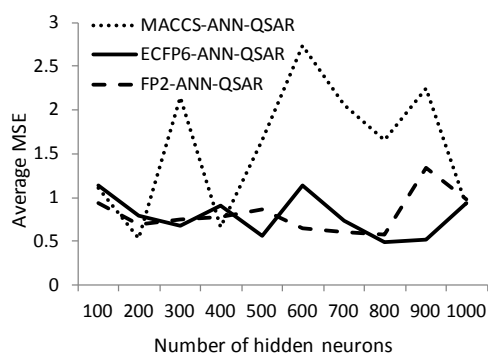
6.0 SUMMARY AND FUTURE WORK

In this dissertation work, several novel QSAR methods were developed and presented. They could be used in virtual screening and hit ranking of chemical ligands. In addition, novel CB2 selective inverse agonist ligands were discovered. To recap, three different QSAR methods were discussed. Each method has slightly different uses and purposes. In particular, the FS-QSAR and PharmShape-QSAR methods are useful to study local changes given a series of compounds with similar scaffolds. Different R-groups can affect the biological activities of analogs and these two methods incorporate such local changes in predictions. On the other hand, the FANN-QSAR method can be used as a high throughput virtual screening tool to screen large and structurally diverse compounds in order to find lead compounds. This method uses 2D molecular fingerprints of whole molecule and no longer requires compounds to have similar scaffolds. Using the approach, new chemical scaffolds having good CB2 binding affinities were discovered. These developed methods could be useful in the early phases of drug discovery process to screen out biologically active chemical ligands to save time and costly experiments. The last chapter discussed the discovery of novel CB2 ligands within nM binding affinities through medicinal chemistry and molecular modeling studies. Using experimental binding and functional assays, these ligands were confirmed to be CB2 selective inverse agonists. These novel scaffolds could be used as probes to investigate CB2 signaling.

In the future, the developed QSAR models can be updated using new SAR studies from literature. In addition to the fingerprints used in the research, addition fingerprints can be further explored to examine if there is any different effect on final QSAR predictions. It is expected that there will not be significant changes in the quality of QSAR model by using different 2D fingerprints. However, using properly defined 3D descriptors may improve the model's quality. But one should keep in mind that 3D descriptors may require additional information such as ligands' conformation and superimposition. Additional QSAR and medicinal chemistry studies will be carried out to make lead compounds better affinity and selectivity. For example, developed 3D-QSAR model can be further improved and updated by incorporating newly synthesized compounds. New SAR studies from other scaffolds can be used as references to infer important molecular moieties which may be useful substituents in our lead compounds. As a proof of principle, only NCI compound database was used for virtual screening in this study. In the future, developed QSAR model can be used to screen several other databases such as ZINC database to find novel lead compounds with different scaffolds as well. As a complimentary approach to ligand-based design, a receptor-based modeling was also used in this study to guide the early design of our lead compounds. In the future, the predicted binding pocket can be further validated via site-directed mutagenesis studies to elucidate important binding residues for successful ligand binding. Results from such studies will provide insights for essential receptor-ligand interactions which can be used in modifying lead compounds for better affinity and selectivity. These lead compounds can be further tested for other in vitro and in vivo studies for potential therapeutic applications.

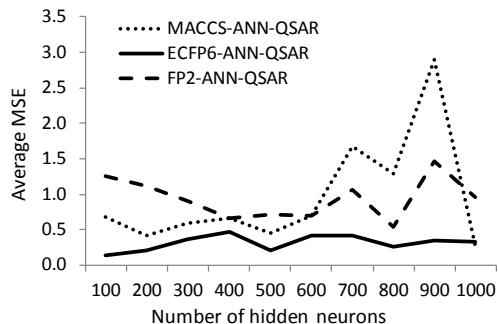
APPENDIX A

SUPPORTING INFORMATION FOR FINGERPRINT-BASED ANN QSAR (FANN-QSAR) METHOD



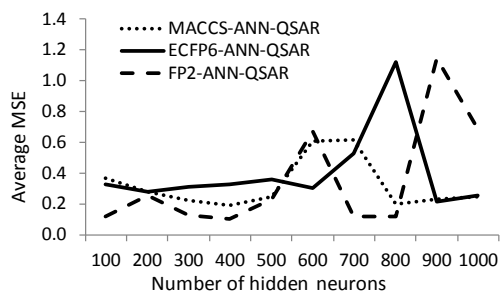
	Optimal no. of hidden neurons	Training Set MSE	Test Set MSE	Average MSE
ECFP6-ANN-QSAR	800	0.208	0.770	0.489
FP2-ANN-QSAR	800	0.368	0.789	0.579
MACCS-ANN-QSAR	200	0.337	0.742	0.539

Figure A1. Cross-validation results of three FANN-QSAR models on ACE data set



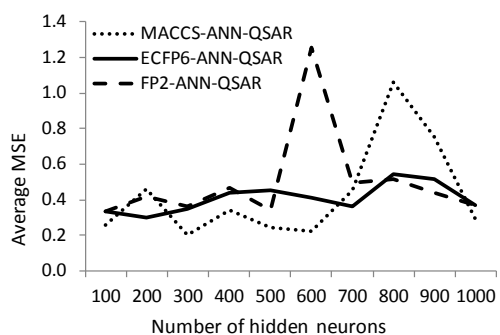
	Optimal no. of hidden neurons	Training Set MSE	Test Set MSE	Average MSE
ECFP6-ANN-QSAR	100	0.100	0.184	0.142
FP2-ANN-QSAR	800	0.381	0.685	0.533
MACCS-ANN-QSAR	1000	0.176	0.391	0.283

Figure A2. Cross-validation results of three FANN-QSAR models on AchE data set



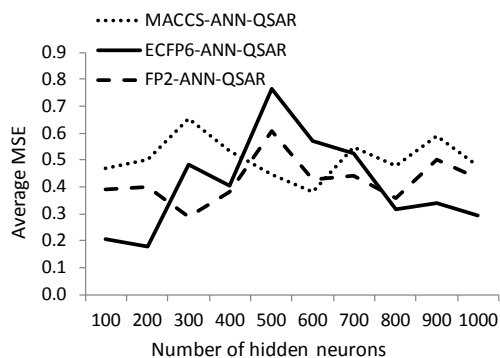
	Optimal no. of hidden neurons	Training Set MSE	Test Set MSE	Average MSE
ECFP6-ANN-QSAR	900	0.100	0.332	0.216
FP2-ANN-QSAR	400	0.100	0.104	0.102
MACCS-ANN-QSAR	400	0.100	0.285	0.193

Figure A3. Cross-validation results of three FANN-QSAR models on BZR data set



	Optimal no. of hidden neurons	Training Set MSE	Test Set MSE	Average MSE
ECFP6-ANN-QSAR	200	0.186	0.418	0.302
FP2-ANN-QSAR	100	0.187	0.482	0.335
MACCS-ANN-QSAR	300	0.112	0.293	0.202

Figure A4. Cross-validation results of three FANN-QSAR models on COX2 data set



	Optimal no. of hidden neurons	Training Set MSE	Test Set MSE	Average MSE
ECFP6-ANN-QSAR	200	0.100	0.261	0.180
FP2-ANN-QSAR	300	0.241	0.340	0.291
MACCS-ANN-QSAR	600	0.207	0.557	0.382

Figure A5. Cross-validation results of three FANN-QSAR models on DHFR data set

Table A1. Twenty rounds of model training and cross-validation results using CB₂ ligand data set before testing compounds from NCI database

Round	Training MSE	Validation MSE	Avg. MSE	Optimal No. of Hidden Neurons	r ² training	r ² validation
1	0.169	0.733	0.451	1000	0.87	0.47
2	0.126	0.832	0.479	500	0.91	0.49
3	0.362	0.645	0.504	600	0.74	0.53
4	0.262	0.555	0.409	400	0.81	0.59
5	0.332	0.776	0.554	800	0.74	0.49
6	0.214	0.830	0.522	200	0.84	0.41
7	0.207	0.622	0.415	500	0.85	0.56
8	0.181	0.764	0.472	400	0.87	0.51
9	0.255	0.563	0.409	800	0.82	0.60
10	0.227	0.713	0.470	900	0.84	0.52
11	0.217	0.636	0.427	600	0.83	0.51
12	0.239	0.832	0.535	600	0.82	0.50
13	0.185	0.729	0.457	800	0.87	0.43
14	0.137	0.656	0.396	400	0.90	0.54
15	0.389	0.565	0.477	900	0.70	0.55
16	0.159	0.727	0.443	500	0.87	0.50
17	0.303	0.709	0.506	500	0.78	0.50
18	0.182	0.515	0.348	600	0.87	0.64
19	0.219	0.681	0.450	900	0.83	0.53
20	0.208	0.567	0.388	1000	0.85	0.53

BIBLIOGRAPHY

1. Esposito EX, Hopfinger AJ, Madura JD: **Methods for Applying the Quantitative Structure-Activity Relationship Paradigm**. In., vol. 275; 2004: 131-213.
2. Bradbury SP: **Quantitative structure-activity relationships and ecological risk assessment: an overview of predictive aquatic toxicology research**. *Toxicology Letters* 1995, **79**(1-3):229.
3. Hansch C, Telzer BR, Zhang L: **Comparative QSAR in Toxicology: Examples from Teratology and Cancer Chemotherapy of Aniline Mustards**. *Critical Reviews in Toxicology* 1995, **25**(1):67-89.
4. Perkins R, Fang H, Tong W, Welsh W: **Quantitative structure-activity relationship methods: perspectives on drug discovery and toxicology**. *Environ Toxicol Chem* 2003, **22**(8):1666-1679.
5. Chen J-Z, Han X-W, Liu Q, Makriyannis A, Wang J, Xie X-Q: **3D-QSAR Studies of Arylpyrazole Antagonists of Cannabinoid Receptor Subtypes CB1 and CB2. A Combined NMR and CoMFA Approach**. *Journal of Medicinal Chemistry* 2006, **49**(2):625-636.
6. Salum L, Andricopulo A: **Fragment-based QSAR: perspectives in drug design**. *Molecular Diversity* 2009, **13**(3):277.
7. Free SJ, Wilson J: **A mathematical contribution to structure-activity studies**. *J Med Chem* 1964, **7**:395-399.
8. Hansch C, Fujita T: **ρ - σ - π Analysis. A method for the correlation of biological activity and chemical structure**. *J Am Chem Soc* 1964, **86**:1616-1626.
9. Tripos-vn.7.6: **Tripos Sybyl (version 7.2) molecular modeling software packages**. www.tripos.com. In. St.Louis, MI63144: TRIPOS, Associates, Inc.
10. Lowis D: **HQSAR: A New, Highly Predictive QSAR Technique**. *Tripos Technical Notes* 1997, **1**(5):17.
11. Castilho MS, Postigo MP, de Paula CBV, Montanari CA, Oliva G, Andricopulo AD: **Two- and three-dimensional quantitative structure-activity relationships for a series of purine nucleoside phosphorylase inhibitors**. *Bioorganic & Medicinal Chemistry* 2006, **14**(2):516.
12. Salum LdB, Polikarpov I, Andricopulo AD: **Structural and chemical basis for enhanced affinity and potency for a large series of estrogen receptor ligands: 2D and 3D QSAR studies**. *Journal of Molecular Graphics and Modelling* 2007, **26**(2):434.
13. Honorio KM, Garratt RC, Andricopulo AD: **Hologram quantitative structure-activity relationships for a series of farnesoid X receptor activators**. *Bioorganic & Medicinal Chemistry Letters* 2005, **15**(12):3119.
14. Castilho MS, Guido RVC, Andricopulo AD: **2D Quantitative structure-activity relationship studies on a series of cholesteryl ester transfer protein inhibitors**. *Bioorganic & Medicinal Chemistry* 2007, **15**(18):6242.

15. Lo Piparo E, Koehler K, Chana A, Benfenati E: **Virtual Screening for Aryl Hydrocarbon Receptor Binding Prediction.** *Journal of Medicinal Chemistry* 2006, **49**(19):5702.
16. Tropsha A, Golbraikh A: **Predictive QSAR Modeling Workflow, Model Applicability Domains, and Virtual Screening** *Current Pharmaceutical Design* 2007, **13**(34):3494–3504.
17. Prakash O, Ghosh I: **Developing an Antituberculosis Compounds Database and Data Mining in the Search of a Motif Responsible for the Activity of a Diverse Class of Antituberculosis Agents.** *Journal of Chemical Information and Modeling* 2005, **46**(1):17.
18. Du Q-S, Huang R-B, Yu-Tuo W, Pang Z-W, Du L-Q, Chou K-C: **Fragment-based quantitative structure-activity relationship (FB-QSAR) for fragment-based drug design.** *Journal of Computational Chemistry* 2009, **30**(2):295-304.
19. Myint K-Z, Ma C, Wang L, Xie XQ: **Fragment-Similarity-Based QSAR (FS-QSAR) Algorithm for Ligand Biological Activity Predictions.** *SAR and QSAR in Environmental Research* 2010, **22**(3-4):1-26.
20. Burden F: **Molecular identification number for substructure searches.** *J Chem Inf Comput Sci* 1989, **29**(3):225-227.
21. Xie XQ, Chen J: **Data-Mining a Small Molecule Drug Screening Representative Subset from NIH PubChem Database.** *Journal of Chemical Information and Modeling* 2008, **48**(3):465-475.
22. Casalegno M, Sello G, Benfenati E: **Top-Priority Fragment QSAR Approach in Predicting Pesticide Aquatic Toxicity.** *Chem Res Toxicol* 2006, **19**(11):1533-1539.
23. Zhokhova N, Baskin I, Palyulin V, Zefirov A, Zefirov N: **Fragmental descriptors with labeled atoms and their application in QSAR/QSPR studies.** *Doklady Chemistry* 2007, **417**(2):282.
24. Baskin II, Halberstam NM, Artemenko NV: **Designing Drugs and Crop Protectants: Processes, Problems, and Solutions.** In: *EuroQSAR-2002*. Melbourne, Blackwell.; 2003: 260-263.
25. Andrade C, Salum Ld, Castilho M, Pasqualoto K, Ferreira E, Andricopulo A: **Fragment-based and classical quantitative structure-activity relationships for a series of hydrazides as antituberculosis agents.** *Molecular Diversity* 2008, **12**(1):47.
26. TALETE_SRL: **DRAGON5.4.** In. Milan, Italy.; 2008.
27. Oliveira DBd, Gaudio AC: **BuildQSAR: A New Computer Program for QSAR Analysis.** *Quantitative Structure-Activity Relationships* 2000, **19**(6):599-601.
28. Infometrix: **Pirouette Multivariate Data Analysis for IBM PC Systems.** In. Seattle, WA.; 2001.
29. Tsygankova I, Zhenodarova S: **Quantitative structure-activity relationship for barbituric acid derivatives: Potential of the fragment approach.** *Russian Journal of General Chemistry* 2007, **77**(5):940.
30. Cramer R, Patterson D, Bunce J: **Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins.** *J Am Chem Soc* 1988, **110**:5959-5967.
31. Klebe G, Abraham U, Mietzner T: **Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity.** *Journal of Medicinal Chemistry* 1994, **37**:4130-4146.

32. Dudek AZ, Arodz O, Galvez J: **Computational Methods in Developing Quantitative Structure-Activity Relationships (QSAR): A Review.** *Combinatorial Chemistry & High Throughput Screening* 2006, **9**(3):213.
33. Cramer RD, Cruz P, Stahl G, Curtiss WC, Campbell B, Masek BB, Soltanshahi F: **Virtual Screening for R-Groups, including Predicted pIC50 Contributions, within Large Structural Databases, Using Topomer CoMFA.** *Journal of Chemical Information and Modeling* 2008, **48**(11):2180.
34. Avram S, Milac AL, Flonta ML: **Computer-Aided Drug Design for Typical and Atypical Antipsychotic Drugs: A Review of Application of QSAR and Combinatorial Chemistry Methods - Tools for New Antipsychotics Design.** *Current Computer - Aided Drug Design* 2005, **1**:347.
35. Patcharawee N, Nahoum GA, Blair FJ, Simon PM, Jiraporn U: **3D-QSAR Studies on Chromone Derivatives as HIV-1 Protease Inhibitors: Application of Molecular Field Analysis.** *Archiv der Pharmazie* 2008, **341**(6):357-364.
36. Labrie P, Maddaford SP, Fortin S, Rakhit S, Kotra LP, Gaudreault RC: **A Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA) of Anthranilamide Derivatives That Are Multidrug Resistance Modulators.** *Journal of Medicinal Chemistry* 2006, **49**(26):7646.
37. Jeong JA, Cho H, Jung SY, Kang HB, Park JY, Kim J, Choo DJ, Lee JY: **3D QSAR studies on 3,4-dihydroquinazolines as T-type calcium channel blocker by comparative molecular similarity indices analysis (CoMSIA).** *Bioorganic & Medicinal Chemistry Letters* 2010, **20**(1):38.
38. Dayan FE, Singh N, McCurdy CR, Godfrey CA, Larsen L, Weavers RT, Van Klink JW, Perry NB: **β -Triketone Inhibitors of Plant p-Hydroxyphenylpyruvate Dioxygenase: Modeling and Comparative Molecular Field Analysis of Their Interactions.** *Journal of Agricultural and Food Chemistry* 2009, **57**(12):5194.
39. Cramer RD: **Topomer CoMFA: A Design Methodology for Rapid Lead Optimization.** *Journal of Medicinal Chemistry* 2003, **46**(3):374.
40. Robinson DD, Winn PJ, Lyne PD, Richards WG: **Self-Organizing Molecular Field Analysis: A Tool for Structure-Activity Studies.** *Journal of Medicinal Chemistry* 1999, **42**(4):573-583.
41. Bravi G, Gancia E, Mascagni P, Pegna M, Todeschini R, Zaliani A: **MS-WHIM, new 3D theoretical descriptors derived from molecular surface properties: A comparative 3D QSAR study in a series of steroids.** *Journal of computer-aided molecular design* 1997, **11**(1):79.
42. Wagener M, Sadowski J, Gasteiger J: **Autocorrelation of Molecular Surface Properties for Modeling Corticosteroid Binding Globulin and Cytosolic Ah Receptor Activity by Neural Networks.** *Journal of the American Chemical Society* 1995, **117**(29):7769-7775.
43. Silverman BD, Platt DE: **Comparative Molecular Moment Analysis (CoMMA): 3D-QSAR without Molecular Superposition.** *Journal of Medicinal Chemistry* 1996, **39**(11):2129-2140.
44. Todeschini R, Gramatica P: **New 3D molecular descriptors: the WHIM theory and QSAR applications.** *Perspectives in Drug Discovery and Design* 1998, **9-11**(0):355.
45. Todeschini R, Lasagni M, Marengo E: **New molecular descriptors for 2D and 3D structures.** *J Chemometrics* 1994, **8**(4):263-272.

46. Pastor M, Cruciani G, McLay I, Pickett S, Clementi S: **GRid-INdependent Descriptors (GRIND): A Novel Class of Alignment-Independent Three-Dimensional Molecular Descriptors.** *J Med Chem* 2000, **43**(17):3233-3243.
47. Connolly M: **Analytical molecular surface calculation.** *Journal of Applied Crystallography* 1983, **16**(5):548-558.
48. Pastor M, Cruciani G, Watson KA: **A Strategy for the Incorporation of Water Molecules Present in a Ligand Binding Site into a Three-Dimensional Quantitative Structure-Activity Relationship Analysis.** *Journal of Medicinal Chemistry* 1997, **40**(25):4089.
49. ALMOND: **Molecular Discovery Ltd.** In. Perugia, Italy. http://www.moldiscovery.com/soft_almond.php.
50. Hopfinger AJ, Wang S, Tokarski JS, Jin B, Albuquerque M, Madhav PJ, Duraiswami C: **Construction of 3D-QSAR Models Using the 4D-QSAR Analysis Formalism.** *Journal of the American Chemical Society* 1997, **119**(43):10509-10524.
51. Scheiber J, Enzensperger C, Lehmann J, Stiefl N, Baumann K: **Alignment-free 4D-QSAR: Applying the XMAP technique in prospective analyses.** In: *QSAR & Molecular Modeling in Rational Design of Bioactive Molecules*. Edited by AKI-SENER E, YALCIN I. Proc. of the 15th EURO-QSAR, Ankara, Turkey.; 2006.
52. Fischer PM: **Computational chemistry approaches to drug discovery in signal transduction.** *Biotechnology Journal* 2008, **3**(4):452-470.
53. Vedani A, Dobler M: **5D-QSAR: The Key for Simulating Induced Fit?** *Journal of Medicinal Chemistry* 2002, **45**(11):2139-2149.
54. Vedani A, Dobler M, Lill MA: **Combining Protein Modeling and 6D-QSAR. Simulating the Binding of Structurally Diverse Ligands to the Estrogen Receptor.** *Journal of Medicinal Chemistry* 2005, **48**(11):3700-3703.
55. Vedani A, Dobler M, Zbinden P: **Quasi-Atomistic Receptor Surface Models: A Bridge between 3-D QSAR and Receptor Modeling.** *Journal of the American Chemical Society* 1998, **120**(18):4471.
56. Biograf: **VirtualToxLab.** In. Basel, Switzerland: <http://www.biograf.ch/downloads/VirtualToxLab.pdf>; 2009.
57. Hillebrecht A, Klebe G: **Use of 3D QSAR Models for Database Screening: A Feasibility Study.** *J Chem Inf Model* 2008, **48**(2):384-396.
58. Matter H, Potter T: **Comparing 3D Pharmacophore Triplets and 2D Fingerprints for Selecting Diverse Compound Subsets.** *Journal of Chemical Information and Computer Sciences* 1999, **39**(6):1211.
59. Khedkar V, Ambre P, Verma J, Shaikh M, Pissurlenkar R, Coutinho E: **Molecular docking and 3D-QSAR studies of HIV-1 protease inhibitors.** *Journal of Molecular Modeling* 2010, **16**(7):1251.
60. Klebe G: **Molecular Similarity Indices in a Comparative Analysis: CoMSIA.** *Perspectives in Drug Discovery and Design* 1998, **12/13/14**:87-104.
61. Klebe G: **Comparative Molecular Similarity Indices Analysis: CoMSIA.** In: *3D QSAR in Drug Design*. 1998: 87-104.
62. Li Q, J rgensen FS, Oprea T, Brunak Sr, Taboureau O: **hERG Classification Model Based on a Combination of Support Vector Machine Method and GRIND Descriptors.** *Molecular Pharmaceutics* 2008, **5**(1):117.

63. Romeiro N, Albuquerque M, Alencastro R, Ravi M, Hopfinger A: **Construction of 4D-QSAR Models for Use in the Design of Novel p38-MAPK Inhibitors.** *Journal of Computer-Aided Molecular Design* 2005, **19**(6):385.
64. Joseph-McCarthy D: **Computational approaches to structure-based ligand design.** *Pharmacology & Therapeutics* 1999, **84**(2):179-191.
65. Goodford PJ: **A computational procedure for determining energetically favorable binding sites on biologically important macromolecules.** *J Med Chem* 1985, **28**(7):849-857.
66. Evensen E, Joseph-McCarthy D, Karplus M: **MCSSv2.** *Cambridge: Harvard University* 1997.
67. Miranker A, Karplus M: **Functionality maps of binding sites: A multiple copy simultaneous search method.** *Proteins: Structure, Function, and Genetics* 1991, **11**(1):29-34.
68. Dean PM: **Computer-aided Design of Small Molecules for Chemical Genomics.** New Jersey: Humana Press Inc.; 2005.
69. DeWitte RS, Shakhnovich EI: **SMoG: de Novo Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. 1. Methodology and Supporting Evidence.** *Journal of the American Chemical Society* 1996, **118**(47):11733-11744.
70. Bohacek RS, McMartin C: **Multiple Highly Diverse Structures Complementary to Enzyme Binding Sites: Results of Extensive Application of a de Novo Design Method Incorporating Combinatorial Growth.** *Journal of the American Chemical Society* 1994, **116**(13):5560-5571.
71. Rotstein SH, Murcko MA: **GenStar: A method for de novo drug design.** *Journal of computer-aided molecular design* 1993, **7**(1):23-43.
72. Moon JB, Howe WJ: **Computer design of bioactive molecules: A method for receptor-based de novo ligand design.** *Proteins: Structure, Function, and Genetics* 1991, **11**(4):314-328.
73. Kitchen DB, Decornez H, Furr JR, Bajorath J: **Docking and scoring in virtual screening for drug discovery: methods and applications.** *Nature Reviews Drug Discovery* 2004, **3**(11):935.
74. Moustakas DT, Lang PT, Pegg S, Pettersen E, Kuntz ID, Brooijmans N, Rizzo RC: **Development and validation of a modular, extensible docking program: DOCK 5.** *Journal of computer-aided molecular design* 2006, **20**(10/11):601-619.
75. Jones G, Willett P, Glen RC: **Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation.** *Journal of Molecular Biology* 1995, **245**(1):43-53.
76. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ: **Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function.** *Journal of Computational Chemistry* 1998, **19**(14):1639-1662.
77. **SYBYL-X 1.2.** In. 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA; .
78. Rester U: **Dock around the Clock - Current Status of Small Molecule Docking and Scoring.** *QSAR & Combinatorial Science* 2006, **25**(7):605-615.
79. Schneider G, Fechner U: **Computer-based de novo design of drug-like molecules.** *Nat Rev Drug Discov* 2005, **4**(8):649.

80. Hessler G, Zimmermann M, Matter H, Evers A, Naumann T, Lengauer T, Rarey M: **Multiple-Ligand-Based Virtual Screening: Methods and Applications of the MTree Approach.** *Journal of Medicinal Chemistry* 2005, **48**(21):6575-6584.
81. Reddy AS, Pati SP, Kumar PP, Pradeep HN, Sastry GN: **Virtual Screening in Drug Discovery - A Computational Perspective.** *Current Protein and Peptide Science* 2007, **8**:329-351.
82. Tripos: **SYBYL8.0.** In: *Discovery software for computational chemistry and molecular modeling.* St. Louis, Missouri, USA 2007.
83. McMartin C, Bohacek RS: **QXP: Powerful, rapid computer algorithms for structure-based drug design.** *Journal of computer-aided molecular design* 1997, **11**(4):333-344.
84. Abagyan R, Totrov M, Kuznetsov D: **ICM - A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation.** *Journal of Computational Chemistry* 1994, **15**(5):488-506.
85. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK *et al*: **Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy.** *Journal of Medicinal Chemistry* 2004, **47**(7):1739-1749.
86. Rarey M, Kramer B, Lengauer T, Klebe G: **A Fast Flexible Docking Method using an Incremental Construction Algorithm.** *Journal of Molecular Biology* 1996, **261**(3):470-489.
87. Schellhammer I, Rarey M: **FlexX-Scan: Fast, structure-based virtual screening.** *Proteins: Structure, Function, and Bioinformatics* 2004, **57**(3):504-517.
88. Welch W, Ruppert J, Jain AN: **Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites.** *Chemistry & Biology* 1996, **3**(6):449.
89. Venkatachalam CM, Jiang X, Oldfield T, Waldman M: **LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites.** *Journal of Molecular Graphics and Modelling* 2003, **21**(4):289-307.
90. McGann MR, Almond HR, Nicholls A, Grant JA, Brown FK: **Gaussian docking functions.** *Biopolymers* 2003, **68**(1):76-90.
91. Lemmen C, Lengauer T, Klebe G: **FlexS: A Method for Fast Flexible Ligand Superposition.** *Journal of Medicinal Chemistry* 1998, **41**(23):4502-4520.
92. Renner S, Schneider G: **Fuzzy Pharmacophore Models from Molecular Alignments for Correlation-Vector-Based Virtual Screening.** *Journal of Medicinal Chemistry* 2004, **47**(19):4653-4664.
93. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ: **Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.** *Advanced Drug Delivery Reviews* 2001, **46**(1-3):3-26.
94. Rush TS, Grant JA, Mosyak L, Nicholls A: **A Shape-Based 3-D Scaffold Hopping Method and Its Application to a Bacterial Protein.** *Journal of Medicinal Chemistry* 2005, **48**(5):1489-1495.
95. Tervo AJ, Ronkko T, Nyronen TH, Poso A: **BRUTUS: Optimization of a Grid-Based Similarity Function for Rigid-Body Molecular Superposition. 1. Alignment and Virtual Screening Applications.** *Journal of Medicinal Chemistry* 2005, **48**(12):4076-4086.
96. Tripos: **Tripos Unity Module.** In. St.Louis, MI 63144.: TRIPOS, Associates, Inc.; 2008.

97. Matter H: **Selecting Optimally Diverse Compounds from Structure Databases: A Validation Study of Two-Dimensional and Three-Dimensional Molecular Descriptors.** *Journal of Medicinal Chemistry* 1997, **40**(8):1219-1229.
98. Martin YC, Bures MG, Danaher EA, DeLazzer J, Lico I, Pavlik PA: **Tripos** In: *DISCO Program*. St. Louis, MO 63144; 2008.
99. Gerlach C, Broughton H, Zaliani A: **FTree query construction for virtual screening: a statistical analysis.** *Journal of computer-aided molecular design* 2008, **22**(2):111-118.
100. Boecker A, Schneider G, Teckentrup A: **NIPALSTREE: A New Hierarchical Clustering Approach for Large Compound Libraries and Its Application to Virtual Screening.** *ChemInform* 2007, **38**(6).
101. Eckert H, Bajorath J: **Determination and Mapping of Activity-Specific Descriptor Value Ranges for the Identification of Active Compounds.** *Journal of Medicinal Chemistry* 2006, **49**(7):2284-2293.
102. MolecularDesign. In: *Maccs II*. 14600 Catalina St. San Leandro, CA 94577.: Ltd.
103. Daylight. In. Mission Viejo, CA, USA. <http://www.daylight.com>; 2008.
104. Krämer A, Horn HW, Rice JE: **Fast 3D molecular superposition and similarity search in databases of flexible molecules.** *Journal of computer-aided molecular design* 2003, **17**(1):13.
105. Mannhold R, Kubinyi H, Folkers G: **Ligand Design for G Protein-coupled Receptors.** Weinheim, Germany: WILEY-VCH; 2006.
106. Gilchrist A: **Second Annual GPCRs: From Orphan to Blockbuster.** *Expert Opinion on Therapeutic Patents* 2004, **8**(5):495-498.
107. Wise A, Gearing K, Rees S: **Target validation of G-protein coupled receptors.** *Drug Discovery Today* 2002, **7**(4):235-246.
108. Pertwee RG: **Cannabinoids and multiple sclerosis.** *Pharmacology & Therapeutics* 2002, **95**(2):165-174.
109. Cravatt BF, Lichtman AH: **The endogenous cannabinoid system and its role in nociceptive behavior.** *Journal of Neurobiology* 2004, **61**(1):149-160.
110. Malan TP, Jr., Ibrahim MM, Lai J, Vanderah TW, Makriyannis A, Porreca F: **CB2 cannabinoid receptor agonists: pain relief without psychoactive effects?** *Current Opinion in Pharmacology* 2003, **3**(1):62-67.
111. McKallip RJ, Lombard C, Fisher M, Martin BR, Ryu S, Grant S, Nagarkatti PS, Nagarkatti M: **Targeting CB2 cannabinoid receptors as a novel therapy to treat malignant lymphoblastic disease.** *Blood* 2002, **100**(2):627-634.
112. Beuming T: **Recent advances in docking and loop modeling accuracy using experimental GPCR structures and homology models.** *Schrodinger Spring 2012 Seminar Series* 2012.
113. Matsuda LA, Lolait SJ, Brownstein MJ, Young AC, Bonner TI: **Structure of a cannabinoid receptor and functional expression of the cloned cDNA.** *Nature* 1990, **346**(6284):561.
114. Munro S, Thomas KL, Abu-Shaar M: **Molecular characterization of a peripheral receptor for cannabinoids.** *Nature* 1993, **365**(6441):61.
115. Montero C, Campillo NE, Goya P: **Homology models of the cannabinoid CB1 and CB2 receptors. A docking analysis study.** *European Journal of Medicinal Chemistry* 2005, **40**(1):75.

116. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE *et al*: **Crystal structure of rhodopsin: A G protein-coupled receptor**. *Science* 2000, **289**(5480):739-745.
117. Baldwin JM: **The probable arrangement of the helices in G protein-coupled receptors**. *The EMBO Journal* 1993, **12**(4):1693-1703.
118. Bramblett RD, Panu AM, Ballesteros JA, Reggio PH: **Construction of a 3D model of the cannabinoid CB1 receptor: determination of helix ends and helix orientation**. *Life Sciences* 1995, **56**(23/24):1971-1982.
119. Schertler GFX, Villa C: **Projection structure of rhodopsin**. *Nature* 1993, **362**:770-772.
120. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools**. *Nucl Acids Res* 1997, **25**(24):4876-4882.
121. Mizuguchi K, Deane CM, Blundell TL, Johnson MS, Overington JP: **JOY: protein sequence-structure representation and analysis**. *Bioinformatics* 1998, **14**(7):617-623.
122. Sali A, Blundell TL: **Comparative Protein Modelling by Satisfaction of Spatial Restraints**. *Journal of Molecular Biology* 1993, **234**(3):779.
123. Fiser A, Do RK, Sali A: **Modeling of loops in protein structures [In Process Citation]**. *Protein Sci* 2000, **9**(9):1753-1773.
124. Luthy R, Bowie JU, Eisenberg D: **Assessment of protein models with three-dimensional profiles**. *Nature* 1992, **356**(6364):83.
125. Laskowski RA, McArthur W, Moss D, Thornton JM: **PROCHECK: a program to check the stereochemical quality of protein structures**. *J Appl Crystallogr* 1993, **26**:283-291.
126. Sali A, Blundell TL: **Definition of general topological equivalence in protein structures : A procedure involving comparison of properties and relationships through simulated annealing and dynamic programming**. *Journal of Molecular Biology* 1990, **212**(2):403.
127. Garcia DE, Brown S, Hille B, Mackie K: **Protein Kinase C Disrupts Cannabinoid Actions by Phosphorylation of the CB1 Cannabinoid Receptor**. *J Neurosci* 1998, **18**(8):2834-2841.
128. Xie XQ, Chen JZ, Billings EM: **3D structural model of the G-protein-coupled cannabinoid CB2 receptor**. *Proteins: Structure, Function, and Genetics* 2003, **53**(2):307-319.
129. MSI-Biosym: **InsightII/Homology v98**. In. San Diego, Ca.
130. Raitio KH, Salo OM, Nevalainen T, Poso A, Jarvinen T: **Targeting the cannabinoid CB2 receptor: mutations, modeling and development of CB2 selective ligands**. *Current Medicinal Chemistry* 2005, **12**(10):1217-1237.
131. Gonzalez A, Duran LS, Araya-Secchi R, Garate JA, Pessoa-Mahana CD, Lagos CF, Perez-Acle T: **Computational modeling study of functional microdomains in cannabinoid receptor type 1**. *Bioorganic & Medicinal Chemistry* 2008, **16**(8):4378.
132. Manfred JS: **Recognition of errors in three-dimensional structures of proteins**. *Proteins: Structure, Function, and Genetics* 1993, **17**(4):355-362.
133. Eisenberg D, Bowie JU, Charles WCJ, Sweet RM: **VERIFY3D: Assessment of protein models with three-dimensional profiles**. In: *Methods in Enzymology*. vol. 277: Academic Press; 1997: 396.

134. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kal L, Schulten K: **Scalable molecular dynamics with NAMD**. *Journal of Computational Chemistry* 2005, **26**(16):1781-1802.
135. Barth F, Rinaldi-Carmona M, Sanofi Recherche MF: **The development of cannabinoid antagonists**. *Curr Med Chem* 1999, **6**(8):745-755.
136. Rinaldi-Carmona M, Barth F, Heaulme M, Shire D, Calandra B, Congy C, Martinez S, Maruani J, Neliat G: **SR141716A, a potent and selective antagonist of the brain cannabinoid receptor**. *FEBS Lett* 1994, **350**(2-31):240-244.
137. Gaoni Y, Mechoulam R: **Isolation, structure, and partial synthesis of an active constituent of Hashish**. *J Am Chem Soc* 1964, **86**:1646.
138. Dewey WL: **Cannabinoid pharmacology**. *Pharmacol Rev* 1986, **38**(21):151-178.
139. Makriyannis A, Rapaka RS: **The molecular basis of cannabinoid activity**. *Life Sci* 1990, **47**:2173-2184.
140. Mechoulam R: **The pharmacohistory of cannabis sativa**. In: *Cannabinoids as Therapeutics Agents*. Edited by Mechoulam R: CRC Press, Boca Raton, FL; 1986: 1-20.
141. Melvin LS, M.R.. J, Rapaka PS, Makriyannis A: **Structure Activity Relationships of the Cannabinoids**. In: *NIDA Research Monograph Series*. vol. **79**: US Dept of Health and Human Services, Rockville, MD; 1987: 31-74.
142. Fride E, Mechoulam R: **Pharmacological activity of the cannabinoid receptor agonist, anandamide, a brain constituent**. *Eur J Pharmacol* 1993, **231**(21):313-314.
143. Varga K, Lake K, Martin BR, Kunos G: **Novel antagonist implicates the CB1 cannabinoid receptor in the hypotensive action of anandamide**. *Eur J Pharmacol* 1995, **278**(31):279-283.
144. Pertwee RG, Stevenson LA, Griffin G: **Cross-tolerance between delta-9-tetrahydrocannabinol and the cannabimimetic agents CP 55,940, WIN 55,212-2 and anandamide**. *Br J Pharmacol* 1993, **110**:1483-1490.
145. Bell MR, D'Ambra TE, Kumar V, Eissenstat MA, Herrmann JL, Jr., Wetzel JR, Rosi D, Pillion RE, Daum SJ: **Antinociceptive (aminoalkyl)indoles**. *J Med Chem* 1991, **34**(31):1099-1110.
146. Hlasta DJ, Luttinger D, Perrone MH, Silbernagel MJ, Ward SJ, Haubrich DR: **α 2-Adrenergic agonists/antagonists: the synthesis and structure-activity relationships of a series of indolin-2-yl and tetrahydroquinolin-2-yl imidazolines**. *J Med Chem* 1987, **30**(91):1555-1562.
147. Pacheco M, Childers SR, Arnold R, Casiano F, Ward SJ: **Aminoalkylindoles: actions on specific G-protein-linked receptors**. *J Pharmacol Exp Ther* 1991, **257**(11):170-183.
148. Ward SJ, Mastriani D, Casino F, Arnold R: **Pravadoline: profile in isolated tissue preparations**. *Pharmacol Exp Ther* 1990, **255**:1230-1239.
149. Razdan RJ, L. H, Martin BR: **2-Arachidonylglycerol and its analogs: their synthesis and pharmacological activity**. In: *ICRS Meeting abstract: 1999; Acapulco, Mexico; 1999*: 3.
150. Mechoulam R, Ben-Shabat S, Hanus L, Ligumsky M, Kaminski NE, ; Schatz AR, Gopher A, Almog S, Martin BR, al. e: **Identification of an endogenous 2-monoglyceride, present in canine gut, that binds to cannabinoid receptors**. *Biochem Pharmacol* 1995, **50**:83-90.
151. Gareau Y, Dufresne C, Gallant M, Rochette C, Sawyer N, Slipetz DM, Tremblay N, Weech PK, Metters KM, Labelle M: **Structure activity relationships of**

- tetrahydrocannabinol analogs on human cannabinoid receptors. *Bioorg Med Chem Lett* 1996, **6**(21):189-194.**
152. Zhang W, Fan P, Makriyannis A, Redda KK: **Synthesis of biphenylic cannabinoid analogs and structure-activity relationship studies.** In: *218th ACS National Meeting: Aug. 22-26 1999; New Orleans; 1999.*
153. Lavey B, Kozlowski JA, Hipkin RW, Gonsiorek W, Lundell DJ, Piwinski J, Narula SK, Lunn CA: **Triaryl bis-sulfones as a new class of cannabinoid CB2 receptor inhibitors: identification of a lead and initial SAR studies.** *Bioorg Med Chem Lett* 2005, **15**:783-786.
154. Lavey B, Kozlowski JA, Shankar B, Spitler J, Zhou G, Yang D-Y, Shu Y, Wong M, Wong S-C, Shin N-Y: **Optimization of triaryl bis-sulfones as cannabinoid-2 receptor ligands.** *Bioorg Med Chem Lett* 2007, **17**:3760-3764.
155. Shankar B, Lavey B, Zhou G, Spitler J, Tong L, Rizvi R, Yang D-P, Wolin R, Kozlowski JA, Shih N-Y: **Triaryl bis-sulfones as cannabinoid-2 receptor ligands: SAR studies.** *Bioorg Med Chem Lett* 2005, **15**:4417-4420.
156. Ross RA, Brockie HC, Stevenson LA, Murphy VL, Templeton F, Makriyannis A, Pertwee RG: **Agonist-inverse agonist characterization at CB1 and CB2 cannabinoid receptors of L759633, L759656 and AM630.** *Br J Pharmacol* 1999, **126**(3):665-672.
157. Yang P, Wang L, Xie XQ: **Latest advances in novel cannabinoid CB(2) ligands for drug abuse and their therapeutic potential.** *Future Med Chem* 2012, **4**(2):187-204.
158. Rogers-Evans M, Alanine AI, Bleicher K, Kube D, Schneider G: **Identification of novel cannabinoid receptor ligands via evolutionary de novo design and rapid parallel synthesis.** *QSAR & Combinatorial Science* 2004, **23**(6):426-430.
159. Schneider G, Lee M-L, Stahl M, Schneider P: **De novo design of molecular architectures by evolutionary assembly of drug-derived building blocks.** *Journal of computer-aided molecular design* 2000, **14**(5):487.
160. Lewell XQ, Judd DB, Watson SP, Hann MM: **RECAP-Retrosynthetic Combinatorial Analysis Procedure: A Powerful New Technique for Identifying Privileged Molecular Fragments with Useful Applications in Combinatorial Chemistry.** *J Chem Inf Comput Sci* 1998, **38**(3):511-522.
161. Chen JZ, Xie XQ: **GPCR Structure-Based Virtual Screening Approach for the CB2 Antagonist Search.** *J Comput Info Modeling* 2007 **47**:1626-1637.
162. Muegge I, Martin YC: **A General and Fast Scoring Function for Protein Ligand Interactions: A Simplified Potential Approach.** *Journal of Medicinal Chemistry* 1999, **42**(5):791-804.
163. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP: **Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes.** *Journal of computer-aided molecular design* 1997, **11**(5):425.
164. Ewing TJA, Kuntz ID: **Critical evaluation of search algorithms for automated molecular docking and database screening.** *Journal of Computational Chemistry* 1997, **18**(9):1175-1189.
165. Jones G, Willett P, Glen RC, Leach AR, Taylor R: **Development and validation of a genetic algorithm for flexible docking.** *Journal of Molecular Biology* 1997, **267**(3):727.

166. Myint K-Z, Xie X-Q: **Recent Advances in Fragment-Based QSAR and Multi-Dimensional QSAR Methods.** *International Journal of Molecular Sciences* 2010, **11**(10):3846-3866.
167. Kubinyi H: **QSAR and 3D QSAR in drug design Part 1: methodology.** *Drug Discovery Today* 1997, **2**(11):457.
168. Sutherland JJ, O'Brien LA, Weaver DF: **Spline-Fitting with a Genetic Algorithm: A Method for Developing Classification Structure-Activity Relationships.** *Journal of Chemical Information and Computer Sciences* 2003, **43**(6):1906.
169. Neelam S, Sonu B, Bhupinder S: **Quantitative correlation between theoretical molecular descriptors and drug-HSA binding affinities for various cox-2 inhibitors.** *Chemical Biology & Drug Design* 2008, **72**(4):297-302.
170. Alvesalo JKO, Siiskonen A: **Similarity Based Virtual Screening: A Tool for Targeted Library Design.** *J Med Chem* 2006, **49**:2353-2356.
171. Bostrom J, Hogner A, Schmitt S: **Do Structurally Similar Ligands Bind in a Similar Fashion?** *Journal of Medicinal Chemistry* 2006, **49**(23):6716-6725.
172. Johnson M, Maggiora G: **Concepts and Applications of Molecular Similarity.** New York: John Wiley; 1990.
173. Johnson M, Maggiora G, Lajiness M, Moon JB, Petke JD, Rohrer DC: **Molecular similarity analysis: applications in drug discovery.** *Methods Principles Med Chem* 1995, **3**:89-110.
174. Lajiness M, Johnson M, Maggiora G: **Molecular similarity: a basis for designing drug screening programs.** *Prog Clin Biol Res* 1989, **291**:167-171.
175. Lewis RA, Pickett SD, Clark DE: **Computer-Aided Molecular Diversity Analysis and Combinatorial Library Design.** In: *Reviews in Computational Chemistry.* Edited by Kenny B. Lipkowitz DBB; 2007: 1-51.
176. Pozzan A: **Molecular descriptors and methods for ligand based virtual high throughput screening in drug discovery.** *Curr Pharm Des* 2006, **12**(17):2099-2110.
177. Willett P: **Similarity-based virtual screening using 2D fingerprints.** *Drug Discovery Today* 2006, **11**(23-24):1046.
178. **Matlab.** In., 7.5.0.342 (R2007b) edn. Natick, MA. USA, 2006; <http://www.mathworks.com/products/matlab/>.
179. Tanimoto TT: **Tanimoto Coefficient.** In: *IBM Internal Report 17th Nov.* 1957.
180. **Tripos Bookshelf 8.0, Tripos International.** In. 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
181. **Template Numerical Toolkit: an interface for scientific computing in C++** [<http://math.nist.gov/tnt/index.html>]
182. Zhong H, Bowen JP: **GALAHAD pharmacophore modeling for drug discovery.** *Journal of the American Chemical Society* 2007, **129**(17):5780-5780.
183. PubChem: **PubChem substance and bioassay data source information.** In. (2010). Available at <http://pubchem.ncbi.nlm.nih.gov/>.
184. Xie XQ: **Exploiting PubChem for Virtual Screening.** *Expert Opinion on Drug Discovery* 2010, **5**(12):1205-1220.
185. Jain AN: **Ligand-Based Structural Hypotheses for Virtual Screening.** *Journal of Medicinal Chemistry* 2004, **47**(4):947-961.
186. Chohan KK, Paine SW, Waters NJ: **Quantitative structure activity relationships in drug metabolism.** *Curr Top Med Chem* 2006, **6**(15):1569-1578.

187. Clark DE: **Rapid calculation of polar molecular surfaces and its application to the prediction of transport phenomena.** *J Pharm Sci* 1999, **88**(815):815-821.
188. Wermuth CG, Ganellin CR, Lindberg P, Mitscher LA: **Glossary of terms used in medicinal chemistry (IUPAC recommendations 1998).** *Pure and Applied Chemistry* 1998, **70**(5):1129-1143.
189. Hawkins PCD, Skillman AG, Nicholls A: **Comparison of Shape-Matching and Docking as Virtual Screening Tools.** *Journal of Medicinal Chemistry* 2007, **50**(1):74-82.
190. Pandit D, So S-S, Sun H: **Enhancing specificity and sensitivity of pharmacophore-based virtual screening by incorporating chem. and shape features-A case study of HIV protease inhibitors.** *Journal of Chemical Information and Modeling* 2006, web release.
191. Singh J, Chuaqui CE, Boriack-Sjodin PA, Lee W-C, Pontz T, Corbley MJ, Cheung H-K, Arduini RM, Mead JN, Newman MN *et al*: **Successful shape-Based virtual screening: The discovery of a potent inhibitor of the type I TGF β receptor kinase (TbRI).** *Bioorganic & Medicinal Chemistry Letters* 2003, **13**(24):4355-4359.
192. Bolton E: **PubChem : An information resource linking chemistry and biology.** In: *232nd ACS National Meeting: Sept. 10-14 2006; San Francisco, CA, United States:* American Chemical Society; 2006.
193. Clark RD, Fox PC: **Statistical variation in progressive scrambling.** *J Comput Aided Mol Des* 2004, **18**(7-9):563-576.
194. Clark RD, Sprous DG, Leonard JM: **Validating models based on large data sets.** In: *Rational Approaches to Drug Design.* Edited by Holtje H-D, Sippl W. Barcelona: Prous Science; 2001: 475-485.
195. Amescua G, Miller D, Alfonso EC: **What is causing the corneal ulcer? Management strategies for unresponsive corneal ulceration.** *Eye* 2011.
196. Cheng F, Yu Y, Shen J, Yang L, Li W, Liu G, Lee PW, Tang Y: **Classification of Cytochrome P450 Inhibitors and Noninhibitors Using Combined Classifiers.** *Journal of Chemical Information and Modeling* 2011, **51**(5):996-1011.
197. Jack DA, Schache B, Smith DE: **Neural network-based closure for modeling short-fiber suspensions.** *Polymer Composites* 2010, **31**(7):1125-1141.
198. Jung E, Choi S-H, Lee N, Kang S-K, Choi Y-J, Shin J-M, Choi K, Jung D: **Machine learning study for the prediction of transdermal peptide.** *Journal of Computer-Aided Molecular Design* 2011, **25**(4):339-347.
199. Kaiserman I, Rosner M, Pe'er J: **Forecasting the Prognosis of Choroidal Melanoma with an Artificial Neural Network.** *Ophthalmology* 2005, **112**(9):1608.e1601-1608.e1606.
200. Meyer B, Hansen T, Nute D, Albersheim P, Darvill A, York W, Sellers J: **Identification of the $^1\text{H-NMR}$ spectra of complex oligosaccharides with artificial neural networks.** *Science* 1991, **251**(4993):542-544.
201. Parhizgar H, Dehghani MR, Khazae a, Dalirian M: **Application of Neural Networks in Prediction of Surface Tensions of Binary Mixtures.** *Industrial & Engineering Chemistry Research* 2012.
202. Vilar S, Santana L, Uriarte E: **Probabilistic Neural Network Model for the In Silico Evaluation of Anti-HIV Activity and Mechanism of Action.** *Journal of Medicinal Chemistry* 2006, **49**(3):1118-1124.

203. González-Díaz H, Bonet I, Terán C, De Clercq E, Bello R, García MM, Santana L, Uriarte E: **ANN-QSAR model for selection of anticancer leads from structurally heterogeneous series of compounds.** *European Journal of Medicinal Chemistry* 2007, **42**(5):580-585.
204. Patra JC, Chua BH: **Artificial neural network-based drug design for diabetes mellitus using flavonoids.** *Journal of Computational Chemistry* 2011, **32**(4):555-567.
205. Molnár L, Keserű GM: **A neural network based virtual screening of cytochrome P450 3A4 inhibitors.** *Bioorganic & Medicinal Chemistry Letters* 2002, **12**(3):419-421.
206. Muresan S, Sadowski J: **“In-House Likeness”:** **Comparison of Large Compound Collections Using Artificial Neural Networks.** *Journal of Chemical Information and Modeling* 2005, **45**(4):888-893.
207. Sutherland JJ, O'Brien LA, Weaver DF: **A Comparison of Methods for Modeling Quantitative Structure–Activity Relationships.** *Journal of Medicinal Chemistry* 2004, **47**(22):5541-5554.
208. DePriest SA, Mayer D, Naylor CB, Marshall GR: **3D-QSAR of angiotensin-converting enzyme and thermolysin inhibitors: a comparison of CoMFA models based on deduced and experimentally determined active site geometries.** *Journal of the American Chemical Society* 1993, **115**(13):5372-5384.
209. Sugimoto H, Tsuchiya Y, Sugumi H, Higurashi K, Karibe N, Iimura Y, Sasaki A, Araki S, Yamanishi Y, Yamatsu K: **Synthesis and structure-activity relationships of acetylcholinesterase inhibitors: 1-benzyl-4-(2-phthalimidoethyl)piperidine, and related derivatives.** *Journal of Medicinal Chemistry* 1992, **35**(24):4542-4548.
210. Sugimoto H, Tsuchiya Y, Sugumi H, Higurashi K, Karibe N, Iimura Y, Sasaki A, Kawakami Y, Nakamura T: **Novel piperidine derivatives. Synthesis and anti-acetylcholinesterase activity of 1-benzyl-4-[2-(N-benzoylamino)ethyl]piperidine derivatives.** *Journal of Medicinal Chemistry* 1990, **33**(7):1880-1887.
211. Haefely W, Kyburz E, Gerecke M, Mohler H: **Recent advances in the molecular pharmacology of benzodiazepine receptors and in the structure-activity relationships of their agonists and antagonists.** *Adv Drug Res* 1985, **14**:165-322.
212. Chavatte P, Yous S, Marot C, Baurin N, Lesieur D: **Three-Dimensional Quantitative Structure–Activity Relationships of Cyclo-oxygenase-2 (COX-2) Inhibitors: A Comparative Molecular Field Analysis.** *Journal of Medicinal Chemistry* 2001, **44**(20):3223-3230.
213. Talley JJ, Brown DL, Carter JS, Graneto MJ, Koboldt CM, Masferrer JL, Perkins WE, Rogers RS, Shaffer AF, Zhang YY *et al*: **4-[5-Methyl-3-phenylisoxazol-4-yl]-benzenesulfonamide, Valdecoxib: A Potent and Selective Inhibitor of COX-2.** *Journal of Medicinal Chemistry* 2000, **43**(5):775-777.
214. Huang H-C, Li JJ, Garland DJ, Chamberlain TS, Reinhard EJ, Manning RE, Seibert K, Koboldt CM, Gregory SA, Anderson GD *et al*: **Diarylspiro[2.4]heptenes as Orally Active, Highly Selective Cyclooxygenase-2 Inhibitors: Synthesis and Structure–Activity Relationships.** *Journal of Medicinal Chemistry* 1996, **39**(1):253-266.
215. Penning TD, Talley JJ, Bertenshaw SR, Carter JS, Collins PW, Docter S, Graneto MJ, Lee LF, Malecha JW, Miyashiro JM *et al*: **Synthesis and Biological Evaluation of the 1,5-Diarylpyrazole Class of Cyclooxygenase-2 Inhibitors: Identification of 4-[5-(4-**

- Methylphenyl)-3-(trifluoromethyl)-1H-pyrazol-1-yl]benzenesulfonamide (SC-58635, Celecoxib).** *Journal of Medicinal Chemistry* 1997, **40**(9):1347-1365.
216. Li JJ, Norton MB, Reinhard EJ, Anderson GD, Gregory SA, Isakson PC, Koboldt CM, Masferrer JL, Perkins WE, Seibert K *et al*: **Novel Terphenyls as Selective Cyclooxygenase-2 Inhibitors and Orally Active Anti-inflammatory Agents.** *Journal of Medicinal Chemistry* 1996, **39**(9):1846-1856.
217. Li JJ, Anderson GD, Burton EG, Cogburn JN, Collins JT, Garland DJ, Gregory SA, Huang H-C, Isakson PC: **1,2-Diarylcyclopentenes as Selective Cyclooxygenase-2 Inhibitors and Orally Active Anti-inflammatory Agents.** *Journal of Medicinal Chemistry* 1995, **38**(22):4570-4578.
218. Reitz DB, Li JJ, Norton MB, Reinhard EJ, Collins JT, Anderson GD, Gregory SA, Koboldt CM, Perkins WE: **Selective Cyclooxygenase Inhibitors: Novel 1,2-Diarylcyclopentenes Are Potent and Orally Active COX-2 Inhibitors.** *Journal of Medicinal Chemistry* 1994, **37**(23):3878-3881.
219. Khanna IK, Yu Y, Huff RM, Weier RM, Xu X, Koszyk FJ, Collins PW, Cogburn JN, Isakson PC, Koboldt CM *et al*: **Selective Cyclooxygenase-2 Inhibitors: Heteroaryl Modified 1,2-Diarylimidazoles Are Potent, Orally Active Antiinflammatory Agents.** *Journal of Medicinal Chemistry* 2000, **43**(16):3168-3185.
220. Khanna IK, Weier RM, Yu Y, Collins PW, Miyashiro JM, Koboldt CM, Veenhuizen AW, Currie JL, Seibert K, Isakson PC: **1,2-Diarylpyrroles as Potent and Selective Inhibitors of Cyclooxygenase-2.** *Journal of Medicinal Chemistry* 1997, **40**(11):1619-1633.
221. Khanna IK, Weier RM, Yu Y, Xu XD, Koszyk FJ, Collins PW, Koboldt CM, Veenhuizen AW, Perkins WE, Casler JJ *et al*: **1,2-Diarylimidazoles as Potent, Cyclooxygenase-2 Selective, and Orally Active Antiinflammatory Agents.** *Journal of Medicinal Chemistry* 1997, **40**(11):1634-1647.
222. Gangjee A, Elzein E, Queener SF, McGuire JJ: **Synthesis and Biological Activities of Tricyclic Conformationally Restricted Tetrahydropyrido Annulated Furo[2,3-d]pyrimidines as Inhibitors of Dihydrofolate Reductases¹.** *Journal of Medicinal Chemistry* 1998, **41**(9):1409-1416.
223. Rosowsky A, Mota CE, Wright JE, Queener SF: **2,4-Diamino-5-chloroquinazoline Analogs of Trimetrexate and Piritrexim: Synthesis and Antifolate Activity.** *Journal of Medicinal Chemistry* 1994, **37**(26):4522-4528.
224. Rosowsky A, Cody V, Galitsky N, Fu H, Papoulis AT, Queener SF: **Structure-Based Design of Selective Inhibitors of Dihydrofolate Reductase: Synthesis and Antiparasitic Activity of 2,4-Diaminopteridine Analogues with a Bridged Diarylamine Side Chain.** *Journal of Medicinal Chemistry* 1999, **42**(23):4853-4860.
225. Graffner-Nordberg M, Kolmodin K, Åqvist J, Queener SF, Hallberg A: **Design, Synthesis, Computational Prediction, and Biological Evaluation of Ester Soft Drugs as Inhibitors of Dihydrofolate Reductase from *Pneumocystis carinii*.** *Journal of Medicinal Chemistry* 2001, **44**(15):2391-2402.
226. Gangjee A, Vidwans AP, Vasudevan A, Queener SF, Kisliuk RL, Cody V, Li R, Galitsky N, Luft JR, Pangborn W: **Structure-Based Design and Synthesis of Lipophilic 2,4-Diamino-6-Substituted Quinazolines and Their Evaluation as Inhibitors of Dihydrofolate Reductases and Potential Antitumor Agents¹.** *Journal of Medicinal Chemistry* 1998, **41**(18):3426-3434.

227. **Cannabinoid Ligand Database** [www.cbligand.org/cbid]
228. **Open Babel, version 2.3.0** [<http://openbabel.org>]
229. O'Boyle N, Banck M, James C, Morley C, Vandermeersch T, Hutchison G: **Open Babel: An open chemical toolbox**. *Journal of Cheminformatics* 2011, **3**:33.
230. Durant JL, Leland BA, Henry DR, Nourse JG: **Reoptimization of MDL Keys for Use in Drug Discovery**. *Journal of Chemical Information and Computer Sciences* 2002, **42**(6):1273-1280.
231. Rogers D, Hahn M: **Extended-Connectivity Fingerprints**. *Journal of Chemical Information and Modeling* 2010, **50**(5):742-754.
232. Morgan HL: **The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service**. *Journal of Chemical Documentation* 1965, **5**(2):107-113.
233. **ChemAxon** [<http://www.chemaxon.com>]
234. Greenidge PA, Carlsson B, Bladh L-G, Gillner M: **Pharmacophores incorporating numerous excluded volumes defined by x-ray crystallographic structure in three-dimensional database searching: Application to the thyroid hormone receptor**. *J Med Chem* 1998, **41**(14):2503-2512.
235. Gertsch J, Leonti M, Raduner S, Racz I, Chen J-Z, Xie X-Q, Altmann K-H, Karsak M, Zimmer A: **Beta-caryophyllene is a dietary cannabinoid**. *Proceedings of the National Academy of Sciences* 2008, **105**(26):9099-9104.
236. Raduner S, Majewska A, Chen J-Z, Xie X-Q, Hamon J, Faller B, Altmann K-H, Gertsch J: **Alkylamides from Echinacea Are a New Class of Cannabinomimetics: cannabinoid type 2 receptor-dependent and -independent immunomodulatory effects**. *Journal of Biological Chemistry* 2006, **281**(20):14192-14206.
237. Zhang Y, Xie Z, Wang L, Schreiter B, Lazo JS, Gertsch J, Xie X-Q: **Mutagenesis and computer modeling studies of a GPCR conserved residue W5.43(194) in ligand recognition and signal transduction for CB2 receptor**. *International Immunopharmacology* 2011, **11**(9):1303-1310.
238. Ferguson AM, Heritage T, Jonathon P, Pack SE, Phillips L, Rogan J, Snaith PJ: **EVA: A new theoretically based molecular descriptor for use in QSAR/QSPR analysis**. *Journal of Computer-Aided Molecular Design* 1997, **11**(2):143-152.
239. Agrafiotis DK, Cedeño W, Lobanov VS: **On the Use of Neural Network Ensembles in QSAR and QSPR**. *Journal of Chemical Information and Computer Sciences* 2002, **42**(4):903-911.
240. Bender A, Jenkins JL, Scheiber J, Sukuru SCK, Glick M, Davies JW: **How Similar Are Similarity Searching Methods? A Principal Component Analysis of Molecular Descriptor Space**. *Journal of Chemical Information and Modeling* 2009, **49**(1):108-119.
241. Glem R, Bender A, Arnby C, Carlsson L, Boyer S, Smith J: **Circular fingerprints: flexible molecular descriptors with applications from physical chemistry to ADME**. *IDrugs* 2006, **9**(3):199-204.
242. Bellis LJ, Akhtar R, Al-Lazikani B, Atkinson F, Bento AP, Chambers J, Davies M, Gaulton A, Hersey A, Ikeda K *et al*: **Collation and data-mining of literature bioactivity data for drug discovery**. *Biochem Soc Trans* 2011, **39**(5):1365-1370.
243. **Drug Synthesis and Chemistry Branch, Developmental Therapeutics Program (DTP), Division of Cancer Treatment and Diagnosis, National Cancer Institute** [<http://dtp.nci.nih.gov/>]

244. Huffman JW, Yu S, Showalter V, Abood ME, Wiley JL, Compton DR, Martin BR, Bramblett RD, Reggio PH: **Synthesis and Pharmacology of a Very Potent Cannabinoid Lacking a Phenolic Hydroxyl with High Affinity for the CB2 Receptor.** *Journal of Medicinal Chemistry* 1996, **39**(20):3875-3877.
245. Huffman JW: **Cannabimimetic indoles, pyrroles and indenes.** *Current Medicinal Chemistry* 1999, **6**(8):705-720.
246. Palmer SL, Thakur GA, Makriyannis A: **Cannabinergic ligands.** *Chem Phys Lipids* 2002, **121**(1-2):3-19.
247. Marriott KS, Huffman JW: **Recent advances in the development of selective ligands for the cannabinoid CB(2) receptor.** *Curr Top Med Chem* 2008, **8**(3):187-204.
248. Rinaldi-Carmona M, Barth F, Millan J, Derocq JM, Casellas P, Congy C, Oustric D, Sarran M, Bouaboula M, Calandra B *et al*: **SR 144528, the first potent and selective antagonist of the CB2 cannabinoid receptor.** *J Pharmacol Exp Ther* 1998, **284**(2):644-650.
249. Iwamura H, Suzuki H, Ueda Y, Kaya T, Inaba T: **In vitro and in vivo pharmacological characterization of JTE-907, a novel selective ligand for cannabinoid CB2 receptor.** *J Pharmacol Exp Ther* 2001, **296**(2):420-425.
250. Lunn CA, Fine JS, Rojas-Triana A, Jackson JV, Fan X, Kung TT, Gonsiorek W, Schwarz MA, Lavey B, Kozlowski JA *et al*: **A novel cannabinoid peripheral cannabinoid receptor-selective inverse agonist blocks leukocyte recruitment in vivo.** *Journal of Pharmacology and Experimental Therapeutics* 2006, **316**(2):780-788.
251. Schuehly W, Paredes JM, Kleyer J, Huefner A, Anavi-Goffer S, Raduner S, Altmann KH, Gertsch J: **Mechanisms of osteoclastogenesis inhibition by a novel class of biphenyl-type cannabinoid CB(2) receptor inverse agonists.** *Chem Biol* 2011, **18**(8):1053-1064.
252. Giblin GM, O'Shaughnessy CT, Naylor A, Mitchell WL, Eatherton AJ, Slingsby BP, Rawlings DA, Goldsmith P, Brown AJ, Haslam CP *et al*: **Discovery of 2-[(2,4-dichlorophenyl)amino]-N-[(tetrahydro- 2H-pyran-4-yl)methyl]-4-(trifluoromethyl)-5-pyrimidinecarboxamide, a selective CB2 receptor agonist for the treatment of inflammatory pain.** *J Med Chem* 2007, **50**(11):2597-2600.
253. Xi Z-X, Peng X-Q, Li X, Song R, Zhang H-Y, Liu Q-R, Yang H-J, Bi G-H, Li J, Gardner EL: **Brain cannabinoid CB2 receptors modulate cocaine's actions in mice.** *Nat Neurosci* 2011, **14**(9):1160-1166.
254. Feng R, Anderson G, Xiao G, Elliott G, Leoni L, Mapara MY, Roodman GD, Lentzsch S: **SDX-308, a nonsteroidal anti-inflammatory agent, inhibits NF-kappaB activity, resulting in strong inhibition of osteoclast formation/activity and multiple myeloma cell growth.** *Blood* 2007, **109**(5):2130-2138.
255. Feng R, Ma H, Hassig CA, Payne JE, Smith ND, Mapara MY, Hager JH, Lentzsch S: **KD5170, a novel mercaptoketone-based histone deacetylase inhibitor, exerts antimyeloma effects by DNA damage and mitochondrial signaling.** *Mol Cancer Ther* 2008, **7**(6):1494-1505.
256. Jaakola V-P, Griffith MT, Hanson MA, Cherezov V, Chien EYT, Lane JR, Ijzerman AP, Stevens RC: **The 2.6 Angstrom Crystal Structure of a Human A2A Adenosine Receptor Bound to an Antagonist.** *Science* 2008, **322**(5905):1211-1217.

257. Chien EYT, Liu W, Zhao Q, Katritch V, Won Han G, Hanson MA, Shi L, Newman AH, Javitch JA, Cherezov V *et al*: **Structure of the Human Dopamine D3 Receptor in Complex with a D2/D3 Selective Antagonist**. *Science* 2010, **330**(6007):1091-1095.
258. Okada T, Fujiyoshi Y, Silow M, Navarro J, Landau EM, Shichida Y: **Functional role of internal water molecules in rhodopsin revealed by x-ray crystallography**. *Proceedings of the National Academy of Sciences* 2002, **99**(9):5982-5987.
259. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK *et al*: **High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor**. *Science* 2007, **318**(5854):1258-1265.
260. Warne T, Moukhametzianov R, Baker JG, Nehme R, Edwards PC, Leslie AGW, Schertler GFX, Tate CG: **The structural basis for agonist and partial agonist action on a [bgr]1-adrenergic receptor**. *Nature* 2011, **469**(7329):241-244.
261. Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AGW, Tate CG, Schertler GFX: **Structure of a beta1-adrenergic G-protein-coupled receptor**. *Nature* 2008, **454**(7203):486.
262. Roy A, Kucukural A, Zhang Y: **I-TASSER: a unified platform for automated protein structure and function prediction**. *Nat Protocols* 2010, **5**(4):725.
263. Zhang Y: **Template-based modeling and free modeling by I-TASSER in CASP7**. *Proteins: Structure, Function, and Bioinformatics* 2007, **69**(S8):108.
264. Salo OMH, Raitio KH, Savinainen JR, Nevalainen T, Lahtela-Kakkonen M, Laitinen JT, Jaervinen T, Poso A: **Virtual Screening of Novel CB2 Ligands Using a Comparative Model of the Human Cannabinoid CB2 Receptor**. *Journal of Medicinal Chemistry* 2005, **48**(23):7166-7171.
265. Zhang R, Hurst DP, Barnett-Norris J, Reggio PH, Song Z-H: **Cysteine 2.59(89) in the Second Transmembrane Domain of Human CB2 Receptor Is Accessible within the Ligand Binding Crevice: Evidence for Possible CB2 Deviation from a Rhodopsin Template**. *Molecular Pharmacology* 2005, **68**(1):69-83.
266. Tuccinardi T, Grazia Cascio M, Di Marzo V, Manera C, Ortore G, Saccomanni G, Martinelli A: **Structure-Based Virtual Screening: Identification of Novel CB2 Receptor Ligands**. *Letters in Drug Design & Discovery* 2007, **4**(1):15-19.
267. Tuccinardi T, Ferrarini PL, Manera C, Ortore G, Saccomanni G, Martinelli A: **Cannabinoid CB2/CB1 Selectivity. Receptor Modeling and Automated Docking Analysis**. *Journal of Medicinal Chemistry* 2006, **49**(3):984-994.
268. El Bakali J, Muccioli GG, Renault N, Pradal D, Body-Malapel M, Djouina M, Hamtiaux L, Andrzejak V, Desreumaux P, Chavatte P *et al*: **4-Oxo-1,4-dihydropyridines as Selective CB2 Cannabinoid Receptor Ligands: Structural Insights into the Design of a Novel Inverse Agonist Series**. *Journal of Medicinal Chemistry* 2010, **53**(22):7918-7931.
269. Xie X-Q, Pavlopoulos S, DiMeglio CM, Makriyannis A: **Conformational Studies on a Diastereoisomeric Pair of Tricyclic Nonclassical Cannabinoids by NMR Spectroscopy and Computer Molecular Modeling**. *J Med Chem* 1998, **41**(21):167-174.
270. Xie X-Q, Melvin LS, Makriyannis A: **The conformational properties of the highly selective cannabinoid receptor ligand CP-55,940**. *J Biol Chem* 1996, **271**(181):10640-10647.

271. Myint KZ, Xie X-Q: **Recent Advances in Fragment-Based QSAR and Multi-Dimensional QSAR Methods.** *International Journal of Molecular Sciences* 2010, **11**(10):3846-3866.
272. Anandarajah AP, Schwarz EM, Totterman S, Monu J, Feng CY, Shao T, Haas-Smith SA, Ritchlin CT: **The effect of etanercept on osteoclast precursor frequency and enhancing bone marrow oedema in patients with psoriatic arthritis.** *Ann Rheum Dis* 2008, **67**(3):296-301.
273. Diaz P, Phatak SS, Xu J, Astruc-Diaz F, Cavasotto CN, Naguib M: **6-Methoxy-N-alkyl Isatin Acylhydrazones as a Novel Series of Potent Selective Cannabinoid Receptor 2 Inverse Agonists: Design, Synthesis, and Binding Mode Prediction.** *Journal of Medicinal Chemistry* 2008, **52**(2):433.
274. Gouldson P, Calandra B, Legoux P, Kerneis A, Rinaldi-Carmona M, Barth F, Le Fur G, Ferrara P, Shire D: **Mutational analysis and molecular modeling of the antagonist SR 144528 binding site on the human cannabinoid CB2 receptor.** *European Journal of Pharmacology* 2000, **401**(1):17-25.
275. Poso A, Huffman JW: **Targeting the cannabinoid CB2 receptor: modelling and structural determinants of CB2 selective ligands.** *Brit J Pharmacol* 2008, **153**(2):335.