

KADIR HAS UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING



PHARMACOPHORE-BASED SCREENING AND DOCKING FOR
THE DISCOVERY OF NOVEL ANTAGONISTS OF BETA-2
ADRENERGIC RECEPTOR

GRADUATE THESIS

RÜYA YAKAR

May,2013

Rüya Yakar

M.S Thesis

2013

PHARMACOPHORE-BASED SCREENING AND DOCKING FOR
THE DISCOVERY OF NOVEL ANTAGONISTS OF BETA-2
ADRENERGIC RECEPTOR

RÜYA YAKAR

Submitted to the Graduate School of Science and Engineering in partial fulfillment
of the requirements for the degree of
Master of Science
In
COMPUTATIONAL BIOLOGY AND BIOINFORMATICS

KADIR HAS UNIVERSITY
May, 2013

KADIR HAS UNIVERSITY
GRADUATE SHOOLOF SCIENCE AND ENGINEERING

PHARMACOPHORE-BASED SCREENING AND DOCKING FOR THE
DISCOVERY OF NOVEL ANTAGONISTS OF BETA-2 ADRENERGIC
RECEPTOR

RÜYA YAKAR

APPROVED BY:

Assist. Prof Dr. E. Demet Akdoğan
(Thesis Supervisor)

Prof. Dr. Kemal Yelekçi

Prof. Dr. Viktorya Aviyente

APPROVAL DATE: 16/May/2013

PHARMACOPHORE-BASED SCREENING AND DOCKING FOR THE DISCOVERY OF NOVEL ANTAGONISTS OF BETA-2 ADRENERGIC RECEPTOR

Abstract

β_2 AR, which is the member of rhodopsin-like GPCR, is the target system for the discovery of novel antagonists using structure-based pharmacophore modeling and docking methods. Initially, a shared pharmacophore model is obtained using the structure of five known inactive β_2 AR complex (PDB ids: 2HR1, 3D4S, 3NY8, 3NY9 and 3NYA). In order to test the discriminatory power of pharmacophore model, a small database consisting of 117 known molecules (53 antagonists against 64 agonists) was screened using LigandScout software tool. The screening yielded 44 antagonists (72% true positives) against 17 agonists (18% false positives), which was found satisfactory. Then, under the same screening conditions, the second database that is the clean drug-like subset of ZINC database was screened. Out of 9,928,465 molecules, 729,413 compounds fulfilled the pharmacophore requirements. All these molecules were docked to one of the inactive crystal structure (PDB id: 2RH1) using GOLD and AutoDock software tools along with the four scoring functions, CHEMPLP, DSX Score, GoldScore, and ChemScore. The docked pose with the highest score of each molecule was further analyzed based on some key residues for antagonist binding. Docking calculations yielded 3 antagonist molecules from our decoy set and 360 compounds from ZINC database as they have been found in the highest score range determined from all four scoring functions. Final test was based on ADMET properties provided by Discovery Studio. ADMET analysis yielded one molecule from decoy set and 62 molecules from ZINC database that were proposed as plausible antagonists for β_2 AR.

FARMAKOFOR VE DOCKING'E DAYALI TARAMA YÖNTEMİNİN β_2 AR ANTAGONİSTLERİN KEŞFİNDE KULLANILMASI

Özet

β_2 AR, rodopsin benzeri GPCR üyesi olup, yapı-tabanlı farmakofor modelleme ve docking yöntemleri kullanılarak yeni antagonistlerinin keşfi için hedef sistemidir. İlk olarak bilinen beş inaktif β_2 AR kompleksi (PDB kodu: 2RH1, 3D4S, 3NY8, 3NY9 ve 3NYA) kullanılarak ortak bir farmakofor modeli elde edilmiştir. Bu farmakofor modelinin ayırt edici özelliğini test edebilmek için bilinen 117 molekülden (53 antagoniste karşılık 64 agonist) oluşan küçük bir veritabanı LigandScout programıyla taranmıştır. Tarama sonucu 44 antagoniste (%72 doğru pozitif) karşılık 17 agonist (%18 yanlış pozitif) tatmin edici bulunmuştur. Daha sonra, aynı koşullar altında ZINC veritabanınının bir alt veritabanı olan temiz ilaç benzeri veritabanı taranmıştır. 9,928,465 molekülden 729,413 tanesi istenilen farmakofor özelliklerini karşılamıştır. Bütün bu moleküller GOLD ve AUTODOCK programları kullanılarak dört ayrı scoring fonksiyonu (CHEMPLP, DSX Score, GoldScore ve ChemScore) ile inaktif kristal yapılarından birine (PDB kodu: 2RH1) dock edilmiştir. Bu dock edilen pozlardan en yüksek skora sahip olanlar antagonistlerin bağlanması için gerekli olan bazı anahtar rezidüleri alınarak analiz edilmiştir. Docking hesaplamaları sonucunda test setten 3 antagonistin ve ZINC veritabanından ise 360 bileşiğin dört scoring fonksiyonunda da en yüksek skora sahip olduğu sonucuna varılmıştır. Son olarak Discovery Studio tarafından sağlanan ADMET özellikleri test edilmiştir. ADMET analizi sonucunda küçük setten bir molekül ve ZINC veritabanından ise 62 molekül β_2 AR için uygun antagonistler olarak belirlenmiştir.

To my Family

Acknowledgements

There are many people who helped to make my years at the graduate school most valuable. First, I thank Ebru Demet Akten, my major professor and dissertation supervisor. Having the opportunity to work with her over the years was intellectually rewarding and fulfilling.

Many thanks to Department computer staff, who patiently answered my questions and problems on word processing. I would also like to thank to my graduate student colleagues who helped me all through the years full of class work and exams.

The last words of my thanks go to my family. I thank my parents Müjdenur Celal and Fikret Yakar for their patience and encouragement. Lastly, I thank my friends for their endless support through this journey.

Table of Contents

Abstract	i
Özet	ii
Dedication	iii
Acknowledgements	iv
Table of Contents	v
List of Tables	vii
List of Figures	viii
List of Abbreviations	x
1 Introduction	1
2 Beta-2-Adrenergic Receptor	4
2.1 Structure of Beta-2-Adrenergic Receptors.....	4
2.2 Functional Role of Beta-2-Adrenergic Receptors.....	6
3 Materials and Methods	8
3.1 Generating a Pharmacophore Model.....	8
3.2 VirtualScreening.....	12
3.2.1 Pharmacophore Based Virtual Screening (PBVS).....	13
3.2.2 Docking Based Virtual Screening (DBVS).....	14
3.3 Docking methodology and Scoring Function.....	14
3.4 Receiver Operating Characteristic (ROC) Curve.....	18
3.5 ZINC Database.....	19
3.6 Rescoring.....	20
4 Results and Discussion	21
4.1 Generation of Pharmacophore Model.....	21
4.2 ZINC Database.....	24
4.3 Set of Compounds with Known Activities.....	25
4.4 Pharmacophore Screening of The Decoy Set and The ZINC	

Database.....	26
4.5 Docking and Binding Evaluations.....	33
4.6 ADMET Filtering.....	38
4.7 A Closer Look at The Hit Compounds.....	41
5 Conclusions	46
References	49
Appendix A Representation of 53 antagonists taken from GLIDA database	54
Appendix B Representation of 64 agonists taken from GLIDA database	61
Appendix C Representation of 55 molecules proposed as possible drug candidates	68
Appendix D Representation of the remaining 7 potential drug candidates that are isomers of the molecules listed in Appendix C	75
Curriculum Vitae	77

List of Tables

Table 3.1	LigandScout default distance constraints.....	9
Table 4.1	Number of pharmacophore features in 5 inactive crystal structures and the shared pharmacophore.....	22
Table 4.2	Physical properties of Clean Drug-Like Subset of ZINC Database.....	25
Table 4.3	Screening results using different values for the maximum number of omitted features.....	27
Table 4.4	Screening results when check exclusion volume option was “off”.....	29
Table 4.5	Screening results of “modified” pharmacophore model using different values for the maximum number of omitted features.....	31
Table 4.6	Scaffold Table.....	43

List of Figures

Figure 2.1 Schematic representation of GPCR.....	5
Figure 2.2 Signaling pathway of β_2 ARs.....	7
Figure 3.1 Pharmacophore representation in LigandScout.....	11
Figure 3.2 Hydrogen bond constraints.....	12
Figure 3.3 Schematic representation of the genetic algorithm.....	15
Figure 3.4 ROC Curve Model.....	19
Figure 4.1 Pharmacophore models of five X-ray crystal structures of human β_2 AR.....	23
Figure 4.2 Pharmacophore features of shared pharmacophore model....	24
Figure 4.3 ROC curves of screening results using different values for the maximum number of omitted features.....	28
Figure 4.4 ROC curves of the screening results when check exclusion volume was “off”.....	30
Figure 4.5 Two exclusion volumes.....	31
Figure 4.6 ROC curve for Run #1.....	33

Figure 4.7 Extracellular view of key residues for binding test #1.....	35
Figure 4.8 Flow chart of the screening process of the Clean Drug-Like ZINC database and the decoy set.....	36
Figure 4.9 Extracellular view of key residues for binding test #2.....	37
Figure 4.10 ADMET Plot of the decoy set.....	39
Figure 4.11 ADMET Plot of ZINC database.....	40
Figure 4.12 Docking poses of 62 molecules.....	41
Figure 4.13 2D representation of Carazolol.....	42
Figure 4.14 Three compounds that hold scaffold #4 in Table 4.6 and alprenolol scaffold.....	44
Figure 4.15 Compound #35 proposed in Tasler <i>et al.</i> study.....	45
Figure 4.16 Compound #3 and Compound #11 in Sabio <i>et al.</i> study.....	45

List of Abbreviations

β_2 AR	Human Beta-2 Adrenergic Receptor
GPCR	G-Protein Coupled Receptor
LBDD	Ligand-Based Drug Design
SBDD	Structure-Based Drug Design
PBVS	Pharmacophore-Based Virtual Screening
DBVS	Docking-Based Virtual Screening
HTS	High-Troughput Screening
VS	Virtual Screening
SF	Scoring Function
FF	Force Field
ROC	Receiver Operating Characteristic

Chapter 1

Introduction

Human β_2 -adrenergic receptors (β_2 -AR) belong to the largest subfamily of G protein-coupled receptors in the human genome, which is the rhodopsin family. Also known as seven transmembrane domain receptors (7TM receptors), they are embedded in the cell membrane and have a crucial role in signal transduction from extracellular side to intracellular side in many different physiological pathways [1]. GPCRs deal with our physiological responses to hormones, neurotransmitters and environmental stimulants and they initiate many signaling pathways [2]. Thus, many diseases such as hypertension, depression, asthma, cardiac dysfunction, and inflammation, are related to the functioning of GPCRs [3], which is among the four gene families targeted by more than 50% of drugs on market [4-6].

In 2007, Rasmussen and coworkers discovered the first X-ray crystal structure of the human β_2 -AR (PDB id: 2RH1) [7], which opened a new gate for computer-aided drug discovery. Novel β_2 -AR inhibitors have been introduced using structure-based and ligand-based computational algorithms [8-11]. Kolb *et al.*[9] screened a library of approximately 1 million compounds via docking using the X-ray structure (PDB id : 2RH1) and introduce twenty-five hits which were tested in a radioligand binding assay. Six confirmed hits were identified with K_i values ranging between 9nM and

3.2 μM . Docking-based virtual screening experiments conducted by Topiol *et al.* [10] produced new chemical classes of hits besides rediscovering the well-known hydroxylamine chemotype.

De Graaf and Rognan [12] modified the rotameric states of (Ser212) S5.43 and (Ser215) S5.46 within the binding site of the first X-ray structure, which represents the inactive state of β_2 -AR and created an “early activated” model, which was found to be more successful in distinguishing partial/full agonists from decoy ligands in docking runs. This study demonstrates that there exist only small but critical differences between agonist- and antagonist-bound structures. Three X-ray crystal structures of β_2 -AR complexed with three antagonists revealed by Wacker *et al.*[13] also demonstrate minor local structural differences that exist in the binding pocket of these complexes.

The docking-based virtual screening study performed by Vilar *et al.* [14] using X-ray structure of β_2 -AR (PDB id: 2RH1) revealed that antagonists (blockers) were preferred over agonists. This was a promising result since the structure of the receptor used as a target was the apo form of the structure complexed with a partial inverse agonist carazolol and thus represents an inactive state. In addition, using an ensemble of alternative conformations of the receptor in order to account protein flexibility, they were able to increase the number of hits within the top 0.5% of the screened database.

In addition to structure-based approaches, a ligand-based drug screening study by Tasler *et al.* [15] revealed a selective and potent human β_2 -AR antagonist. The screening was based on a pharmacophore alignment on known β_3 -adrenoceptor ligands, which generated a set of β -adrenoceptor ligands. Their binding affinities were measured in various binding assays. Upon further optimization of these ligands, a selective and potent human β_2 -AR antagonist with a K_i value of 0.3 nM was introduced.

In this thesis, we conducted a pharmacophore-based virtual screening experiment in order to reveal potential human β_2 -AR antagonists. The novelty of this work is the pharmacophore model, which has been generated using five different X-ray crystal structures of β_2 -AR complexed with five different antagonists. As of today, no virtual screening study based on structural information of the receptor has been reported. The screened database was the “clean drug-like” subset of ZINC database [16]. In addition 64 known agonists and 53 known antagonists obtained from GLIDA database [17] was used in order to test the discriminatory power of the pharmacophore model. A series of docking experiments have been conducted for the compounds that satisfied the pharmacophore requirements using the apo form of the X-ray complex structure (PDB id: 2RH1). Compounds with highest binding affinities have been extracted and finally evaluated based on their predicted ADMET properties.

Chapter 2

Beta-2-Adrenergic Receptor

2.1 Three-Dimensional Structure of Beta-2 Adrenergic Receptors

Beta-2 adrenergic receptor (β_2 AR) is the member of G protein-coupled receptors (GPCRs), which is the largest protein family in the human genome. GPCRs were classified according to their physiological and structural features, position of the binding pocket and the sequence length of the receptors. This so-called GRAFS classification system combines all GPCRs into five main families [18]: **G**lutamate-like (Family C), **R**hodopsin-like (Family A), **A**dhesion-like (Family B), **F**rizzled/Taste2 (Family O) and **S**ecretin-like (Family B). Bovine rhodopsin was the first GPCR of which the X-ray crystal structure was determined. β_2 ARs belong to rhodopsin-like family because of the structural similarity [19].

β_2 AR has seven transmembrane (TM) domains, which is the main characteristics of the structure (Figure 2.1a). These seven TM domains are connected to each other with three extracellular loops and three intracellular loops. GPCRs are membrane-embedded proteins. Amino terminus is located in extracellular milieu and carboxyl terminus is located in cytosol (Figure 2.1b).

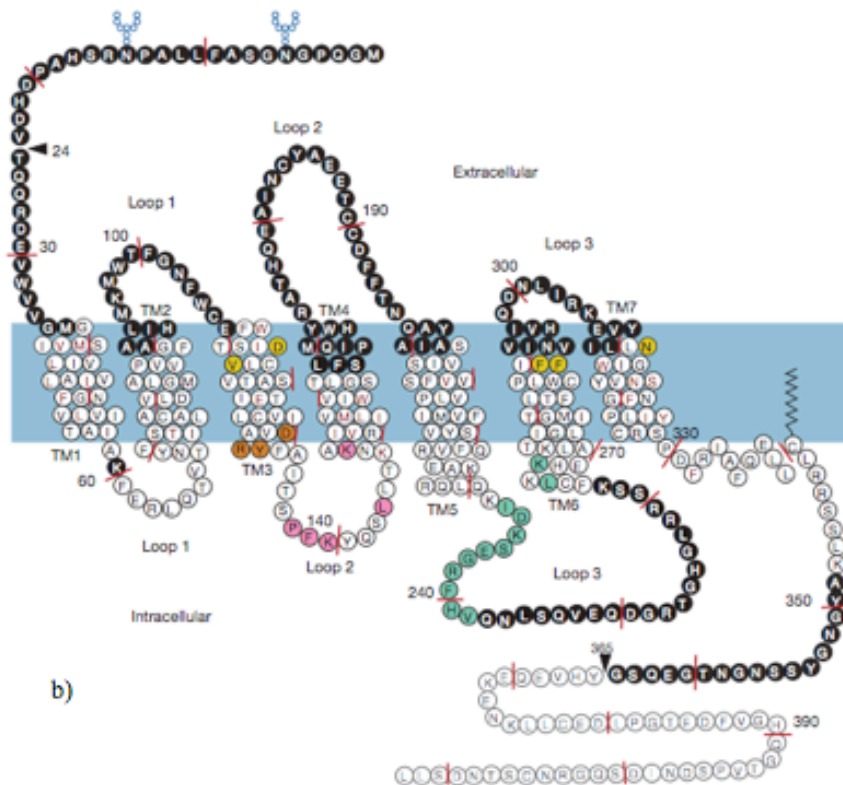
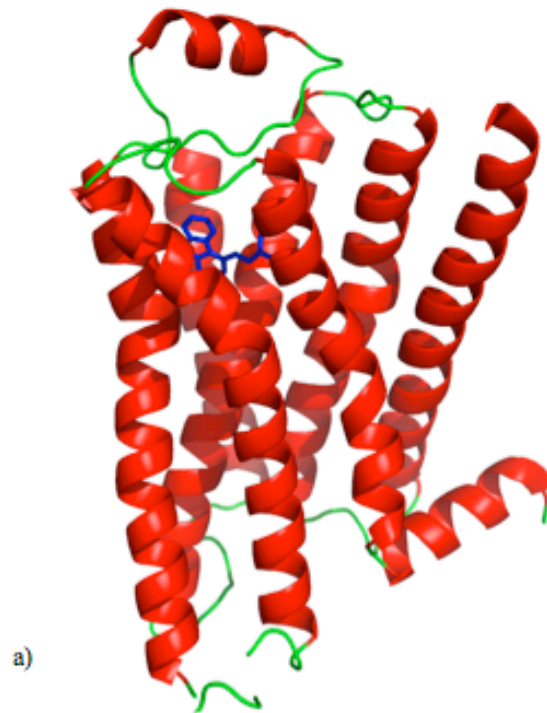


Figure 2.1. Schematic representation of GPCRs. (a) 3D modeling of β_2 AR in cartoon representation. Red shows the 7TMs, greens are extracellular and intracellular loops and the inverse agonist Carazolol in blue with stick representation. The figure was prepared with Pymol. (b) 2D representation of β_2 AR embedded in the membrane [2.16].

2.2 Functional Role of Beta-2-Adrenergic Receptors

β_2 ARs are responsible for signal transduction between the cell and the environment. They become active by hormones and neurotransmitters [20]. Figure 2.2 demonstrates the activation mechanism of β_2 AR [21]. Upon activation of the receptor by a bound agonist diffusing from extracellular side, β_2 AR interacts with a G protein nearby. G-protein consists of three subunits: α , β and γ . Followed by contact, guanosine diphosphate (GDP) inside the α -subunit is replaced by guanosine triphosphate (GTP). The affinity of α -subunit is dramatically reduced for β_2 AR, thus dissociates and couples with adenylate cyclase (A-C in Figure 2.2). Adenylyl cyclase causes an increase in cyclic adenosine monophosphate (cAMP) level, which catalyzes the activation of protein kinase A (PKA). PKA induces phosphorylation (Pi) of β_2 AR and uncoupling of G_s . The phosphorylated β_2 AR bound to β and γ subunits of G protein, couples to G_i proteins. β and γ subunits act as a scaffold for intracellular proteins such as β -arrestin, Src, SOS and RAS, which leads to the activation of mitogen-activated protein kinase (MAPK) pathway.

β_2 AR acts on smooth muscle relaxation, therefore its activation through agonists are used as treatment of asthma. It is believed that pKA phosphorylates key regulatory proteins involved in the control of muscle tone. In addition, cAMP decreases the Ca^{+2} level inside the cell, leading to muscle relaxation in the pulmonary airways. β_2 AR also increases heart rate and cardiac muscle contractility with a minor degree compared to β_1 AR. In addition, its stimulation increases pressure inside eye, therefore for glaucoma patients, its inhibition with an antagonist such as timolol, is required.

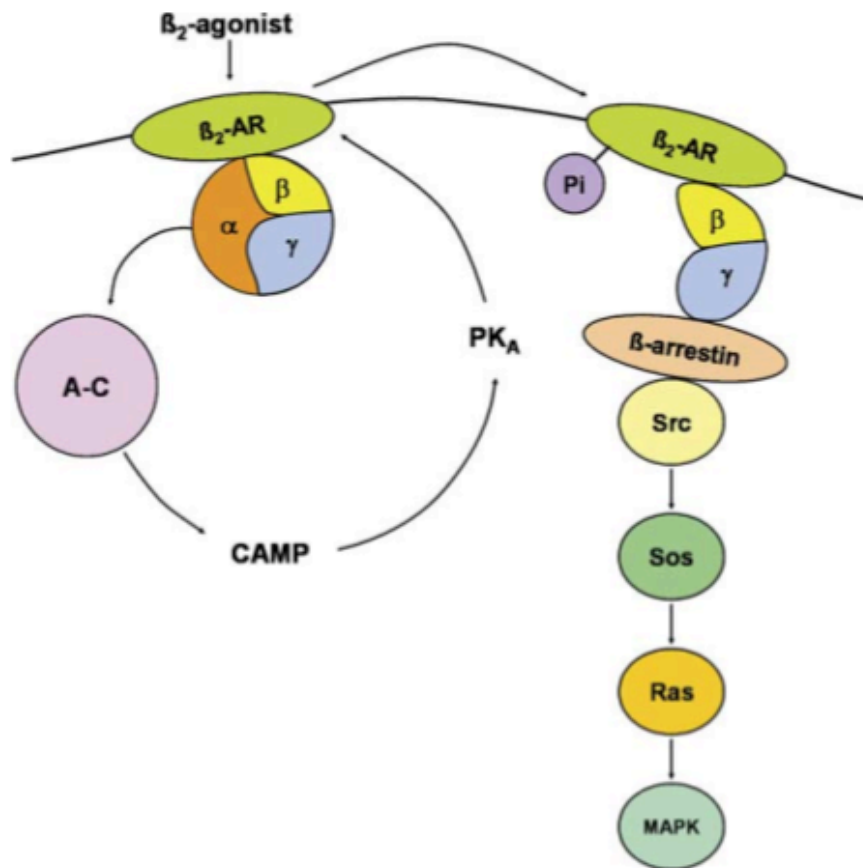


Figure 2.2. Intracellular signaling pathways for β_2 ARs.

Chapter 3

Materials and Methods

3.1 Generating a Pharmacophore Model

The term pharmacophore was defined as, “*the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response*” by Wermunt *et al.* in 1968 [22]. Since then, it became a standard definition for pharmacophore concept accepted worldwide. Pharmacophores are not real molecules or they are not functional groups. They just represent the essential points of drug molecules for an optimal interaction with the target protein [23] and these interactions are Hydrogen-bonding, charge transfer, electrostatic and hydrophobic interactions. The pharmacophore model combined all these interaction in 3D space.

In modern drug discovery, there are two types of techniques, which help to design a drug: ligand-based drug design (LBDD) and structure-based drug design (SBDD). LBDD techniques rely on a set of drug molecules that are known to bind to a certain target with a known activity. The pharmacophore model consists of the chemical features shared by all known drug molecules. On the other hand, SBDD techniques rely on the knowledge of a three-dimensional structure of the protein [1] and the native state of the bound ligand. The pharmacophore model is composed of the interacting groups of the ligand with surrounding residues of the target protein.

With the increase in the amount of experimental data about the protein's crystal structures, SBDD has become more powerful than LBDD [24]. Many popular software programs that generate pharmacophore models emerged and a few of them are Phase by Schrödinger [25], GASP (Genetic Algorithm Superposition Program) by Jones *et al.* and Catalyst by Accelrys [26] and finally LigandScout by inte:ligand [27], which was used in thesis.

LigandScout is using a structure-based approach to generate the pharmacophore model. It also provides a powerful visualization interface. It represents pharmacophore points by using special features, which are illustrated in Figure 3.1. These features determined based on two constraints: distance and direction. LigandScout selects two points. One point must be located on the protein while the other one must be located on ligand. It analyzes the relationship between these two points by distance constraints. Based on the information in LigandScout, default distance constraints are listed as in Table 3.1. On the other hand, to analyze the relationship between two atom groups instead of points, LigandScout uses direction constraints.

Table 3.1. LigandScout's default distance constraints.

Aromatic interaction with positive ionizable	3.5 – 5.5 Å
Aromatic interaction with ring (parallel)	2.8 – 4.5 Å
Aromatic interaction with ring (orthogonal)	2.8 – 4.5 Å
H-Bond interaction	2.2 – 3.8 Å
Hydrophobic interaction	1.0 – 5.9 Å
Iron binding location	1.3 – 3.5 Å
Magnesium binding location	1.5 – 3.8 Å
Negative ionizable interaction	1.5 – 5.5 Å
Positive ionizable interaction with negative ionizable	1.5 – 5.5 Å
Positive ionizable with aromatic ring	1.0 – 10.0 Å
Zinc binding location	1.0 – 4.0 Å

As shown in Figure 3.1, hydrogen bond features are represented by both distance and direction constraints while hydrophobic interactions and ionizable areas are shown with only a distance constraint. LigandScout can distinguish between flexible hydrogen bonds and rigid hydrogen bonds. Rigid hydrogen bond donors occur at sp^2 hybridized heavy atoms and they are illustrated by a cone with cutoff apex (Figure 3.2.a). Flexible hydrogen bond donors occur at sp^3 hybridized heavy atoms and they are illustrated by a torus (Figure 3.2.b). For a valid hydrogen bond, the angle between Hydrogen-bond atom vector and hydrogen vector must be below the defined constraints (default: 50.00° for rigid H-bond, 34.00° for flexible H-bond). To define a hydrogen bond acceptor, LigandScout takes neighboring atoms in considerations and calculates the weighted vector. This vector's angle must be below the defined constraint (default: 85.00°).

Hydrophobic features are depicted with distance constraints. There are two thresholds that change the effect of hydrophobicity: minimum threshold and surface accessibility threshold. Minimum threshold analyzes the hydrophobic moiety and determines the minimum hydrophobic threshold of groups, rings and chains. Less hydrophobic regions can be determined when the minimum threshold value is increased. Surface accessibility threshold determines the sterical accessibility due to the hydrophobic moiety. Increasing the surface accessibility threshold causes more restrict identification.

LigandScout defines aromatic features according to the interaction of the ligand and its environment. Orthogonal, parallel (π - π) and cation- π interactions are considered to determine the aromatic region. To determine excluded volumes, LigandScout determines the shape of the active site and the conformation of the protein. Excluded volumes increase the selectivity of the pharmacophore model.

LigandScout also provides two different alignment options to its users when dealing with more than one pharmacophore model: a “shared” pharmacophore model consists of the pharmacophore features shared by all models, whereas the “merged” pharmacophore combines all pharmacophore features encountered in any model.

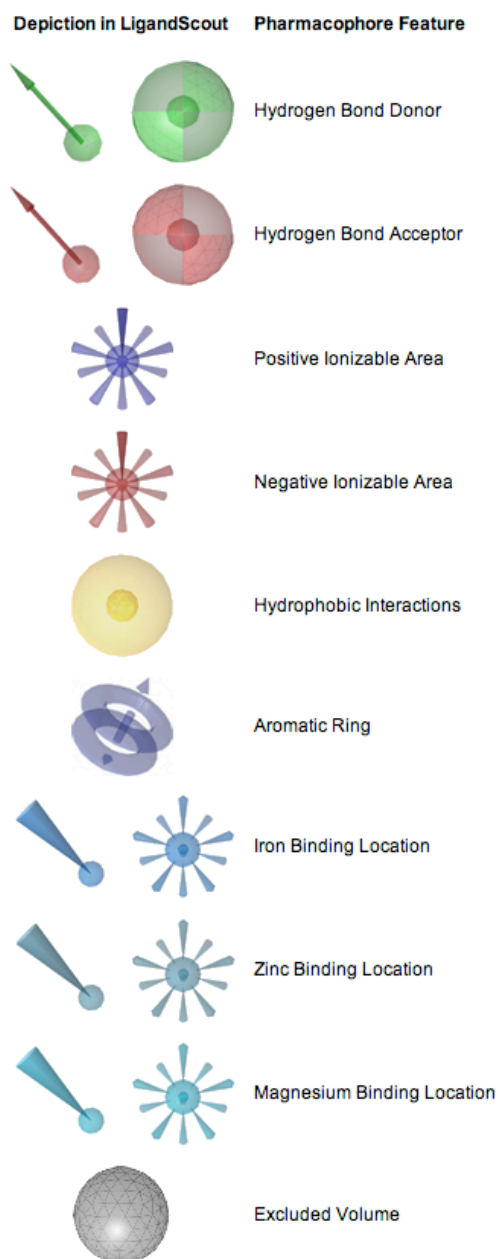


Figure 3.1. Pharmacophore representation in LigandScout. Picture taken from inteligand website.

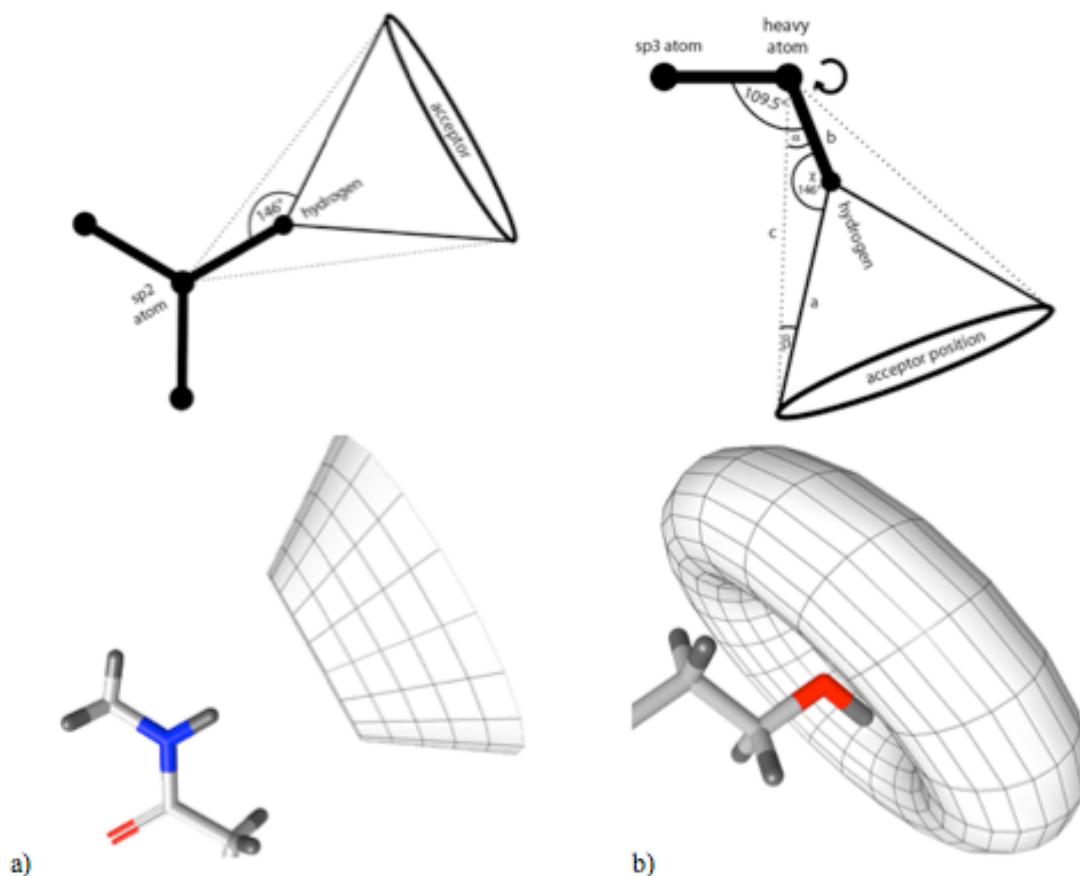


Figure 3.2. Hydrogen bond constraints. (a) Rigid H-bonds constraint of an sp^2 hybridized amide nitrogen. (b) Flexible hydrogen bonding constraint of an sp^3 hybridized hydroxy group. Pictures are taken from manual of LigandScout.

3.2 Virtual Screening

Discovering a new drug is a long term and expensive process. To make this process faster and cheaper, molecules that do not have the desired activities must be eliminated in early stages. For that purpose, screening methodologies have been developed. Ligands with desired activities can be identified experimentally, by high-throughput screening (HTS) of databases including millions of compounds. However, this is a time-consuming and expensive process. Therefore, as an alternative to HTS, virtual screening (VS) became a useful tool for drug discovery [28]. The aim of VS is to determine proper molecules with required properties by screening massive chemical databases in a short period of time at low costs [29]. Generally, VS techniques can be divide into two

groups: Pharmacophore-based virtual screening and Docking-based virtual screening, which are described in detail in the following two sections.

3.2.1 Pharmacophore-based Virtual Screening (PBVS)

Pharmacophore-based screening is useful when the three-dimensional structure of the target protein is not known. Moreover, if 3D structure of a target is known, both docking-based and pharmacophore-based virtual screening methods can be used [30]. A screened molecule is counted as a hit if it contains the required pharmacophore features of the proposed model. Pharmacophore-based screenings are around ten thousand times faster than docking-based screenings, thus they are usually preferred over docking as an initial filter to remove the molecules which do not have essential features for binding [31]. Docking can be used in later stages for a more detailed evaluation of the hit compounds.

In this thesis, the ZINC database was screened using a pharmacophore-based virtual screening tool provided by LigandScout. The program accepts the proprietary LDB file format for molecule libraries. More than one database can be screened simultaneously against the selected pharmacophore. Also, it provides the users to mark their databases as active or inactive to the activity of molecules in the database. In addition, users can customize several screening settings [29]. For instance, LigandScout has two types of scoring functions options. One of them is Pharmacophore-Fit (PF) scoring function. It considers just pharmacophoric features and their RMSD values. The other one is Relative Pharmacophore Fit (RPF) scoring function. RPF scores the amount of the matching pharmacophoric features and normalized the aligned pharmacophores to [0..1]. Moreover, maximum number of features that can be omitted during the screening process, can be specified. If maximum number of omitted features is defined as zero than all pharmacophoric features should be matched to be a valid hit. Check exclusion volume option enables the consideration of exclusion volumes during the screening process.

3.2.2 Docking-based Virtual Screening (DBVS)

Nowadays, docking-based virtual screening techniques are widely used in drug discovery and many important drugs on the market are discovered *in silico* [29]. DBVS generates all possible poses of the small molecule to find out the best configuration to form a protein-ligand complex [32]. Hence, two characteristics of the docking tool are critical: configuration search algorithm and scoring function (SF). The search algorithm predicts all possible binding modes of the ligand in the receptor's binding pocket. SF ranks them according to the score value and accepts the conformation with the highest score as the most plausible binding mode [33]. At the end of screening, all docked ligand molecules are sorted based on their score values, and the ligand molecules at the top of the list are selected as hit molecules. The top of the list can be defined either as a small percentage of the whole database or using a threshold score value.

3.3 Docking methodology and Scoring Functions

Currently, many software tools exist for docking. Every tool has its own pose generation methods and scoring functions [28]. Both AutoDock and GOLD software tool used in this thesis, use genetic algorithm to search the conformational space. In genetic algorithm, the ligand is represented by a set of torsion angles of rotatable bonds resembling to the set of genes in a chromosome. As shown in Figure 3.3, genetic algorithm starts by creation of a population, composed of randomly generated ligand conformations, the so-called "parents" [33]. Then, these parent conformations reproduce new conformations by transmitting genetic information between each other by crossover and mutation, e.g. by exchanging randomly their torsion angles as shown in Figure 3.3. As a result, children ligand poses are generated. Next, the energies of these new conformations were evaluated. The ones with the lowest energies (or highest scores) were selected as members of the next generation. This iteration continues until the best conformation with the highest score is obtained.

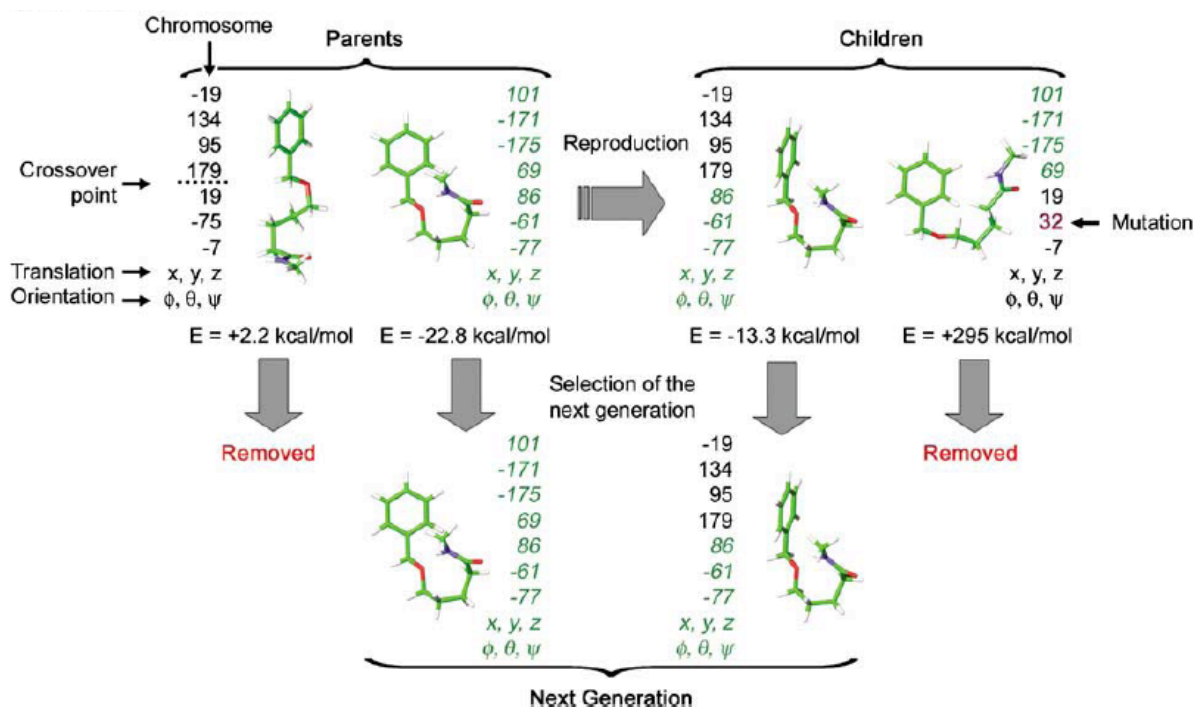


Figure 3.3 Schematic representation of the genetic algorithm.

Scoring functions are the important part of a docking tool because the binding poses that are generated from DBVS must be evaluated to recognize the correct pose which would resemble the native binding conformation. There are three principles of scoring [28]: (i) to recognize native binding mode, (ii) to estimate binding affinities correctly, and (iii) to recognize active compounds from poor binders. Current SF's are divided into three groups as force-field based (FF), empirical and knowledge-based [34].

Force-field scoring functions are based on different force field parameter sets, such as CHARMM [35] and AMBER [36]. The interaction between a ligand and a receptor is most often described by sum of van der Waals (VDW), hydrogen bond and electrostatic terms, which are defined with 12-6 Lennard-Jones potential, 12-10 Lennard-Jones potential and Coulombic formulation, respectively. For example, AutoDock is based on the AMBER force field [37] as given in equation 1.

$$\begin{aligned}
\Delta G = & \\
& \Delta G_{vdW} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \\
& \Delta G_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} + E_{hbond} \right) + \Delta G_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + \Delta G_{tor} N_{tor} + \\
& \Delta G_{sol} \sum_{i,j} S_i V_j e^{\left(-\frac{r_{ij}^2}{2\sigma^2}\right)} \tag{3.1}
\end{aligned}$$

The first term of this equation is the 12-6 Lennard-Jones potential. r_{ij} is the distance between protein atom i and ligand atom j , A_{ij} and B_{ij} are the VDW parameters. The second term is the 12-10 Lennard-Jones potential. $E(t)$ function provides directionality based on the angle t from ideal hydrogen bonding geometry. The third term is the Coulomb formulation for electrostatics. $\varepsilon(r_{ij})$ stands for distance-dependent dielectric constant, q_i and q_j are the atomic charges. Final term is a desolvation potential. V stands for the volume of the atoms and S represents the solvation parameter.

GoldScore is also a force-field based scoring function and consists of four components namely, protein-ligand hydrogen bond energy and van der Waals energy, ligand internal van der Waals energy and ligand torsional energy. GoldScore scoring function is defined as:

$$f = S_{hb_ext} + S_{VDW_ext} + S_{hb_int} + S_{VDW_int} \tag{3.2}$$

where, S_{hb_ext} is the protein-ligand hydrogen bonding score, and S_{hb_int} is the internal hydrogen bonding of the ligand. S_{VDW_ext} and S_{VDW_int} are the van der Waals scores.

Empirical scoring functions are derived using the experimentally determined binding affinities and structural information. They are the sum of uncorrelated terms with coefficients obtained from regression analysis. X-Score, GlideScore and ChemScore are the examples. ChemScore scoring function defined as [38]:

$$\tag{3.3}$$

$$\Delta G_{binding} = \Delta G_0 + \Delta G_{\text{bond}} + \Delta G_{\text{metal}} + \Delta G_{\text{lipo}} + \Delta G_{\text{rot}}$$

Each component of this equation is the product of a term dependent on the magnitude of a particular physical contribution to free energy and regression coefficients. According to this information, those components can be written as:

$$\Delta G_{binding} = v_0 + v_1 P_{\text{bond}} + v_2 P_{\text{metal}} + v_3 P_{\text{lipo}} + v_4 P_{\text{rot}} \quad (3.4)$$

where, v terms represent regression coefficients and P terms are various types of physical contributions to binding.

ChemScore value is obtained by adding in a clash penalty and internal torsion terms. Covalent and constraint scores may also be added.

$$\text{ChemScore} = \Delta G_{binding} + P_{\text{clash}} + c_{\text{internal}} P_{\text{internal}} + (c_{\text{covalent}} P_{\text{covalent}} + P_{\text{constraint}}) \quad (3.5)$$

Knowledge-based scoring functions's parameters are derived from experimentally determined X-ray structures of protein-ligand complexes. They are calculated using Boltzmann distribution law, as:

$$w(r) = -k_B T \ln \left[\frac{\rho(r)}{\rho^*(r)} \right] \quad (3.6)$$

where $w(r)$ represents the all potential parameters, k_B is the Boltzmann constant, T is the absolute temperature of the system, $\rho(r)$ is the number density of the protein-ligand atom pair at distance r in the training set and $\rho^*(r)$ is the pair density in a reference state where the interatomic interaction are zero. DrugScore and GOLD/ASP are examples of knowledge-based SFs [33,39].

3.4 Receiver Operating Characteristic (ROC) Curve

Pharmacophore-based virtual screening (PBVS) is a widely used method in drug discovery. A pharmacophore model that is generated for PBVS should be able to enrich as many active compounds as possible for selected database. There are several methods that ensure this enrichment. Receiver Operating Characteristic (ROC) curve is one of these methods [29]. ROC curves can be easily analyzed visually. It is used for selecting appropriate threshold values, which gives the maximum ratio of active molecules to inactive molecules. [40]. ROC curve graph is derived from sensitivity (Se) value on the y -axis and 1-specificity (Sp) value on the x -axis. Se is the ratio of the number of selected active molecules and the number of all active molecules in the database. In other words, Se gives the rate of true positives and is calculated as [29]:

$$Se = \frac{\text{Selected actives}}{\text{All actives}} = \frac{TP}{TP + FN} \quad (3.7)$$

where TP is true positives, FN is false negatives. Sp is the ratio of inactive molecules not selected by the model and all inactive molecules in the database. It defines the capability of pharmacophore model to eliminate inactive molecules. Sp is defined as [29]:

$$Sp = \frac{\text{Discarded inactives}}{\text{All inactives}} = \frac{TN}{TN + FP} \quad (3.8)$$

where TN is true negatives, FP is false positives. Accordingly, $1-Sp$ becomes [29]:

$$1 - Sp = \frac{FP}{TN + FP} \quad (3.9)$$

ROC curve represents all possible Se and $1-Sp$ values of a molecule obtained from PBVS, using different thresholds, such as a scoring function defining the goodness of a hit molecule. ROC curve starts from the origin. The diagonal line represents the case where the pharmacophore model cannot make a distinction and molecules are selected

randomly. If ROC curve is located above the diagonal line, it represents a pharmacophore model that enriches more active compounds than inactive compounds. If the ROC curve is below the diagonal line, it indicates simply that the pharmacophore model's ability to distinguish actives from inactives is worse than random (Figure 3.4) [29].

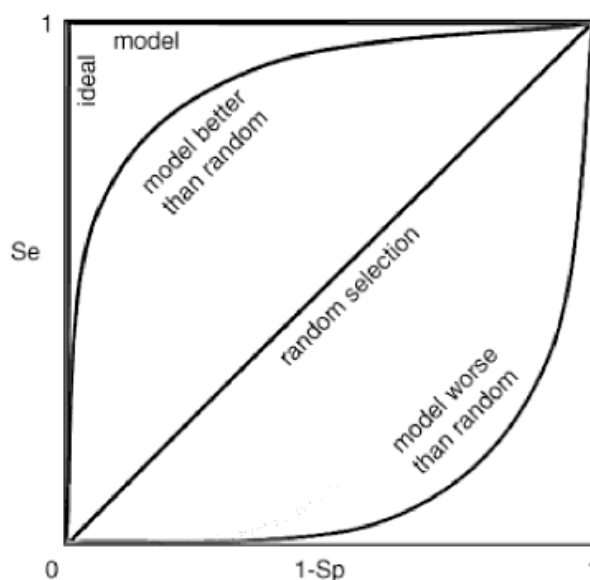


Figure 3.4 ROC curve model.

3.5 ZINC Database

Since structure-based drug design has become a powerful technique for drug discovery, the need for suitable databases has increased. These databases have to be easily accessible and they must be comprised of proper data. Lots of databases were generated to supply these needs, such as ChemNavigator [41], Ligand.Info [42], The ChemBank project [43]. However, these databases do not meet the exact requirements. For instance, the ChemBank contains only 2D molecules. Availability of 3D structures of each molecule that are going to be screened is very important for structure based virtual screening [16].

In 2005, John Irwin and Brian Shoichet introduces ZINC database [16], which includes 3D structure of each molecule. Also, some physical properties such as molecular weight, predicted logP, number of hydrogen bond donors, number of hydrogen acceptors, number of rotatable bonds, net charge, number of rigid fragments, number of chiral centers, number of chiral double bonds and a polar desolvation energy value. ZINC is a freely accessible database. Multiple file formats are available such as SMILE, mol2, 3D SDF and DOCK feasible format. Users can download the entire ZINC database or they can download a set of molecules that satisfy certain physical properties listed above, based on user's research interest. Moreover, they can upload their own molecules to ZINC database as well.

3.6 Rescoring

Docking programs generate a score value for every pose according to their scoring functions and rank them accordingly. Yet, these scores may not be strong enough to make good predictions in all cases. This is the main reason for the use of rescoring functions. These functions simply rescore all the poses obtained from a docking program and reevaluated them. These scoring functions try to enhance predictive power of docking tools [44]. In this thesis, DSX (Drug Score eXtended), which is a knowledge-based scoring function, is used for rescoring the poses obtained from AutoDock. The pair potentials are derived from protein-ligand complexes as stored in the PDB DataBank.

Chapter 4

Results and Discussions

4.1 Generation of the Pharmacophore model

Five crystal structures of β_2 -AR in the inactive state were used for the generation of a pharmacophore model via LigandScout software tool [27]. Their PDB entries are 2RH1 (inverse agonist: carazolol), 3D4S (inverse agonist: timolol), 3NY8 (inverse agonist: ICI118,551), 3NY9 (inverse agonist: JSZ), and 3NYA (antagonist: Alprenolol). LigandScout determines the pharmacophoric features according to some distance constraints between the macromolecule and its ligand. These constraints were listed in Table 3.1 in previous chapter.

First, a pharmacophore model was created for each complex structure as shown in Figure 4.1, and listed in Table 4.1. 2RH1 has a total of nine pharmacophoric features: three hydrogen bond (H-bond) donors, two H-bond acceptors, three hydrophobic regions, and one positive ionizable area (Figure 4.1a). 3D4S consists of three H-bond donors, three H-bond acceptors, three hydrophobic areas, and one positive ionizable area (Figure 4.1b). 3NY8 includes one H-bond donor, two H-bond acceptors, four hydrophobic interactions, and one positive ionizable area (Figure 4.1c). 3NY9 has three H-bond donors, two H-bond acceptors, four hydrophobic interactions, and one positive ionizable area (Figure 4.1d). 3NYA consists of ten pharmacophoric regions: three H-bond donors, three H-bond acceptors, three hydrophobic interactions and one positive ionizable area (Figure 4.1e).

Next, a shared pharmacophore model was obtained after aligning all five models via LigandScout. The shared pharmacophore model consists of five pharmacophoric features that are common to all five pharmacophore models (Figure 4.2). As listed in Table 4.1, these features include one H-bond donor, one H-bond acceptor, two hydrophobic interactions, and one positive ionizable area. Figure 4.2 shows the model with and without excluded volumes.

Table 4.1 Number of pharmacophore features in five inactive crystal structures and the shared pharmacophore model.

	Hydrogen Bond Donor	Hydrogen Bond Acceptor	Hydrophobic Interactions	Positive Ionizable Area	Total
2RH1	3	2	3	1	9
3D4S	3	3	3	1	10
3NY8	1	2	4	1	8
3NY9	3	2	4	1	10
3NYA	3	3	3	1	10
Shared	1	1	2	1	5

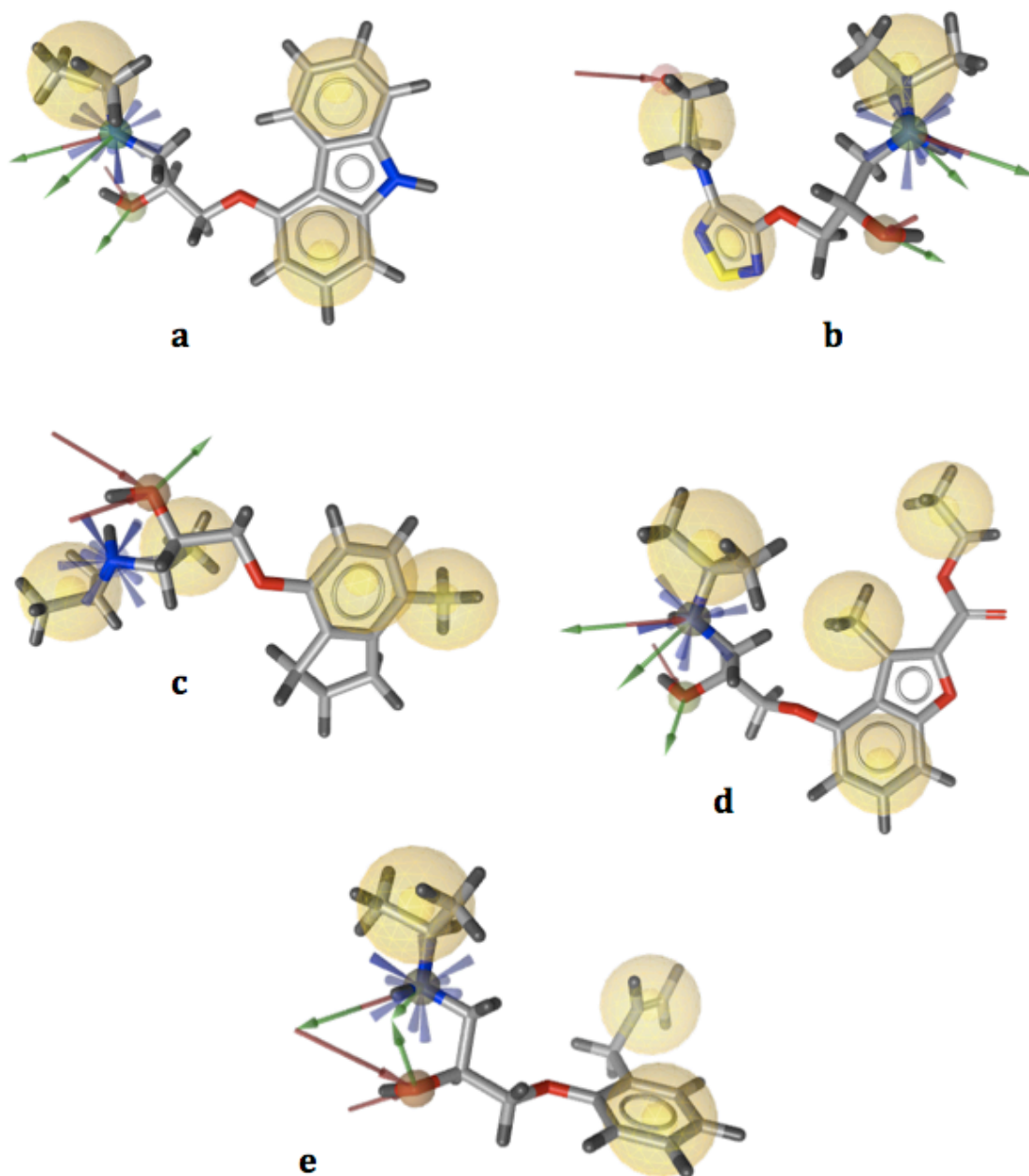


Figure 4.1 Pharmacophore models of five X-ray crystal structures of human- β_2 AR with PDB ids (a) 2RH1, (b) 3D4S, (c) 3NY8, (d) 3NY9, (e) 3NYA (a-e). Pharmacophore models generated via LigandScout. Yellow and blue spheres represent hydrophobic areas, and positive ionizable areas, respectively. Green and red arrows represent H-bond donors and H-bond acceptors, respectively.

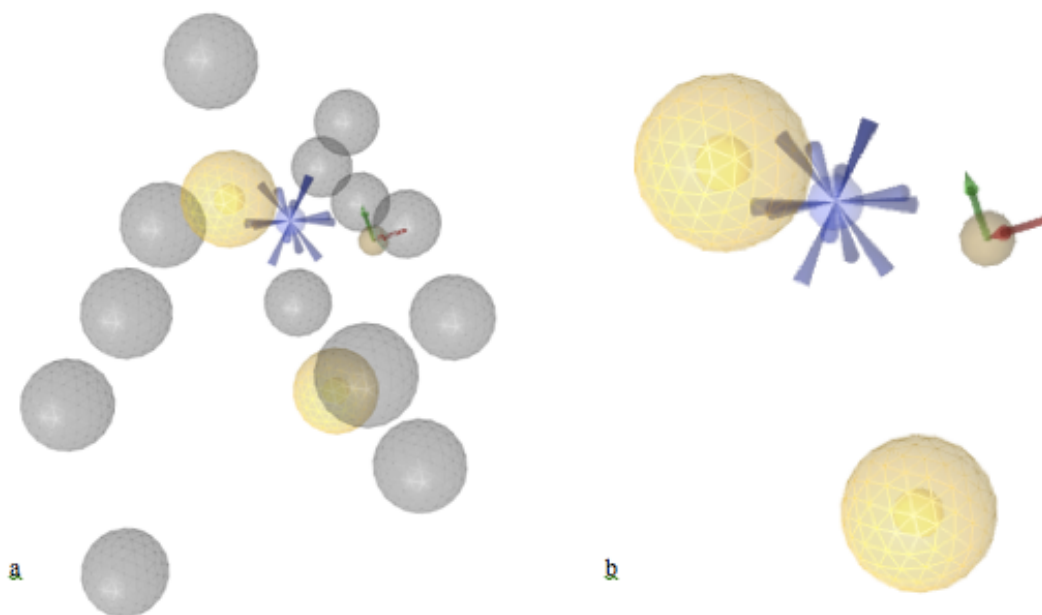


Figure 4.2. Pharmacophore features of the shared pharmacophore model (a) with and (b) without excluded volume. See the caption of Figure 4.1 for more explanation about illustration.

4.2 ZINC Database

ZINC Database includes 19,607,982 purchasable molecules as of January 2013. According to some physical properties, these molecules were categorized into three subsets: Lead-Like, Fragment-Like and Drug-Like. The Lead-like subset includes molecules with molecular weight between 250 and 350 daltons, logP value less than or equal to 3.5, and the number of rotatable bonds less than or equal to 7. The Fragment-Like subset consists of molecules having molecular weight 250 or less, calculated logP less than or equal to 3.5 and number of rotatable bonds less than or equal to 5. The Clean Drug-Like subset, selected for screening process of our study in this thesis, consists of 9,928,465 molecules and complies with the Lipinski's Rule of five. In addition, other desirable physical properties for drug likeliness such as partition coefficient (xlogP) and number of rotatable bonds are also observed in these molecules, as listed in Table 4.2.

Table 4.2. Physical Properties of Clean Drug-Like subset of ZINC Database.

Molecular Weight (g/mol)	$150 \leq \text{mwt} \leq 500$
xlogP	$\text{xlogP} \leq 5$
Number of Rotatable Bonds	$\text{Rb} \leq 7$
Polar Surface Area (\AA^2)	$\text{tPSA} < 150$
Number of H-bond Donors	$\text{Hbd} \leq 5$
Number of H-bond Acceptors	$\text{Hba} \leq 10$

4.3 Set of Compounds with Known Activities

A decoy set was created to validate the reliability of the shared pharmacophore model using GLIDA GPCR-Ligand Database [45]. There exist many databases related to GPCRs, such as GPCRDB [46], IUPHAR [47], which comprise data about biological aspects of proteins [17]. However, GLIDA is different from other databases, in the fact that it includes novel and GPCR oriented molecules. It focuses on the importance of ligands as drug leads. GLIDA provides three kinds of data: biological information about GPCRs, the chemical information about their ligands, and, the binding information of specific GPCR ligands. All that data about these GPCRs was obtained from three species: human, rat and mouse since these species have an effective role in drug discovery process. In addition, GLIDA provides a user-friendly interface [48], over which users can search a molecule according to a ligand or a GPCR protein. It demonstrates the ligand structures and lists the correlation maps that show the relation between the structure of the protein and the property of the ligand. Moreover, users can acquire information on the activity types of the ligand.

GLIDA provides a total 317 agonists and 259 antagonists for human β_2 -adrenergic receptor. During the creation of our decoy set, the structural diversity was taken into account. Thus, compounds that have similar structures were eliminated. The final decoy set consisted of a total 117 distinct molecules of which 53 were antagonists and of which 64 were agonists. Corina [49] program was used to transform these molecules into 3D mol2 format. These molecules were depicted in Appendix A and Appendix B.

4.4 Pharmacophore-Based Screening of the decoy set and the Clean Drug-Like Subset of ZINC Database

In order to validate how well the shared pharmacophore model discriminates the antagonists from agonists or non-inhibitors, the decoy set was first screened using the shared pharmacophore model using LigandScout software tool. LigandScout presents two different scoring functions: Pharmacophore Fit and Relative Pharmacophore Fit. Furthermore, the maximum number of omitted features, which signifies the maximum number of query features that can be left out during screening can be adjusted. Finally, the exclusion volume can be left checked or unchecked. Exclusion volume represents the sterically occupied regions by the receptor. All of these variables affect the result of a virtual screening. To determine an optimum solution, several runs have been applied to the decoy set using various combinations of these variables.

Initially, a screening was performed using the default values that are given at the third entry of Table 4.3 (Scoring function = Relative Pharmacophore Fit, Maximum Number of Omitted Features = 2, Check exclusion volume = On). Accordingly, molecules were counted as hits, if they had at least three pharmacophoric features out of five and if they fit properly in the area that was defined by exclusion volumes. As a result of this screening, 82 hits were found. 45 of them were antagonists (active), and 37 of them were agonists (inactive). After that, screening was repeated three times using maximum number of omitted features as 0,1, and 3 and the results were listed in Table 4.3. As expected, increasing the number of omitted features increased the number of hits as well, since less number of pharmacophoric features are required to become a hit. On the other hand, the number of false positives (agonists) has increased as well. ROC (Receiver Operating Characteristic) curves for each screening were determined using LigandScout and illustrated in Figure 4.3.

Table 4.3 Screening results using different values for the maximum number of omitted features.

RUN #	Scoring Function	Max. Number of Omitted Features	Check Exclusion Volumes	Number of Antagonists/ Total Antagonists¹	Number of Agonists / Total Agonists²	Total Hits / Total Compounds³
1	RPF ⁴	0	On	19 / 53	1 / 64	20 / 117
2	RPF	1	On	37 / 53	12 / 64	49 / 117
3	RPF	2	On	45 / 53	37 / 64	82 / 117
4	RPF	3	On	45 / 53	38 / 64	83 / 117
5	PF ⁵	0	On	19 / 53	1 / 64	20 / 117
6	PF	1	On	37 / 53	12 / 64	49 / 117
7	PF	2	On	45 / 53	37 / 64	82 / 117
8	PF	3	On	45 / 53	38 / 64	83 / 117

¹ Total number of antagonists in the decoy set was 53.

² Total number of agonists in the decoy set was 64.

³ Total compounds in the decoy set was 117.

⁴ Relative Pharmacophore Fit

⁵ Pharmacophore Fit

Next, the scoring function was changed to Pharmacophore Fit to see the effect on the screening results. Check exclusion volume option was “on” during this run as well. The number of omitted features option was changed similarly. Results were listed in Table 4.3. Same results were obtained with the same ROC curves, which means the scoring function did not have any effect on the screening.

Next, screenings were performed to understand the effect of exclusion volumes. Check exclusion volumes option was set to “off”. First, Relative Pharmacophore fit scoring function was chosen and the maximum omitted features were increased from zero to three (Table 4.4). Then, the scoring function was changed to Pharmacophore Fit scoring and the number of omitted features was changed. Similar to previous results, the scoring function had no effect on the results. However, the number of hits changed with the maximum number of omitted features, as seen previously with excluded volumes. The corresponding ROC curves are illustrated in Figure 4.4 for run numbers 9 through 12.

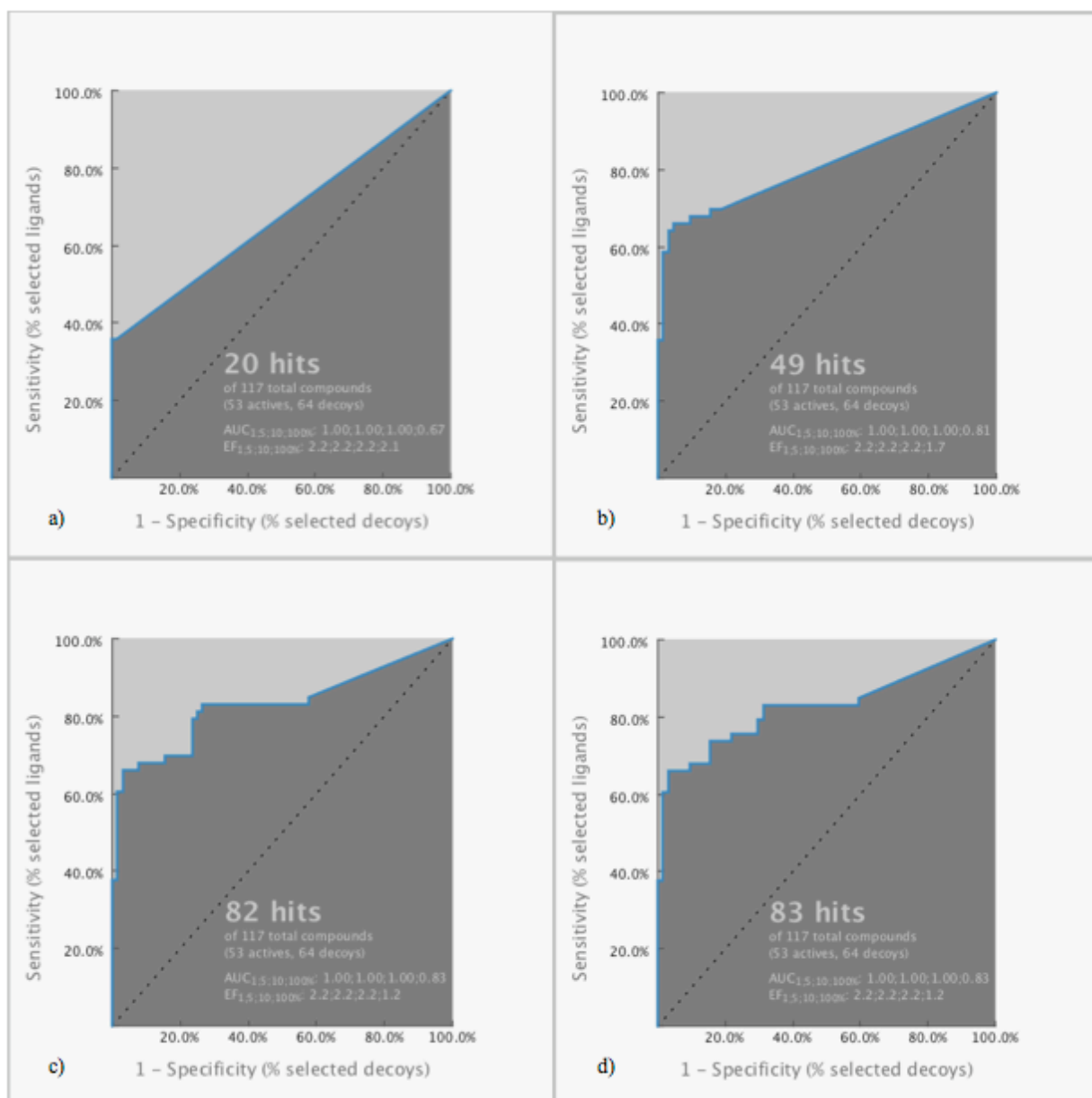


Figure 4.3 ROC curves of the screening results using different values for the maximum number of omitted features. (a) RUN#1. $AUC_{1,5,10,100\%}: 1.00; 1.00; 1.00; 0.67$ and $EF_{1,5,10,100\%}: 2.2; 2.2; 2.2; 2.1$. (b) RUN#2. $AUC_{1,5,10,100\%}: 1.00; 1.00; 1.00; 0.81$ and $EF_{1,5,10,100\%}: 2.2; 2.2; 2.2; 1.7$. (c) RUN#3. $AUC_{1,5,10,100\%}: 1.00; 1.00; 1.00; 0.83$ and $EF_{1,5,10,100\%}: 2.2; 2.2; 2.2; 1.2$. (d) RUN#4. $AUC_{1,5,10,100\%}: 1.00; 1.00; 1.00; 0.83$ and $EF_{1,5,10,100\%}: 2.2; 2.2; 2.2; 1.2$.

Table 4.4 Screening results when check exclusion volumes option was “off”.

RUN #	Scoring Function	Max. Number of Omitted Features	Check Exclusion Volumes	Number of Antagonists/ Total Antagonists¹	Number of Agonists / Total Agonists²	Total Hits / Total Compounds³
9	RPF ⁴	0	Off	24 / 53	3 / 64	27 / 117
10	RPF	1	Off	47 / 53	39 / 64	86 / 117
11	RPF	2	Off	51 / 53	63 / 64	114 / 117
12	RPF	3	Off	51 / 53	63 / 64	114 / 117
13	PF ⁵	0	Off	24 / 53	3 / 64	27 / 117
14	PF	1	Off	47 / 53	39 / 64	86 / 117
15	PF	2	Off	51 / 53	63 / 64	114 / 117
16	PF	3	Off	51 / 53	63 / 64	114 / 117

¹ Total number of antagonists in the decoy set was 53.

² Total number of agonists in the decoy set was 64.

³ Total compounds in the decoy set was 117.

⁴ Relative Pharmacophore Fit

⁵ Pharmacophore Fit

As expected, more hits were obtained when check exclusion volumes option was off. However, the number of agonists (false positives) as hit molecules, has increased as well. Accordingly, more accurate hits were acquired when check exclusion volumes option was on.

LigandScout allows the user to select any excluded volume and discard it. To increase the discriminatory power of the pharmacophore model, two outmost excluded volumes were deleted as shown in Figure 4.5. New screenings were performed with Relative Pharmacophore Fit scoring function and this new “modified” model (Table 4.5). The main goal of this trial was to obtain more active molecules (antagonist) than inactive ones (agonists).

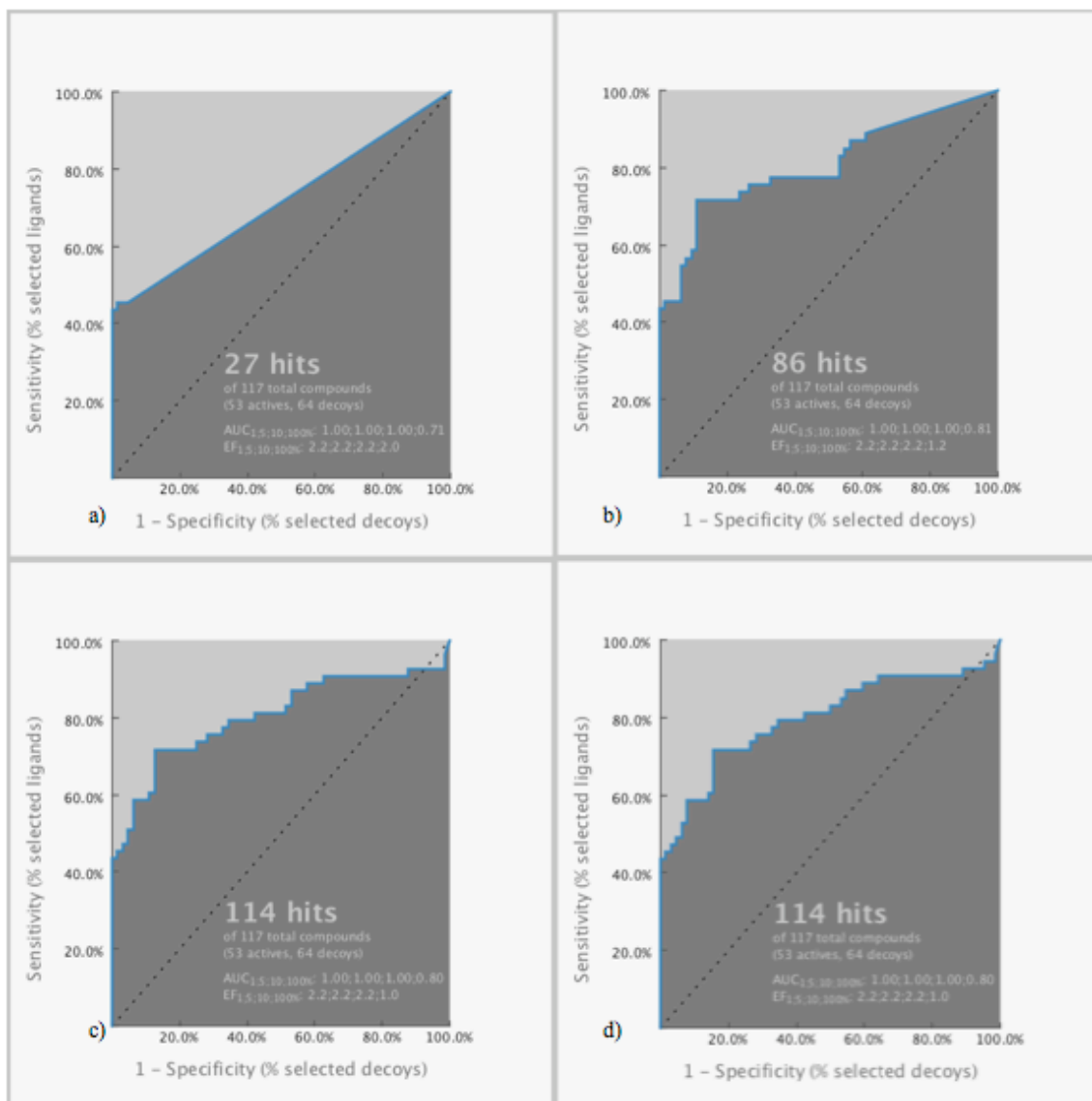


Figure 4.4 ROC curves of the screening results when check exclusion volume option was “off”. (a) RUN#9. AUC_{1,5,10,100%}: 1.00; 1.00; 1.00; 0.71 and EF_{1,5,10,100%}: 2.2; 2.2; 2.2; 2.0. (b) RUN#10. AUC_{1,5,10,100%}: 1.00; 1.00; 1.00; 0.81 and EF_{1,5,10,100%}: 2.2; 2.2; 2.2; 1.2. (c) RUN#11. AUC_{1,5,10,100%}: 1.00; 1.00; 1.00; 0.80 and EF_{1,5,10,100%}: 2.2; 2.2; 2.2; 1.0. (d) RUN#12. AUC_{1,5,10,100%}: 1.00; 1.00; 1.00; 0.80 and EF_{1,5,10,100%}: 2.2; 2.2; 2.2; 1.0.

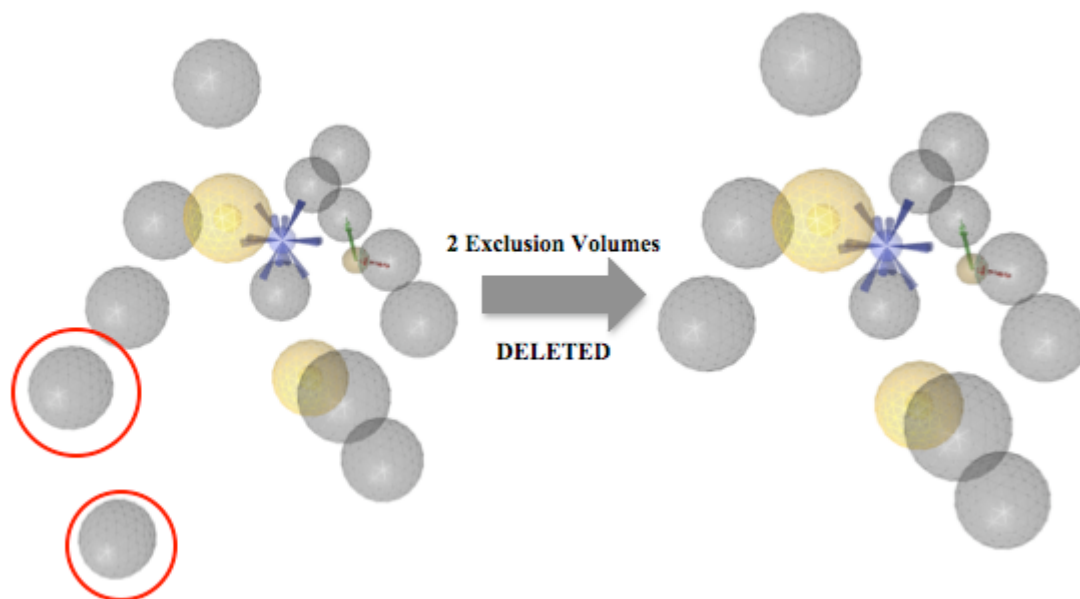


Figure 4.5 Two exclusion volumes, which were circled in red color, were deleted. Picture was taken from LigandScout.

Table 4.5 Screening results of “modified” pharmacophore model using different values for the maximum number of omitted features.

RUN #	Scoring Function	Max. Number of Omitted Features	Check Exclusion Volumes	Number of Antagonists/ Total Antagonists ¹	Number of Agonists / Total Agonists ²	Total Hits / Total Compounds ³
1	RPF ⁴	0	On	19 / 53	1 / 64	20 / 117
2	RPF	1	On	37 / 53	12 / 64	49 / 117
3	RPF	2	On	45 / 53	37 / 64	82 / 117
4	RPF	3	On	46 / 53	38 / 64	83 / 117

¹ Total number of antagonists in the decoy set was 53.

² Total number of agonists in the decoy set was 64.

³ Total compounds in the decoy set was 117.

⁴ Relative Pharmacophore Fit

⁵ Pharmacophore Fit

According to the results in Table 4.5, deleting two exclusion volumes did not affect the screening process. Equal number of hits was obtained from both original pharmacophore model and the modified one. After careful evaluations of ROC curves, ROC curves with the highest AUC (Area Under the ROC Curve) values of 0.83 were obtained using the Relative Pharmacophore Fit score when check exclusion volume option was on and the maximum number of omitted features was either 2 or 3 (RUN#3 and #4 in Table 4.3). As the third run provides less inactive compounds (37 agonists) than the fourth run (38 agonists) with the same number of active compounds (45 antagonists), the set of parameters in RUN #3 (Relative Pharmacophore Fit scoring function, # of omitted features: 2, check exclusion

volume: on) has been selected for further screening of the ZINC database. They were also the default parameters of LigandScout software tool.

Prior to screening ZINC database, a threshold value, which corresponds to a so-called “model exhaustion” or “cut-off” point on the selected model’s ROC curve, must be specified. Hit molecules that have a Relative Pharmacophore Fit score value less than this threshold will be eliminated. This value is determined from the point on the ROC curve where there exists a considerable amount of true positives (antagonists) with respect to false positives (agonists) among the hit molecules. As shown in Figure 4.6 the exhaustion point (red dot on the curve) was determined as (0.27,0.83), which represents the percentage of true positives as 83 (44 antagonists) and the ratio of false positives as 27 (17 agonists). The lowest score value of these 61 (44 + 17) molecules was determined as 0.64, which was taken as the threshold value for the screening of ZINC database.

Using the shared pharmacophore model with default parameters of LigandScout mentioned above a total of 9.928.495 molecules of the Clean Drug-Like subset of ZINC was screened and hit molecules were eliminated if their Relative Pharmacophore Fit score value was under 0.64. Finally, a total of 729.413 molecules with a score value above 0.64 remained and was further evaluated with several docking experiments, alongside the decoy set of 61 molecules. These results are discussed in the next part

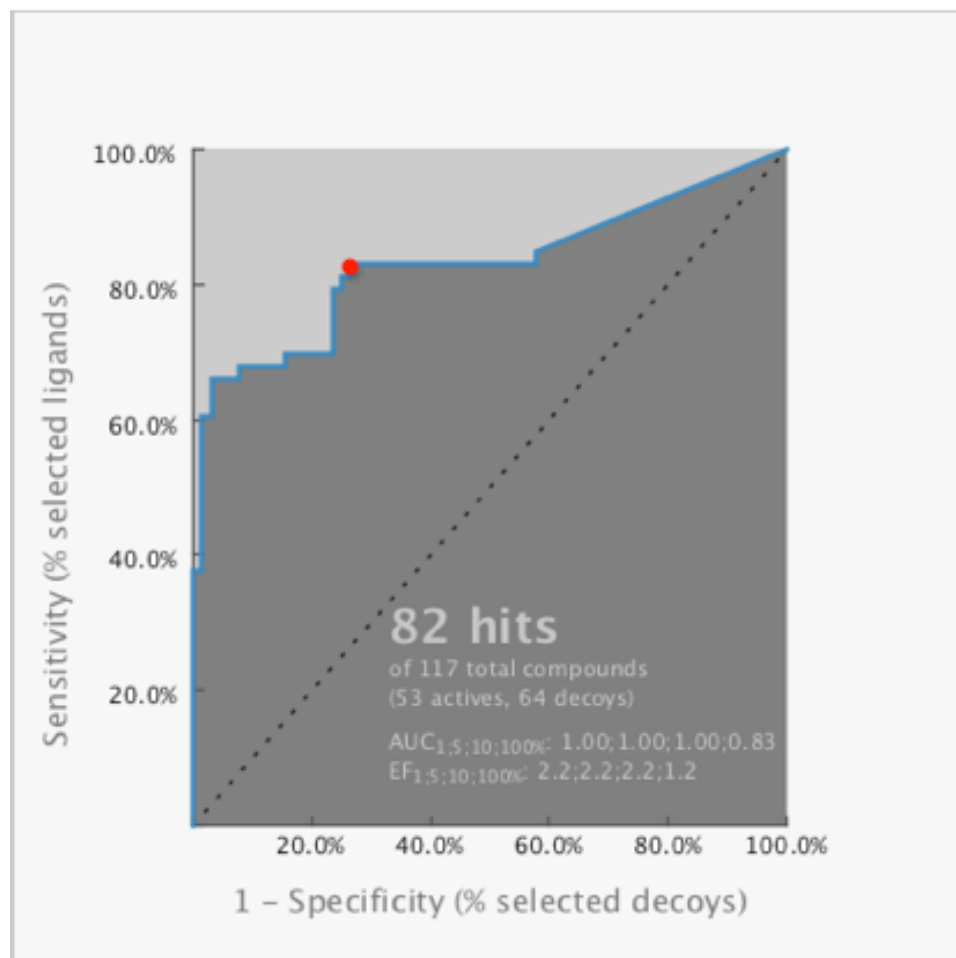


Figure 4.6 ROC Curve for Run #1. Red dot illustrates the model exhaustion point, which corresponds to 83% true positives and 27% false positives.

4.5 Docking and Binding Evaluations

To further evaluate the hit molecules, 61 molecules (44 antagonists and 17 agonists) from decoy set and 729,413 compounds from clean drug-like subset of ZINC database were docked to the inactive crystal structure (PDB id: 2RH1) using GOLD and AutoDock software tools. The crystal structure with PDB id 2RH1 was selected due to its high resolution, which is 2.40 Å. On the other hand, resolutions of structures with PDB ids 3D4S, 3NY8, 3NY9 and 3NYA are reported as 2.80 Å, 2.84 Å, 2.84 Å, and 3.16 Å, respectively.

The first docking was performed using GOLD with CHEMPLP scoring function. Default features were used (radius: 10 Å, ga run: 10, ga search: SLOW). The docked poses with the highest scores were analyzed based on some key residues for antagonist binding grouped into two groups as illustrated in (Figure 4.7). First group

includes Ser203, Ser204 and Ser207 while second group consists of Asp113, Val114 and Asn312. According to the binding test, the docked pose of each compound with the highest score has to be interacting with at least one residue from each group. Otherwise, it is eliminated. A total of 58 molecules (41 antagonists and 17 agonists) from the decoy set have passed the binding test (designed as binding test #1 in Figure 4.8). On the other hand, 610.490 compounds out of 729.413 compounds from ZINC database have passed the binding test. Next, a threshold of 85 for ChemPLP score value was selected for further elimination. 23.568 ZINC compounds, 17 antagonists and 3 agonists have been selected as they had score values above 85. These results correspond to an enrichment factor of 11.9 for 3.2% database coverage calculated using the equation below:

$$ER = \frac{\left(\frac{TP}{n}\right)}{\left(\frac{A}{N}\right)} \quad (4.1)$$

where TP is the number of true positives (antagonists), which is 17, n is the total number of compounds with a score value above 85 and is equal to 23.588 (=23.568 ZINC compounds + 17 antagonists + 3 agonists), A is the number of active compounds (antagonists) which is 44, and N is the total number of compounds screened which is 729.474 (=729.413 + 44 antagonists + 17 agonists). 3.2% simply corresponds to the $n/N \times 100$.

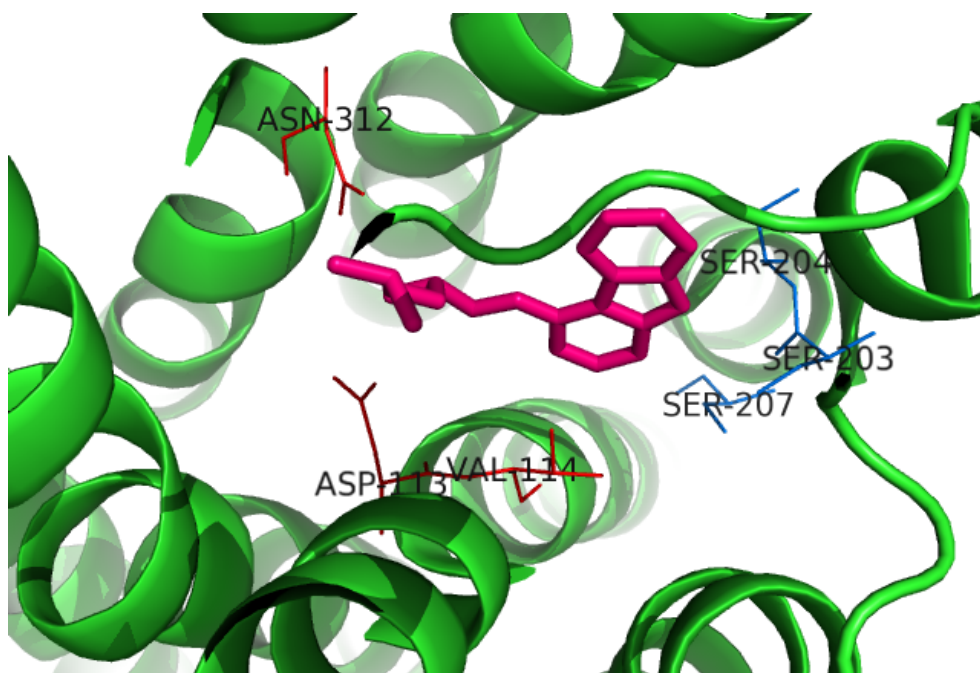


Figure 4.7 Extracellular view of key residues for binding test #1. Blue illustrates the first group (Ser204, Ser203 and Ser 207). Red illustrates the second group (Asn312, Asp113 and Val114). Pink illustrates the partial inverse agonist Carazolol.

The selected 23.568 molecules were redocked to the inactive crystal structure (PDB id: 2RH1) with AutoDock and GOLD. For every molecule, a total of 20 poses were generated from AutoDock and rescored with DSX scoring function. The best pose (with the highest score) was further evaluated based on its binding mode (designed as binding test #2 in Figure 4.8). In addition, each molecule was docked with GOLD using GoldScore and ChemScore as the scoring function. Similarly, the best pose of every molecule from each docking run was further evaluated based on its interacting residues (binding test #2). According to the second binding test, the most essential key residues for antagonist binding were determined and divided into two groups: Ser203, Asp113, Asn312, Tyr316 were found in the first group while Tyr286 and Asn293 formed the second group (Figure 4.9). Molecules have to interact with all the residues from first group and with at least one of the residues from the second group in order to pass the binding test #2. Otherwise, they were eliminated.

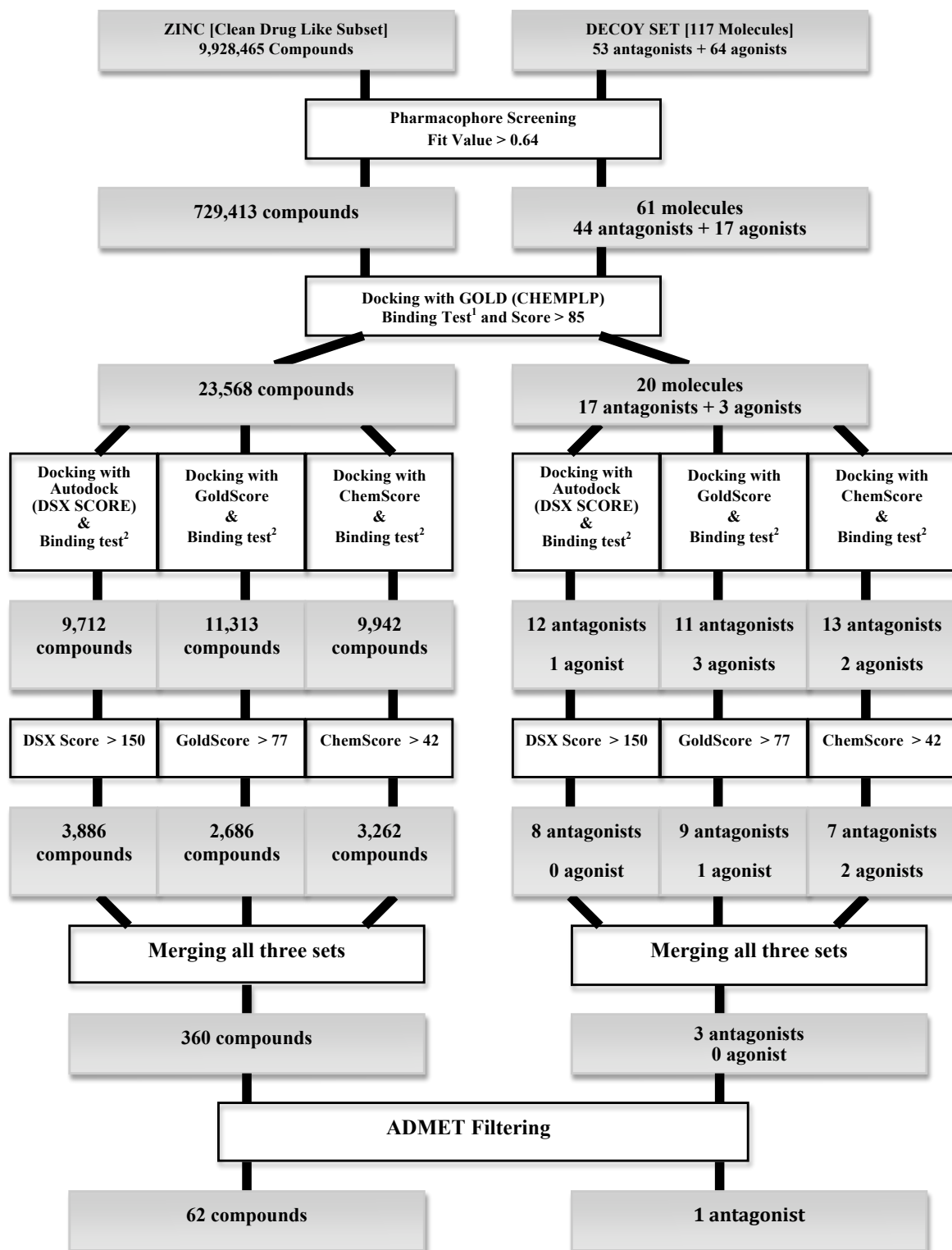


Figure 4.8 Flowchart of the screening process of the Clean Drug-Like ZINC database and the decoy set.

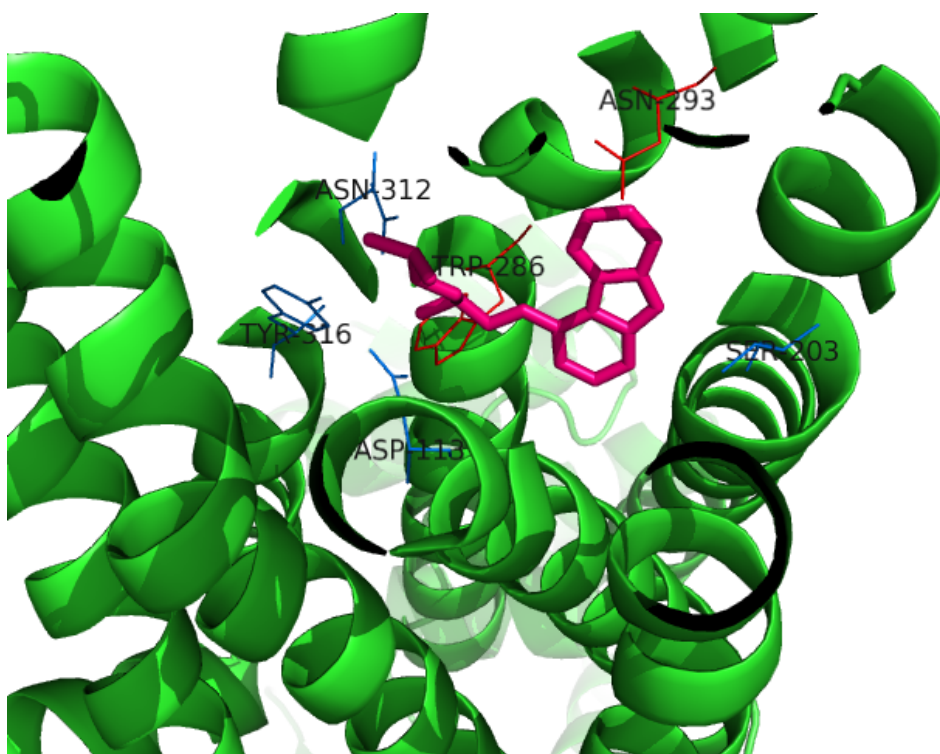


Figure 4.9 Extracellular view of key residues for binding test #2. Red represents the first group, blue represents the second group, and pink represents the partial inverse agonist Carazolol.

Based on these criteria, docking all 23,588 molecules with AutoDock and rescoring with DSX yielded 13 molecules (12 antagonists, 1 agonist) from decoy set and 9,712 compounds from ZINC database that have passed the binding test #2. Similarly, docking with GOLD using GoldScore led to 11 antagonists and 3 agonists from decoy set and 11,313 compounds while docking with GOLD using Chemscore led to 13 antagonists and 2 agonists from decoy set and 9,942 compounds from ZINC database.

For further elimination, threshold values were determined for each scoring function used previously. Among those that have passed the binding test #2, molecules that have a score value above the threshold values have been selected for the next evaluation step. The threshold values were set to 150, 77 and 42 for DSX, GoldScore and ChemScore, respectively. Based on these threshold values, 8 antagonists from decoy set and 3,886 compounds from ZINC database had a DSX score value above 150. On the other hand, 9 antagonists and 1 agonist from decoy set and 2,686 compounds from ZINC database satisfied the threshold value for GoldScore and

finally, 7 antagonists and 2 agonists from decoy set and 3.262 compound from ZINC have been selected as they had a ChemScore over 42.

All these selected molecules were simply merged and a molecule was counted as a hit if it was found in all three groups, i.e., if it satisfied all the three threshold values. Consequently, 360 compounds from ZINC and 3 antagonist molecules from the decoy set satisfied the threshold values and were further evaluated based on their ADMET (Absorption, Distribution, Metabolism, Extraction and Toxicity) properties, using Discovery Studio tool by Accelrys [50].

4.6 ADMET Filtering

ADMET (Absorption, Distribution, Metabolism, Elimination, Toxicity) analysis is a crucial step in drug design. Molecules cannot be counted as drug candidates if they do not fulfill the requirement of ADMET properties. In this thesis, Discovery Studio tool from Accelrys was used for ADMET predictions. ADMET module of Discovery Studio makes predictions based solely on the chemical structure of the molecule. Human intestinal absorption (HIA) and blood brain barrier penetration are the two properties estimated by the program and used for our filtering.

3 antagonist molecules from decoy set and 360 compounds from ZINC database were analyzed based on these features. Figure 4.10 and 4.11 depicts the plot of polar surface area (ADMET_PSA_2D) versus the partition coefficient between n-octanol and water (ADMET_AlogP98) for 3 antagonists from the decoy set and 360 compounds from ZINC, respectively. Since Carvedilol (compound #1, antagonist from the decoy set) falls within all four ellipses in Figure 4.10, it fully satisfies the requirements of intestinal absorption and blood brain penetration. On the other hand, L023760 (compound #2) and L004107 (compound #3) do not fall into any ellipses, thus they were eliminated automatically. At the end of the ADMET filtering of decoy set, only one molecule was remained.

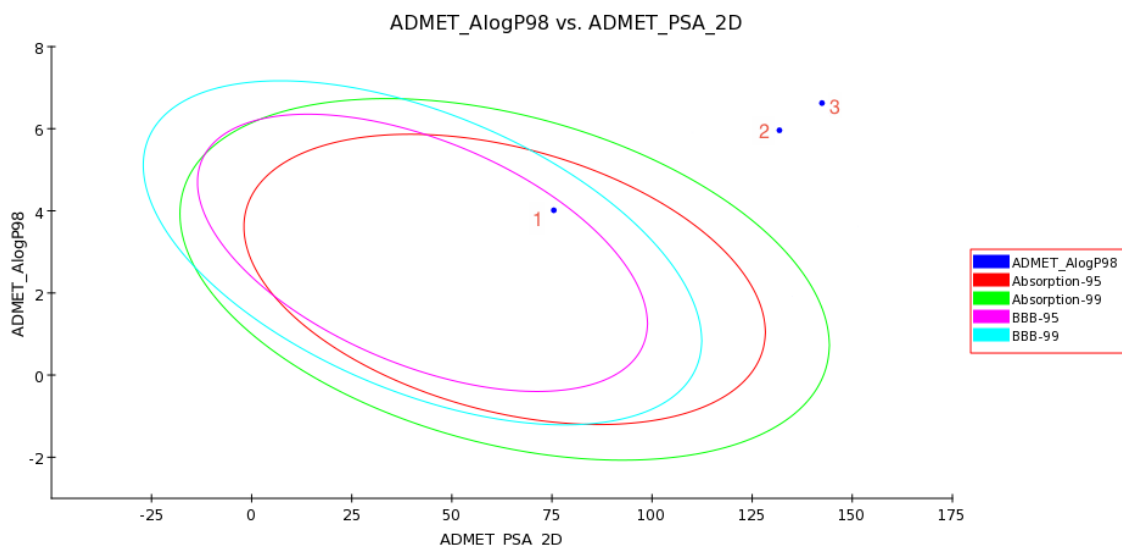


Figure 4.10 ADMET Plot of the decoy set. Carvedilol, L023760 and L004107 are represented with numbers 1, 2, and 3, respectively.

Figure 4.11 shows the ADMET plot of 360 molecules from ZINC database. All 360 molecules satisfied at least one of the required conditions that are blood brain penetration and intestinal absorption, but 62 of them were found inside all four ellipses, thus satisfied all two requirements. Accordingly, these molecules were proposed as possible drug candidates.

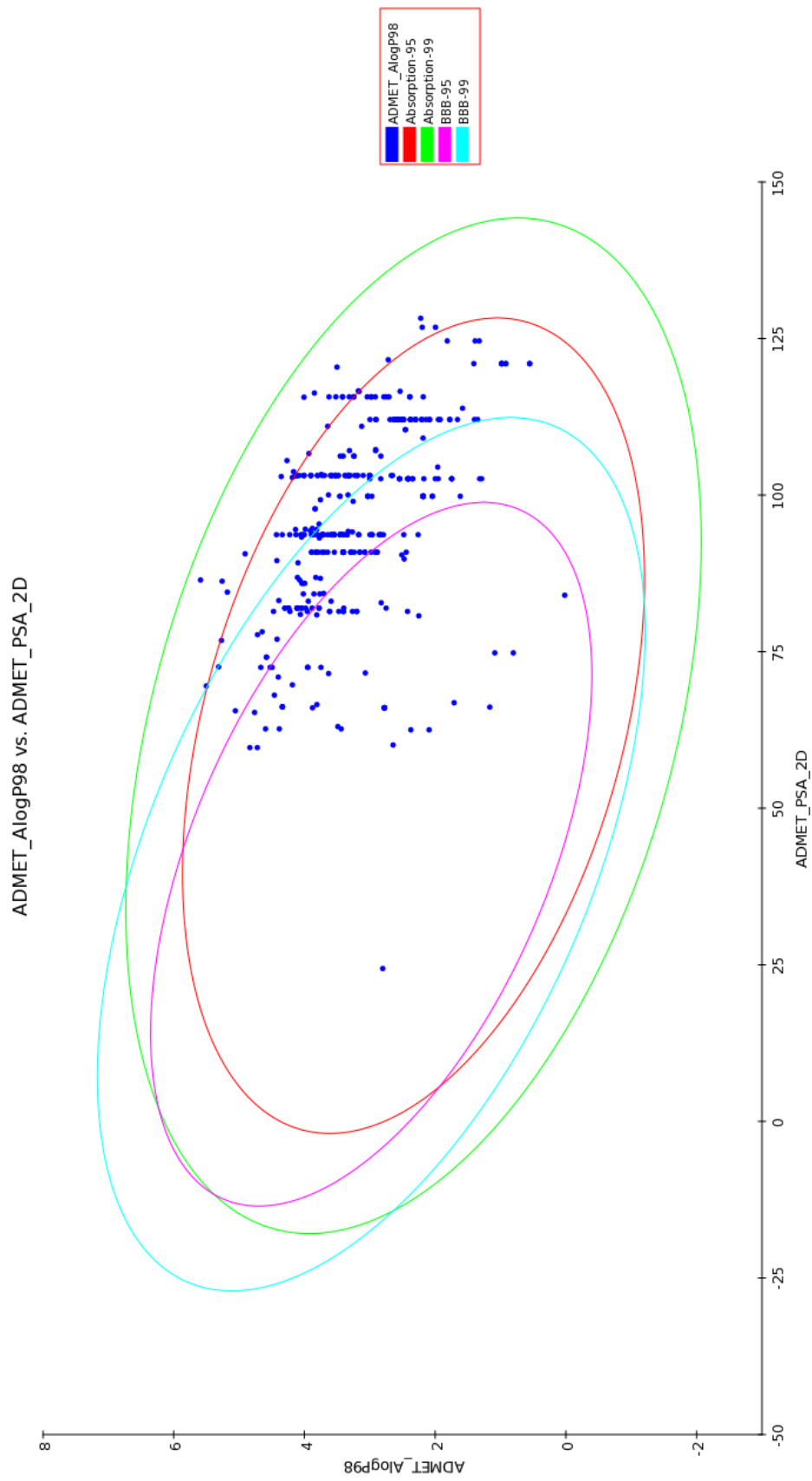


Figure 4.11 ADEMT Plot of ZINC database. 62 molecules out of 360 fall within all four ellipses.

4.7 A Closer Look At The Hit Compounds

As mentioned in the previous section, 62 molecules of ZINC database were proposed as possible drug candidates. Figure 4.12 illustrates these molecules' best poses from AutoDock (rescored with DSX), ChemScore and GoldScore to validate that they bind properly in the binding pocket as the partial inverse agonist Carazolol in the active crystal structure (PDB id: 2RH1).

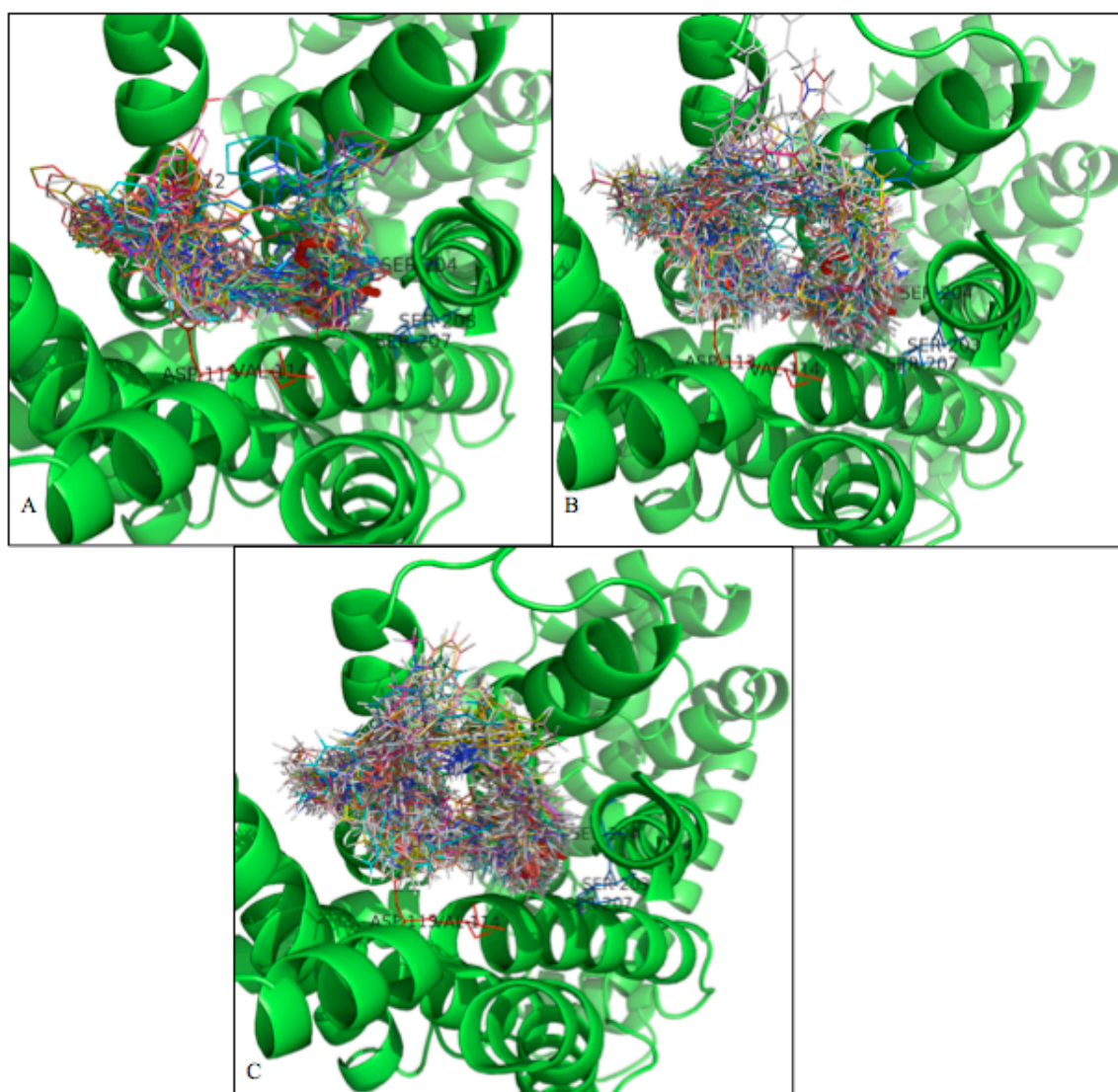


Figure 4.12 Docking poses of 62 molecules. (A) AutoDock- DSX poses. (B) GoldScore poses. (C) ShemScore poses. Carazolol demonstrated in red and stick representation.

After detailed examination of the chemical structures of these 62 molecules, 7 of them were found to be isomers of one of the remaining 55 molecules. Therefore, 55 molecules and its 7 isomers were listed in separately in Appendix C and Appendix

D, respectively. 55 molecules were further classified according to their fixed part or the so-called “scaffolds”. Table 4.6 demonstrates four scaffolds that have been determined so far. 25 out of 55 molecules hold the first scaffold structure, while 10 compounds have scaffold #2, 6 of them have scaffold #3 and other 8 of them scaffold #4. The remaining 6 compounds have unique structures. For each scaffold in Table 4.6, an example hit compound that holds the corresponding scaffold was illustrated. When the scaffolds of the known antagonists are considered, it was found that there was a great similarity between carazolol (See Figure 4.13) except instead of NH, there is a cyclic ring in scaffold #3 and #4.

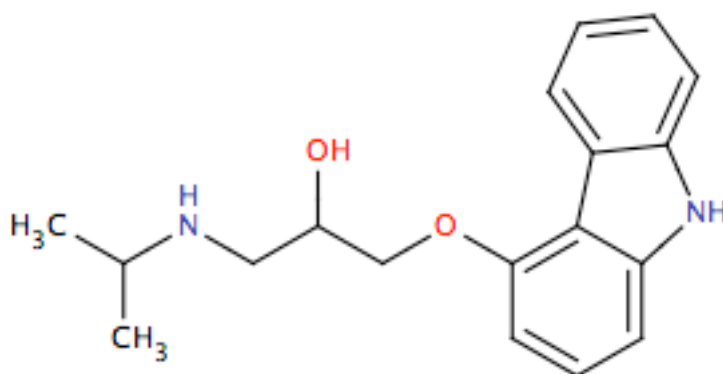
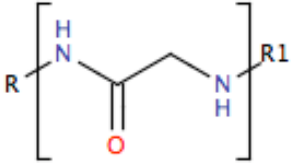
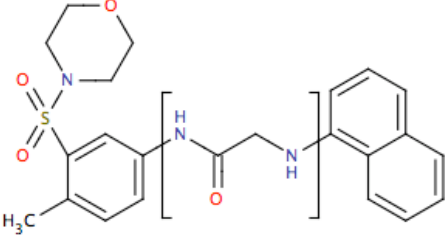
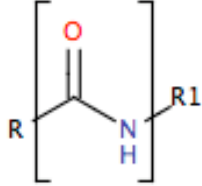
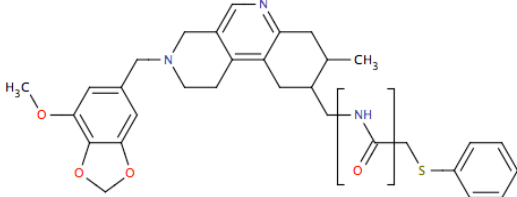
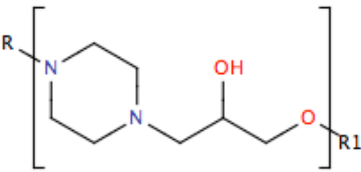
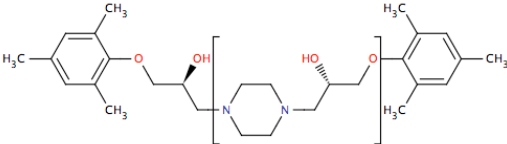
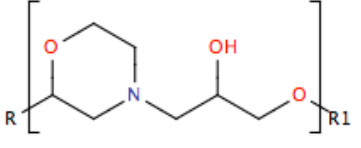
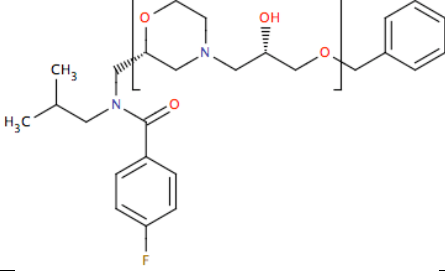
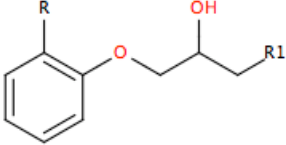
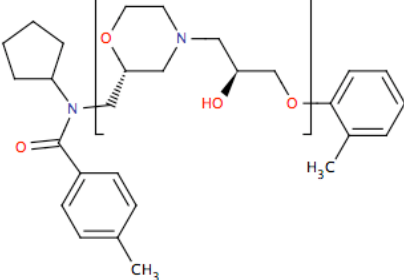


Figure 4.13 2D representation of partial inverse agonist carazolol.

Moreover, Tasler *et al.* [15] proposed a new compound with a strong binding affinity to human s_2AR measured with an experimental K_i value of 1.2 nM (see Figure 4.15). It has been shown to have the alprenolol scaffold as illustrated in Table 4.6. Surprisingly, it also holds the scaffold #4 listed in Table 4.6. Among 55 selected compounds, there exist three compounds with scaffold #4 and that alprenolol scaffold. They are illustrated in Figure 4.14 with database id values of ZINC34691722, ZINC35865918, and ZINC35866100.

Finally, Sabio *et al.* [11] published a study, which proposed two novel compounds illustrated in Figure 4.16. These compounds showed strong binding affinities with experimentally measured K_i values of 0.311 ± 0.09 nM and 57.3 ± 1.6 nM, respectively and amazingly, they hold the scaffold #3 in Table 4.6.

Table 4.6 Scaffold Table

Scaffold #	Scaffold	An example hit compound
1		
2		
3		
4		
Alprenolol Scaffold		

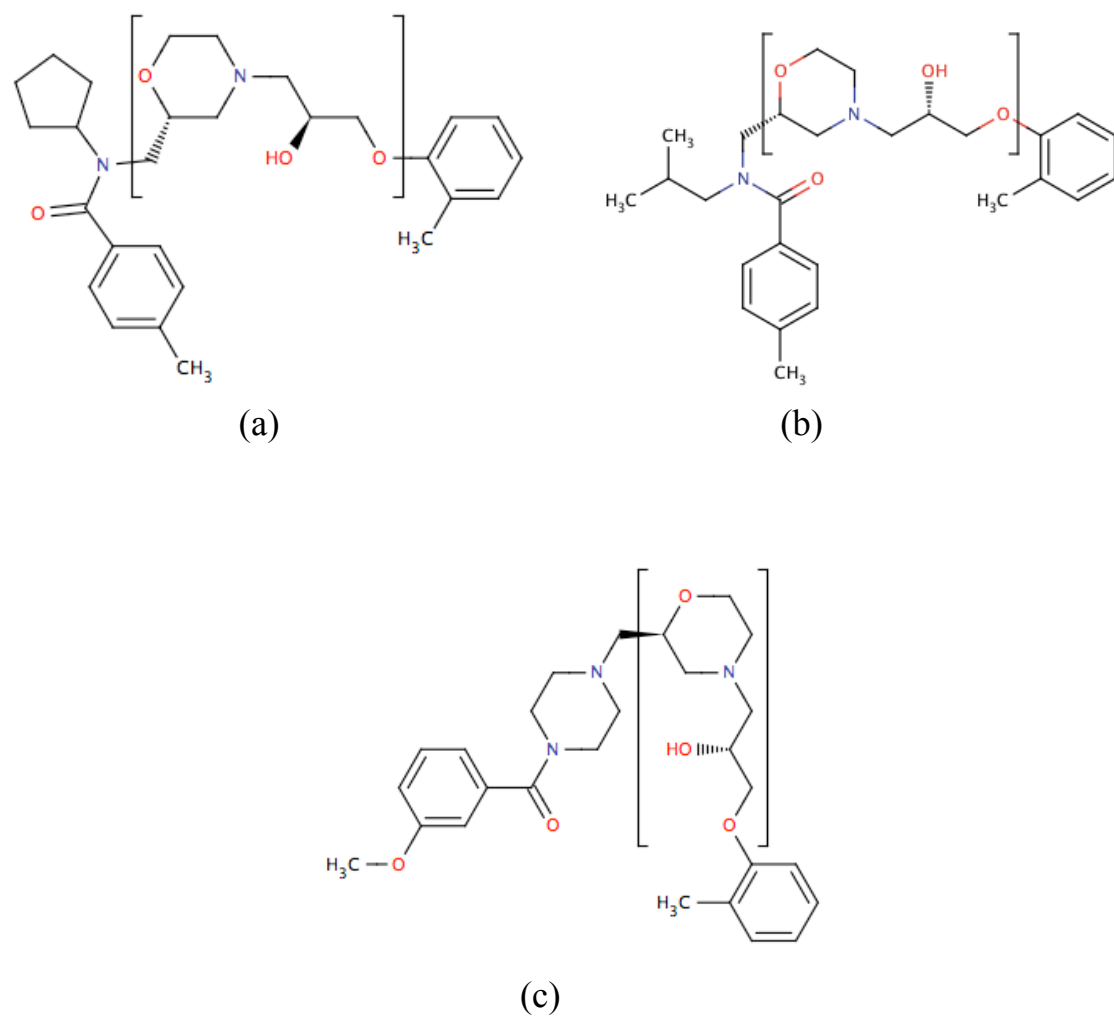


Figure 4.14 These compounds that hold scaffold #4 in Table 4.6 and alprenolol scaffold. (a) ZINC34691722, (b) ZINC35865918, and (c) ZINC35866100

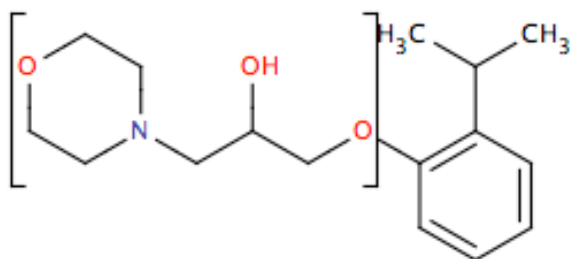
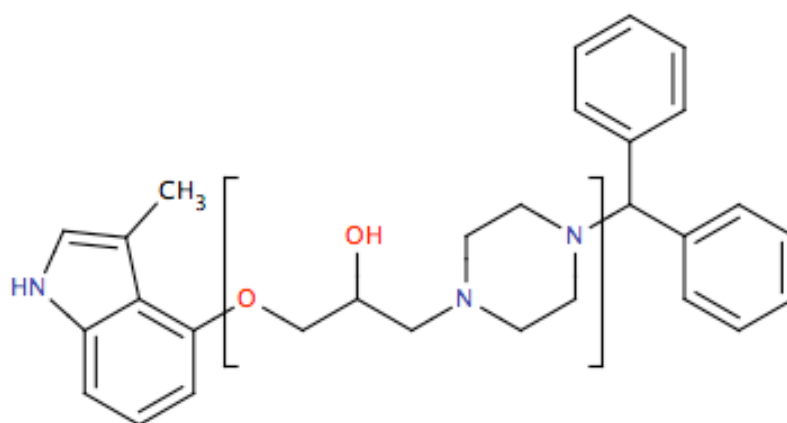
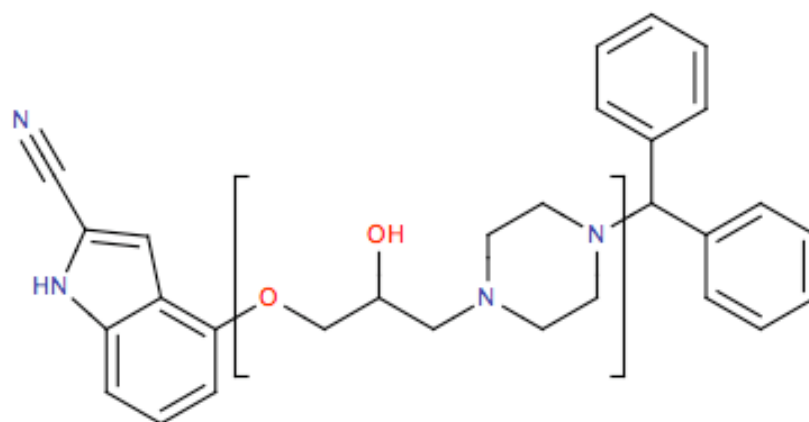


Figure 4.15 Compound #35 proposed in Tasler *et al.* study [15].



a



b

Figure 4.16 (a) Compound #3 and (b) Compound #11 in Sabio *et al.* study [11]

Chapter 5

Conclusion

In this thesis, a shared pharmacophore model was generated from five known inactive crystal structures (PDB ids: 2RH1, 3D4S, 3NY8, 3NY9, 3NYA) of human β_2 AR by LigandScout software tool. The pharmacophore model was composed of five pharmacophoric features including one hydrogen bond donor, one hydrogen bond acceptor, two hydrophobic interactions and one positive ionizable area. The model was then used to screen the clean-drug like subset of ZINC database that includes 9.928.465 compounds for the discovery of novel antagonists.

In order to test the discriminatory power of the pharmacophore model, a decoy set that consists of 53 antagonists and 64 agonists were screened first. A total of 82 hits were obtained from decoy set. The ROC curve obtained after the screening gave a cut-off point where the slope of the curve starts to become lower than 1. At this so-called “model exhaustion” or cut-off value, the threshold value of 0.64 was determined as the pharmacophore fit value for which 61 molecules (44 antagonists and 17 agonists) with a fit value above 0.64 remained. This gives 72% true positives versus 18% false positives, which is found to be a satisfying discriminatory power for the model. Taken the same threshold value for the screening of ZINC database, a total of 729.413 compounds with a score value above 0.64 were left for the next round of screening.

The next high-throughput screening is based on docking with AutoDock and GOLD software tools using various scoring functions. 61 molecules from decoy set and 729,413 compounds from ZINC database were docked to the inactive crystal structure (PDB id: 2RH1), first with GOLD using its ChemPLP scoring function. The docked poses with highest scores were evaluated based on some key residues for antagonist binding. As a result, 23,568 compounds from ZINC database and 20 molecules (17 antagonists and 3 agonists) from decoy set satisfied the binding test requirements in addition to holding a score value higher than the threshold value which is set to 85.

The remaining molecules were redocked to the same inactive crystal structure with three different docking tools, AutoDock (rescore with DSX), GOLD/ChemScore and GOLD/GoldScore. The best pose from each docking was similarly evaluated based on a stricter binding test followed by further elimination based on score value, which is set to 150, 77, 42 for DSX, GoldScore and ChemScore, respectively. All three docking results were merged and a total of 360 compounds from ZINC database and 3 antagonists from decoy set satisfied the binding test requirements and all the three threshold values. Finally, all 360 molecules and 3 antagonists were sent to ADMET filtering. As a result, 61 molecules from ZINC database and 1 antagonist from decoy set have fulfilled the requirements for human intestinal absorption (HIA) and blood brain barrier penetration and thus were proposed as potential drug candidates.

When 62 compounds were further analyzed, 7 of them were found to be isomers. The remaining 55 compounds were classified based on their molecular scaffold. Four distinct scaffolds have been detected. 49 out of 55 compounds contain one of these four scaffolds. The remaining 6 molecules had unique structures. In addition, 3 out of 55 compounds hold the well-known alprenolol scaffold. Tasler *et al.* [15] proposed a novel compound, which possess the alprenolol scaffold and our proposed scaffold listed as #4 in Table 4.6. This compound had a strong binding affinity with K_i value of 1.2nM. In addition, Sabio *et al.* [11] proposed two novel compounds from their screening studies for which their experimental binding affinities were measured as

0.311 ± 0.09 nM and 57.3 ± 1.6 nM. All two compounds possess the scaffold #3 given in Table 4.6.

References

- [1] Marjin, P.A.; Verhoeven, S.; de Graaf, C.; Roumen, L.; Vroling, B.; Nabuurs, B.S.; de Vlieg, J.; Klomp, P.G., "SNOOKER: A Structure-Based Pharmacophore Generation Tool Applied To Class A GPCRs" *J. Chem. Inf. Model*, 51: 2277-2292, 2011.
- [2] Rosenbaum, D.M.; Rasmussen, S.G.; Kobilka, B.K., "The Structure and function of G-protein-coupled Receptors" *Nature*, 459: 356-363, 2009.
- [3] Becker, M.O.; Marantz, Y.; Shacham, S.; Inbal, B.; Helfetz, A.; Kalid, O.; Bar-Halm, S.; Warshaviak, D.; Fichman, M.; Noiman, S., " G Protein-Coupled Receptors: In silico drug discovery in 3D" *Pans*, 101: 11304-11309, 2004
- [4] Lundstrom, K. "Latest development in drug discovery on G protein-coupled receptors" *Curr Protein Pept Sci*, 7:465-470, 2006.
- [5] Overington, J.P.; Al-Lazikani, B.; Hopkins, A.L., "How many drug targets are there?" *Nature reviews: drug discovery*, 5: 993-996, 2006.
- [6] Schlyer, S.; Horuk, R., "I want a mew drug: G-protein-coupled receptors in drug development" *Drug Discovery Today*, 11: 481-493, 2006.
- [7] Rasmussen, S.G.F.; Choi, H.J.; Rosenbaum, M.D.; Kobilka, T.S.; Thian, S.F.; Edwards, C.P.; Burghammer, M.; Ratnala, R.P.V.; Sanishvili, R.; Fischetti, F.R.; Schertler, F.X.G.; Weis, I.W.; Kobilka, B.K., "Crystal structure of the human β_2 adrenergic G-protein-coupled receptors" *Nature*, 450: 383-388, 2007.
- [8] Wada, M.; Kanamori, E.; Nakamura, H.; Fukunishi, Y., "Selection of In Silico Drug Screening Results for G-Protein-Coupled Receptors by Using Universal Active Probes" *J. Chem. Inf. Model*, 51: 2398-2407, 2011.

- [9] Kolb, P.; Rosenbaum, D.M.; Irwin, J.J.; Fung, J.J.; Kobilka, B.K.; Shoichet, B. K., "Structure-based discovery of beta2-adrenergic receptor ligands" *Proceedings of the National Academy of Sciences of the United States of America*, 106(16): 6843–6848, 2009.
- [10] Topiol, S.; Sabio, M., "Use of the X-ray structure of the Beta2-adrenergic receptor for drug discovery" *Bioorganic & Medicinal Chemistry Letters*, 18(5): 1598-1602, 2008.
- [11] Sabio, M.; Jones, K.; Topiol, S., "Use of X-ray structure of the beta2-adrenergic receptor for drug discovery. Part2: Identification of active compounds" *Bioorganic & Medicinal Chemistry Letters*, 18(20): 5391-5395, 2008.
- [12] de Graaf, C.; Rognan, D., "Selective structure-based virtual screening for full and partial agonists of the beta2 adrenergic receptor" *Journal of Medicinal Chemistry*, 51(16): 4978-4985, 2008
- [13] Wacker, D.; Fenalti, G.; Brown, M.A.; Katritch, V.; Abagyan, R.; Cherezov, V.; Stevens, R.C., "Conserved binding mode of human beta2 adrenergic receptor inverse agonists and antagonists revealed by X-ray crystallography" *J. Am. Chem. Soc.*, 132(33): 11443-11445, 2010.
- [14] Vilar, S.; Ferino, G.; Phatak, S.S.; Berka, B.; Cavasotto, C.N.; Costanza, S., *Journal of Molecular Graphics and Modeling*, 29; 614-623, 2011.
- [15] Tasler, S. *et.al.*, "A vHTS approach for the identification of β -adrenoceptor ligands" *Bioorganic & Medicinal Chemistry Letter*, 20: 3399-3404, 2010.
- [16] Irwin, J.J.; Shoichet, K.B., "ZINC-A Free Database of Commercially Available Compounds for Virtual Screening" *J Chem Inf Model.*, 45(1): 177-182, 2005.
- [17] Okuno, Y.; Yang, J.; Taneishi, K.; Yabuuchi, H.; Tsujimoto, G., "GLIDA: GPCR-ligand database for chemical genomics drug discovery" *Nucleic Acid Research*, 34: 673-677, 2006.
- [18] Rognan, D. "Ligand Design for G Protein-coupled Receptors", Volume 30, 2006.

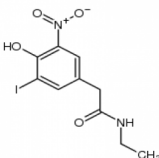
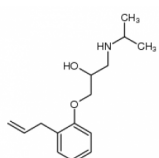
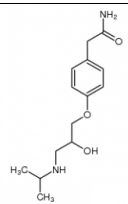
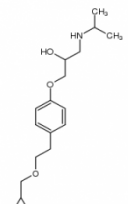
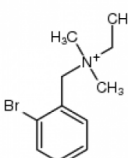
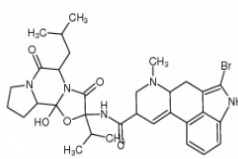
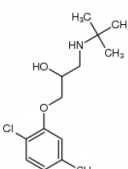
- [19] Costanzi, S., "Modeling G protein-coupled receptors and their interactions with ligands" *Curr. Opin. Struct. Biol.*, 23: 1-6, 2013.
- [20] Rasmussen, G.F.S.; Choi, H.J.; Rosenbaum, M.D.; Kobilka, S.T.; Thian, S.F.; Edwards, C.P.; Burghammer, M.; Ratnala, R.P.V.; Sanishvili, R.; Fischetti, F.R.; Schertler, F.X.G.; Weis, I.W.; Kobilka, K.B., "Crystal Structure of the human β_2 adrenergic G-protein coupled receptor" *Nature*, 450: 383-388, 2007.
- [21] Johnson, M. "Molecular mechanisms of β_2 -adrenergic receptor function, response, and regulation" *J. Allergy Clin. Immunol.*, 117: 18-24, 2006.
- [22] Wermuth, G.C.; Gonellin, C.R.; Lindber, P.; Mitscher, L.A., "Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1997)" *Annu. Rep. Med. Chem.*, 33: 385-395, 1998.
- [23] Langer, T.; Hoffman, D.R., "Pharmacophores and Pharmacophore Searches", Volume 32, 2006.
- [24] Goto, J.; Kataoka, R.; Hirayama, N., "Ph4Dock: Pharmacophore-Based Protein-Ligand Docking", *J. Med. Chem.*, 47: 6804-6811, 2004.
- [25] Phase. Schrödinger, Portland, OR. <http://www.schrodinger.com/>
- [26] Catalyst. Accelrys Software, San Diego, CA. <http://www.accelrys.com>
- [27] Wolber, G.; Langer, T., "LigandScout: 3-D Pharmacophores Derived from Protein Bound Ligands and their use as Virtual Screening Filters" *J. Chem. Inf. Model.*, 45: 160-169, 2005.
- [28] Bielska, E.; Lucas, X.; Czerwoniec, A.; Kasprzak, M.J.; Kaminska, H.K.; Bujnicki, M.J., "Virtual Screening strategies in drug design-methods and applications" *J. Biotechnology, Computational Biology and Bionanotechnology*, 92: 249-264, 2011.
- [29] Sottriffer, C., "Virtual Screening: Principles, Challenges and Practical Guidelines", Volume 32, 2006.

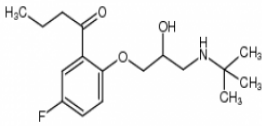
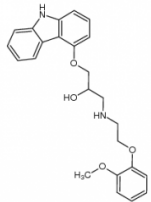
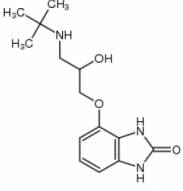
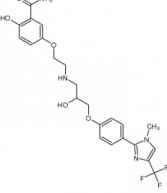
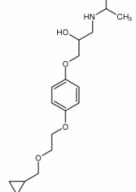
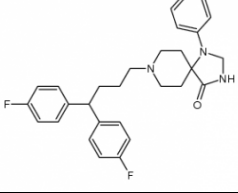
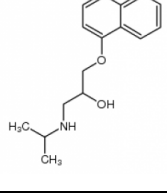
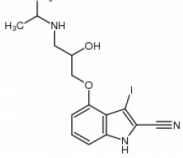
- [30] Chen, Z.; Li, H.; Zhang, Q.; Bao, X.; Yu, K.; Lou, X.; Zhu, X.; Jiang, H., "Pharmacophore-based virtual screening versus docking-based screening: a benchmark comparison against eight targets" *Acta Pharmacologica Sinica*, 30: 1694-1708, 2009.
- [31] McInnes, C., "Virtual screening strategies in drug discovery" *Curr. Opin Chem. Biol.* 11: 494-502, 2007.
- [32] Hou, T.; Xu, X., "Recent Development and Application of Virtual Screening in Drug Discovery: An Overview" *Curr. Phar. Des.*, 10: 1011-1033, 2004.
- [33] Mositessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C.R., "Towards the development of universal, fast and highly accurate docking/scoring methods :a long way to go" *British Journal of Pharmacology*, 153: 7-26, 2008.
- [34] Ktichen, B.D.; Decornez, H.; Furr, R.J.; Bajorath, J., " Docking and Scoring in Virtual for Drug Discovery: Methods and Applications " *Nature*, 3; 935-949, 2004.
- [35] Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M., "CHARMm: A program for macromolecular energy minimization, and Dynamics calculations" *J. Comput. Chem.*, 4: 187-217, 1983.
- [36] Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A., "An all atom force field for simulations of protiens and numcleic acids." *J. Comput. Chem.*, 7: 230-252, 1986.
- [37] Autodock Version 4.2 Online Manual. <http://autodock.scripps.edu/downloads>
- [38] GOLD User Guide. <http://www.ccdc.cam.ac.uk/Solutions/GoldSuite>
- [39] Zhong, S.; Zhang, Y.; Xiu, Z., "Rescoring ligand docking poses." *Current Opinion in Drug Discovery and Development*, 13: 326-334, 2010.
- [40] Rizzi, A.; Fioni, A., "Virtual Screening Using PLS Discriminant Analysis and ROC Curve Approach: An Application Study on PDE4 Inhibitors", *J. Chem. Inf. Model*, 48: 1686-1692, 2008.

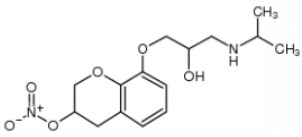
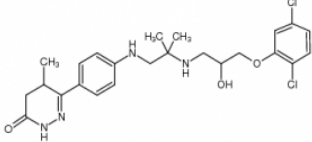
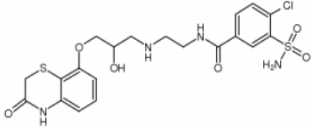
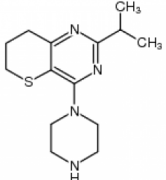
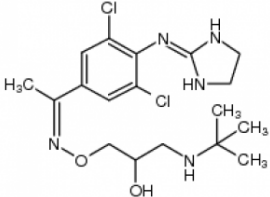
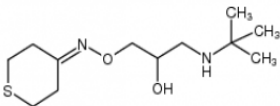
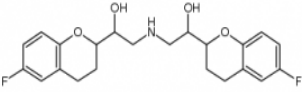
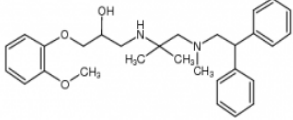
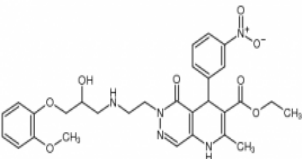
- [41] Chemnavigator. <http://www.chemnavigator.com>
- [42] Von Grotthus, M.; Pas, J.; Rychleski, L., "Ligand-Info, searching for similar small compounds using index profiles" *Bioinformatics*, 19: 1041-1042, 2003.
- [43] Strausberg, R.L.; Schreiber, S.L., "From knowing to controlling: a path from genomics to drug design using small molecule probes" *Science*, 300: 294-295, 2003.
- [44] Kroeme, T.R., "Structure-Based Drug Design: Docking and Scoring" *Current Protein and Peptide Science*, 8: 312-328, 2007.
- [45] Yang, J; Okuno, Y.; Tsujimoto, G., "GLIDA: GPCR and Ligand Database" *Genome Informatics*. 15: 57, 2004.
- [46] Horn, F.; Bettler, E.; Oliviera, L; Campagne, F.; Cohen F.E.; Vriend, G., "GPCRDB information system for G-protein coupled receptors", *Nucleic Acids Res.*, 31: 294-297, 2003.
- [47] Fredholm, B.B.; Fleming, W.W.; Vanhoutte, P.M.; Godfraind, T., "The role of pharmacology in drug discovery" *Nature Rev. Drug Discov.* 1: 237-248, 2002.
- [48] Okuno, Y.; Tamon, A.; Yabuuchi, H. Niijima, S.; Minova, Y.; Tonomura, K.; Kunimoto, R.; Fencg, C., "GLIDA: GPCR-ligand database for chemical genomics drug discovery-database and updates" *Nucleic Acids Res.*, 36: 907-912, 2008.
- [49] CORINA Molecular Networks. www.molecular-networks.de
- [50] Prentis, R.A.; Lis, Y.; Walker, S.R., *Br. J. Clin. Pharmac.* 25: 387-396, 1988.

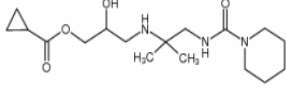
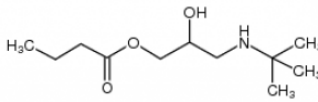
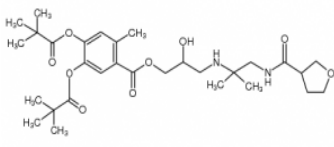
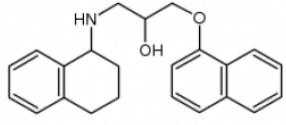
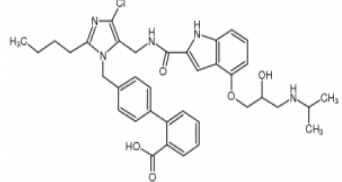
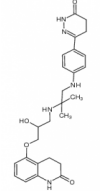
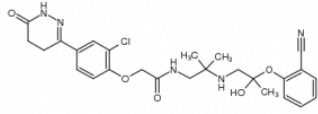
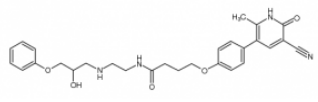
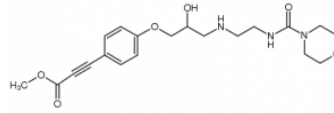
APPENDIX A

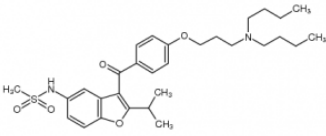
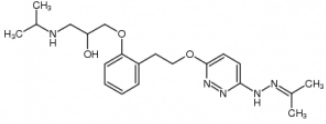
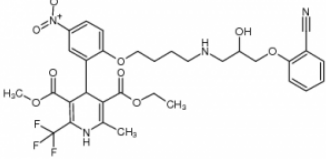
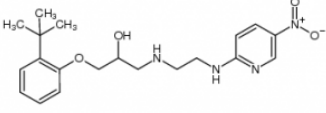
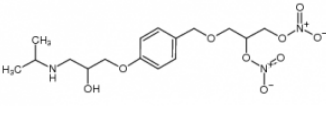
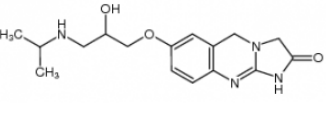
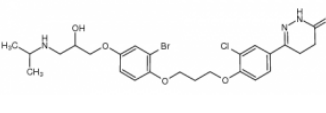
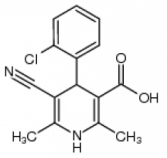
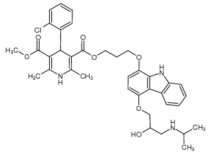
Representation of 53 antagonists taken from GLIDA database

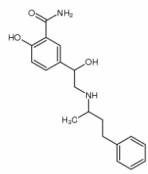
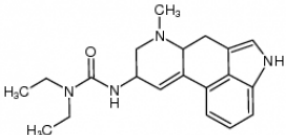
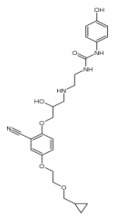
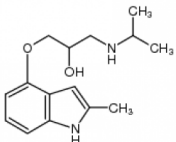
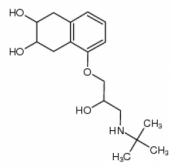
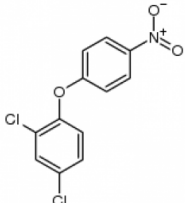
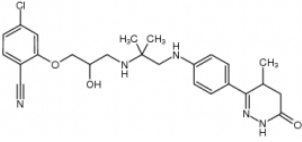
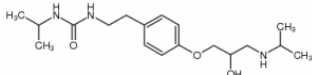
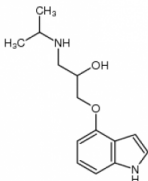
Compound Number	Compound Name	Chemical Formula	Chemical Sketches	MW (g/mol)
1	4 Hydroxy	$C_{10}H_{11}IN_2O_4$		350.11
2	Alprenolol	$C_{15}H_{23}NO_2$		249.173
3	Atenolol_Tenoretic	$C_{14}H_{22}N_2O_3$		266.163
4	Betaxolol	$C_{18}H_{29}NO_3$		307.214
5	Bretylum	$C_{11}H_{17}NBr$		242.054
6	Bromocriptine	$C_{32}H_{40}N_5O_5Br$		653.221
7	Bupranolol	$C_{14}H_{22}NO_2Cl$		271.134

Compound Number	Compound Name	Chemical Formula	Chemical Sketches	MW (g/mol)
8	Butofilolol	$C_{17}H_{26}NO_3F$		311.189
9	Carvedilol	$C_{24}H_{26}N_2O_4$		406.189
10	CGP12177	$C_{14}H_{21}N_3O_3$		279.158
11	CGP20712A	$C_{23}H_{25}N_4O_5F_3$		494.177
12	Cicloprolol	$C_{18}H_{29}NO_4$		323.209
13	Fluspirilene	$C_{29}H_{31}N_3OF_2$		475.243
14	Inderex_Propranolol	$C_{16}H_{21}NO_2$		259.157
15	Iodocyanopindolol	$C_{15}H_{18}N_3O_2I$		399.044

Compound Number	Compound Name	Chemical Formula	Chemical Sketches	MW (g/mol)
16	K351	$C_{15}H_{22}N_2O_6$		326.148
17	L001727	$C_{24}H_{30}N_4O_3Cl_2$		492.169
18	L001755	$C_{20}H_{23}N_4O_6S_2Cl$		492.169
19	L002134	$C_{14}H_{22}N_4S$		492.169
20	L002301	$C_{18}H_{27}N_5O_2Cl_2$		415.154
21	L002666	$C_{12}H_{24}N_2O_2S$		260.156
22	L003304	$C_{22}H_{25}NO_4F_2$		405.175
23	L003361	$C_{29}H_{38}N_2O_3$		415.154
24	L003769	$C_{29}H_{33}N_5O_8$		579.232

Compound Number	Compound Name	Chemical Formula	Chemical Sketches	MW (g/mol)
25	L003915	C ₁₇ H ₃₁ N ₃ O ₄		341.231
26	L003916	C ₁₁ H ₂₃ NO ₃		217.168
27	L003926	C ₃₀ H ₄₆ N ₂ O ₉		578.320
28	L003966	C ₂₃ H ₂₅ NO ₂		347.188
29	L004107	C ₃₇ H ₄₂ N ₅ O ₅ Cl		671.287
30	L004502	C ₂₆ H ₃₃ N ₅ O ₄		479.253
31	L004939	C ₂₆ H ₃₀ N ₅ O ₅ Cl		527.193
32	L005112	C ₂₈ H ₃₂ N ₄ O ₅		504.237
33	L005221	C ₂₀ H ₂₇ N ₃ O ₆		405.190

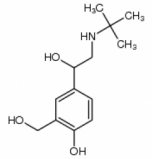
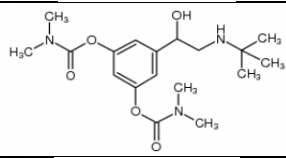
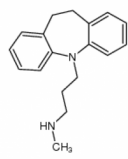
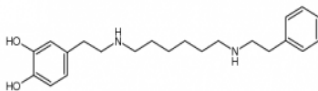
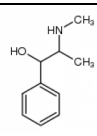
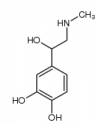
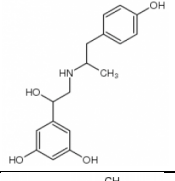
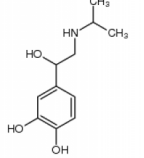
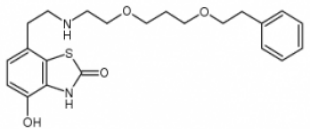
Compound Number	Compound Name	Chemical Formula	Chemical Sketches	MW (g/mol)
34	L006500	$C_{30}H_{42}N_2O_5S$		542.281
35	L011199	$C_{21}H_{31}N_5O_3$		401.242
36	L013581	$C_{32}H_{35}N_4O_9F_3$		676.235
37	L017579	$C_{20}H_{28}N_4O_4$		388.211
38	L019300	$C_{16}H_{25}N_3O_9$		403.159
39	L023024	$C_{16}H_{22}N_4O_3$		318.169
40	L023037	$C_{25}H_{31}N_3O_5BrCl$ 1		567.113
41	L023755	$C_{15}H_{13}N_2O_2Cl$		288.066
42	L023760	$C_{37}H_{42}N_3O_7Cl$		675.271

Compound Number	Compound Name	Chemical Formula	Chemical Sketches	MW (g/mol)
43	Labetalol	$C_{19}H_{24}N_2O_3$		328.178
44	Lisuride	$C_{20}H_{26}N_4O$		338.210
45	LK204-545	$C_{25}H_{32}N_4O_6$		484.232
46	Mepindolol	$C_{15}H_{22}N_2O_2$		262.168
47	Nadolol	$C_{17}H_{27}NO_4$		309.194
48	Nitrofen	$C_{12}H_7O_3Cl_2$		282.980
49	Oberadilol	$C_{25}H_{30}N_5O_3Cl$		483.203
50	Pafendolol	$C_{18}H_{31}N_3O_3$		337.236
51	Pindolol	$C_{14}H_{20}N_2O_2$		248.153

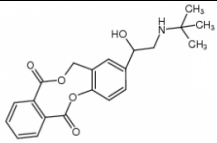
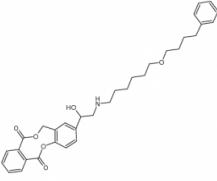
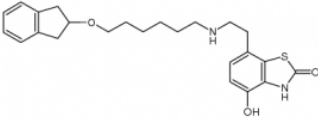
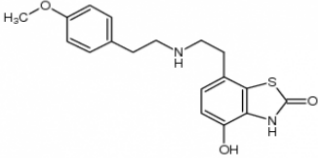
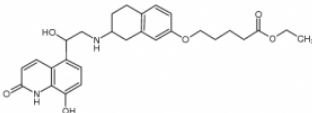
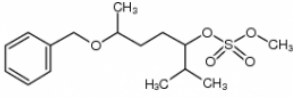
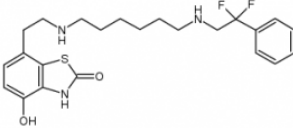
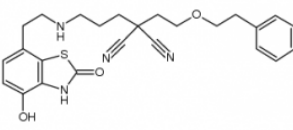
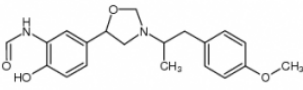
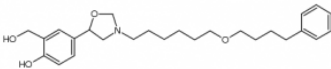
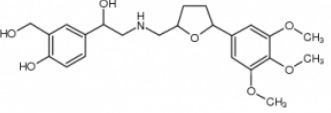
Compound Number	Compound Name	Chemical Formula	Chemical Sketches	MW (g/mol)
52	SR59230A	$C_{21}H_{27}NO_2$		325.204
53	Timolol	$C_{13}H_{24}N_4O_3S$		316.157

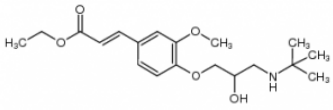
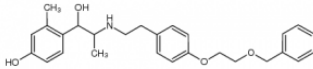
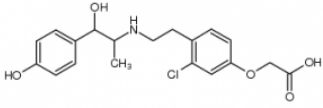
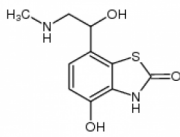
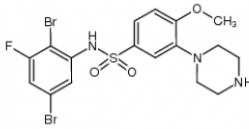
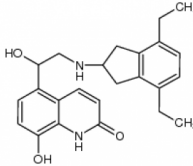
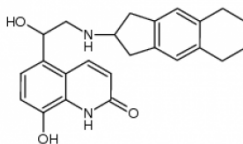
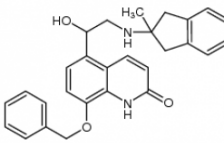
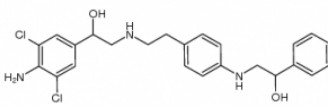
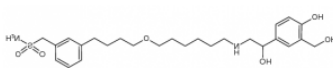
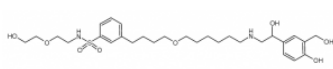
APPENDIX B

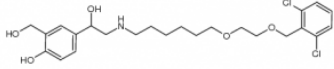
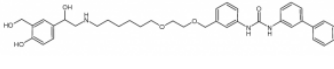
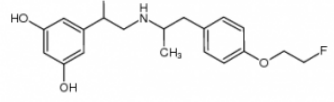
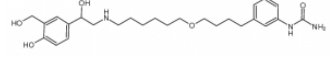
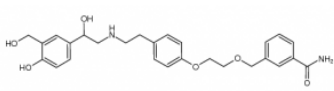
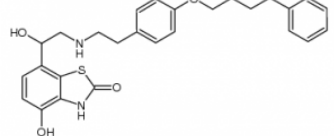
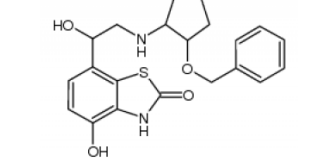
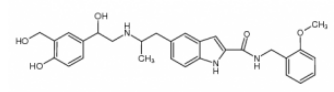
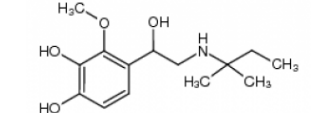
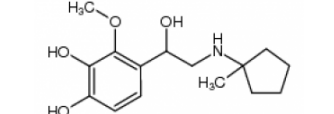
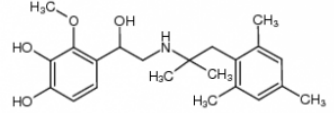
Representation of 64 agonists taken from GLIDA database

Compound Number	Compound Name	Chemical Formula	Chemical Sketches	MW (g/mol)
1	Albuterol	$C_{13}H_{21}NO_3$		239.152
2	Bambuterol	$C_{18}H_{29}N_3O_5$		367.210
3	Desipramine	$C_{18}H_{22}N_2$		266.178
4	Dopexamine	$C_{22}H_{32}N_2O_2$		356.246
5	Ephedrine	$C_{10}H_{15}NO$		165.115
6	Epinephrine	$C_9H_{13}NO_3$		183.090
7	Fenoterol	$C_{17}H_{21}NO_4$		303.147
8	Isoprotenerol	$C_{11}H_{17}NO_3$		211.121
9	L001484	$C_{22}H_{28}N_2O_4S$		416.177

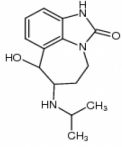
Compound Number	Compound Name	Chemical Formula	Chemical Sketches	MW (g/mol)
10	L001619	C ₁₉ H ₂₃ NO ₅		345.157
11	L002148	C ₂₂ H ₂₈ N ₂ O ₂ ClF ₃		444.179
12	L002189	C ₂₅ H ₃₇ NO ₄ S		447.244
13	L002190	C ₂₈ H ₄₁ N ₃ O ₃ Cl ₂		537.252
14	L002585	C ₂₈ H ₃₉ N ₃ O ₅ S		529.261
15	L004003	C ₁₈ H ₂₂ N ₄ O		310.179
16	L004012	C ₁₉ H ₂₁ N ₅ O		335.174
17	L004861	C ₂₅ H ₃₅ NO ₄		413.256
18	L004863	C ₃₃ H ₅₁ N ₃ O ₆ S		617.349
19	L004864	C ₂₁ H ₃₄ N ₂ O ₃ Cl ₂		432.194
20	L004930	C ₂₉ H ₃₇ NO ₄		463.272

Compound Number	Compound Name	Chemical Formula	Chemical Sketches	MW (g/mol)
21	L005041	C ₂₁ H ₂₃ NO ₅		369.157
22	L005043	C ₃₃ H ₃₉ NO ₆		545.277
23	L007006	C ₂₄ H ₃₀ N ₂ O ₃ S		426.197
24	L007013	C ₁₈ H ₂₀ N ₂ O ₃ S		344.119
25	L009579	C ₂₈ H ₃₄ N ₂ O ₆		494.241
26	L009610	C ₁₆ H ₂₆ O ₅ S		330.150
27	L009867	C ₂₃ H ₂₉ N ₃ O ₂ SF ₂		449.195
28	L009870	C ₂₅ H ₂₈ N ₄ O ₃ S		464.188
29	L010630	C ₂₀ H ₂₄ N ₂ O ₄		356.173
30	L013462	C ₂₆ H ₃₇ NO ₄		427.272
31	L013880	C ₂₃ H ₃₁ NO ₇		433.210

Compound Number	Compound Name	Chemical Formula	Chemical Sketches	MW (g/mol)
32	L017717	C ₁₉ H ₂₉ NO ₅		351.204
33	L018030	C ₂₇ H ₃₃ NO ₄		435.241
34	L018574	C ₁₉ H ₂₂ NO ₅ Cl		379.118
35	L018634	C ₁₀ H ₁₂ N ₂ O ₃ S		240.057
36	L020233	C ₁₇ H ₁₈ N ₃ O ₃ SBr ₂ F		520.942
37	L20300	C ₂₄ H ₂₈ N ₃ O ₃		392.210
38	L20301	C ₂₄ H ₂₆ N ₂ O ₃		390.194
39	L020302	C ₂₈ H ₂₈ N ₂ O ₃		440.210
40	L020651	C ₂₄ H ₂₇ N ₃ O ₂ Cl ₂		459.148
41	L021740	C ₂₆ H ₄₀ N ₂ O ₆ S		508.260
42	L021744	C ₂₉ H ₄₆ N ₂ O ₈ S		582.297

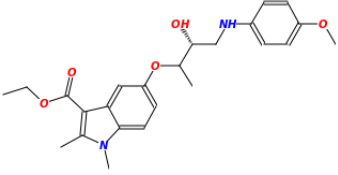
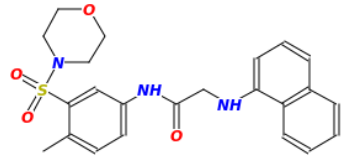
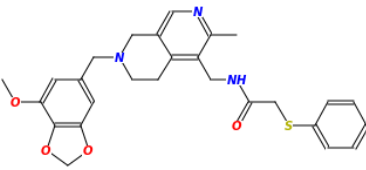
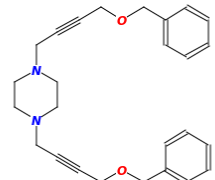
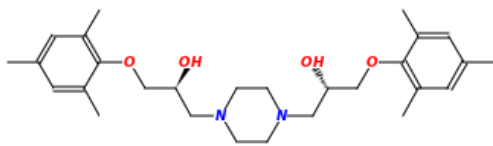
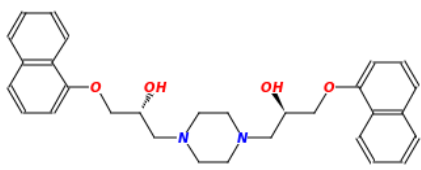
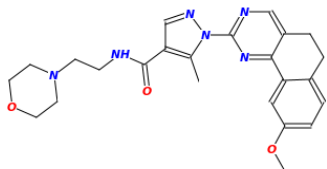
Compound Number	Compound Name	Chemical Formula	Chemical Sketches	MW (g/mol)
43	L022137	$C_{24}H_{33}NO_5Cl$		485.173
44	L022143	$C_{37}H_{45}N_3O_6$		627.330
45	L022292	$C_{19}H_{24}NO_4F$		349.169
46	L022352	$C_{27}H_{41}N_3O_5$		487.304
47	L022477	$C_{27}H_{32}N_2O_6$		480.226
48	L022740	$C_{27}H_{30}N_2O_4S$		478.192
49	L022749	$C_{21}H_{24}N_2O_4S$		400.145
50	L022822	$C_{29}H_{33}N_3O_5$		503.242
51	L022866	$C_{14}H_{23}NO_4$		269.163
52	L022867	$C_{15}H_{23}NO_4$		281.162
53	L022868	$C_{22}H_{31}NO_4$		373.225

Compound Number	Compound Name	Chemical Formula	Chemical Sketches	MW (g/mol)
54	L022921	C ₃₄ H ₄₇ N ₂ O ₆ S		610.307
55	L023245	C ₃₆ H ₃₉ N ₃ O ₄		577.294
56	L023254	C ₂₀ H ₂₀ N ₂ O ₃		336.147
57	L023257	C ₂₂ H ₃₀ N ₂ O ₃		370.225
58	L023259	C ₂₂ H ₂₄ N ₂ O ₄		380.173
59	L023260	C ₂₂ H ₃₁ N ₃ O ₃		385.236
60	L023322	C ₃₈ H ₄₈ N ₄ O ₄		624.367
61	L023497	C ₂₈ H ₃₀ N ₂ O ₄ F ₄		534.214
62	Norepinephrine	C ₈ H ₁₁ NO ₃		169.074
63	Salmeterol	C ₂₅ H ₃₇ NO ₄		415.272

Compound Number	Compound Name	Chemical Formula	Chemical Sketches	MW (g/mol)
64	Zilpaterol	$C_{14}H_{19}N_3O_2$	 <p>The chemical structure of Zilpaterol is a complex heterocyclic molecule. It features a benzene ring fused to a seven-membered ring containing a nitrogen atom. The nitrogen atom is part of a cyclic amide group (N-C=O). A hydroxyl group (-OH) is attached to the benzene ring. Another nitrogen atom is attached to the seven-membered ring, and this nitrogen is further substituted with an isopropyl group (-CH(CH₃)₂).</p>	261.148

APPENDIX C

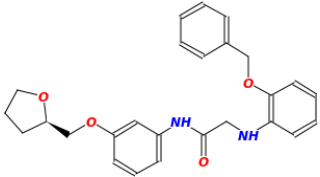
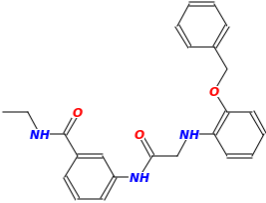
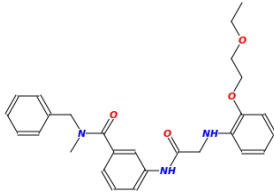
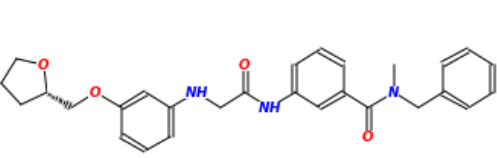
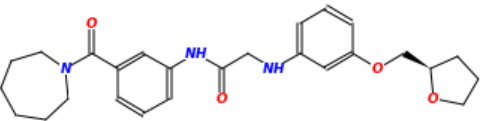
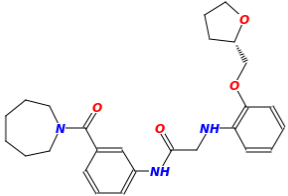
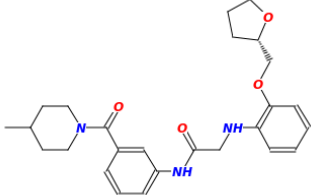
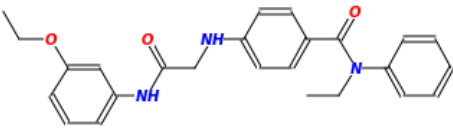
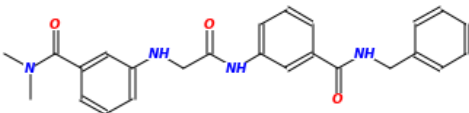
Representation of 55 molecules proposed as possible drug candidates.

ZINC ID	NAME	Scaffold ID	2D Representation
ZINC 06732830	Ethyl 5-{2-hydroxy-3-[(4-methoxyphenyl)amino]propoxy}-1,2-dimethylindole-3-carboxylate	Unique	
ZINC 15086276	N-(4-methyl-3-morpholinosulfonylphenyl)-2-(1-naphthylamino)acetamide	1	
ZINC 19229410	N-({7-[(7-methoxy-1,3-benzodioxol-5-yl)methyl]-3-methyl-5,6,7,8-tetrahydro-2,7-naphthyridin-4-yl)methyl)-2-(phenylthio)acetamide	2	
ZINC 19367103	1,4-di(4-benzyloxy-2-butynyl)piperazine; hydrochloride	Unique	
ZINC 19901106	1-{4-[2-hydroxy-3-(mesityloxy)propyl]-1-piperazinyl}-3-(mesityloxy)-2-propanol	3	
ZINC 19904796 (1 isomer)	1,1'-piperazine-1,4-diylbis[3-(naphthalen-1-yloxy)propan-2-ol]	3	
ZINC 20451620	1-(9-methoxy-5,6-dihydrobenzo[h]quinazolin-2-yl)-5-methyl-N-[2-(4-morpholinyl)ethyl]-1H-pyrazole-4-carboxamide	2	

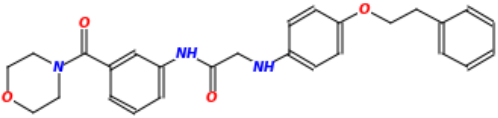
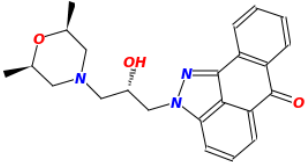
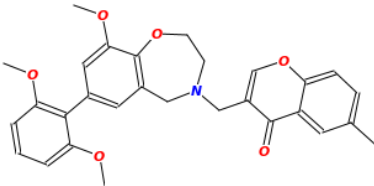
ZINC ID	NAME	Scaffold ID	2D Representation
ZINC 22130659	(2S)-1-[4-[(2R)-2-hydroxy-3-(2-naphthylloxy)propyl]piperazin-1-yl]-3-(2-naphthylloxy)propan-2-ol	3	
ZINC 22289083 (3 isomers)	1-[4-[2-hydroxy-3-(1-naphthylloxy)propyl]-1-piperazinyl]-3-(2-naphthylloxy)-2-propanol	3	
ZINC 22924215	1,1'-piperazine-1,4-diylbis[3-(naphthalen-1-yloxy)propan-2-ol]	3	
ZINC 32676044 (1 isomer)	N-[(2S)-2-(dimethylamino)-2-(2-thienyl)ethyl]-4-[(2-naphthylsulfonylamino)methyl]benzamide	2	
ZINC 33034896	3-[(3R)-1-(4-ethylbenzoyl)-3-piperidyl]-N-(2-morpholinoethyl)benzamide	3	
ZINC 34667246	N-[(2R)-4-[(2S)-3-benzyloxy-2-hydroxy-propyl]morpholin-2-yl]methyl]-4-fluoro-N-isobutylbenzamide	4	
ZINC 34667284	N-[(2R)-4-[(2S)-3-(4-chlorophenoxy)-2-hydroxy-propyl]morpholin-2-yl]methyl]-N-isobutyl-2-phenylacetate	4	
ZINC 34691493	N-[(2S)-4-[(2S)-3-(4-fluorophenoxy)-2-hydroxy-propyl]morpholin-2-yl]methyl]-N-isobutyl-2-phenylacetate	4	
ZINC 34691722	N-cyclopentyl-N-[(2R)-4[(2S)-2-hydroxy-3-(2-methylphenoxy)propyl]morpholin-2-yl]methyl]-4-methyl-b	4 Alprenol Scaffold	

ZINC ID	NAME	Scaffold ID	2D Representation
ZINC 34691828	(2,6-dimethoxyphenyl)-[4-[(2S)-4-[(2-hydroxyphenyl)methyl]morpholin-2-yl]methyl]piperazin-1-yl]meth	Unique	
ZINC 34717124	N-[(4-fluorophenyl)methyl]-2-[[[(1R)-1-(m-tolyl)-2-propanoyl-3,4-dihydro-1H-isoquinolin-7-yl]oxy]acet	2	
ZINC 35854704	(2S)-2-[[[(1R)-2-(3-methylbutanoyl)-1-phenyl-3,4-dihydro-1H-isoquinolin-7-yl]oxy]-N-(2-pyridylmethyl)	2	
ZINC 35865478	N-[[[(2R)-4-[(2S)-3-benzyloxy-2-hydroxypropyl]morpholin-2-yl]methyl]-N-isobutyl-2-phenoxyacetamide	4	
ZINC 35865581	N-[[[(2R)-4-[(2S)-3-(2-furylmethoxy)-2-hydroxypropyl]morpholin-2-yl]methyl]-N-(2-methoxyethyl)-4-met	4	
ZINC 35865918	N-[[[(2R)-4-[(2S)-2-hydroxy-3-(2-methylphenoxy)propyl]morpholin-2-yl]methyl]-N-isobutyl-4-methyl-benz	4 Alprenol Scaffold	
ZINC 35866100	[4-[[[(2R)-4-[(2R)-2-hydroxy-3-(2-methylphenoxy)propyl]morpholin-2-yl]methyl]piperazin-1-yl]-(3-metho	4 Alprenol Scaffold	
ZINC 37602474	N-benzyl-2-[[[(1S)-2-(cyclopropanecarbonyl)-1-(p-tolyl)-3,4-dihydro-1H-isoquinolin-7-yl]oxy]acetamide	2	
ZINC 38714609	3,3-diphenyl-N-(2-propylsulfonyl-3,4-dihydro-1H-isoquinolin-7-yl)propanamide	2	

ZINC ID	NAME	Scaffold ID	2D Representation
ZINC 40721209	5-[(1R)-2-[4-[[2-(2-fluorophenyl)ethylamino]methyl]phenoxy]-1-hydroxyethyl]-1,3-dimethylbenzimidaz	Unique	
ZINC 51450845	[4-[(2S)-3-(2-allylphenoxy)-2-hydroxypropyl]piperazin-1-yl]-(2,5-dimethyl-3-furyl)methanone	3	
ZINC 58167132	6-cyclopropyl-3-methyl-N-[1-(2-morpholinoethyl)indol-5-yl]isoxazolol[5,4-b]pyridine-4-carboxamide	2	
ZINC 60105533	2-[(4-fluorophenyl)methoxy]-N-[[3-(2-morpholinoethoxy)phenyl]methyl]benzamide	2	
ZINC 66062891 (1 isomer)	3-[[2-[2-(2-methylallyloxy)anilino]acetyl]amino]-N-[[[(2R)-tetrahydrofuran-2-yl]methyl]benzamide	1	
ZINC 66062896	2-[2-(2-methylallyloxy)anilino]-N-[3-(morpholine-4-carbonyl)phenyl]acetamide	1	
ZINC 66064410	3-[[2-(3-benzyloxyanilino)acetyl]amino]-N,N-diethylbenzamide	1	
ZINC 66064457	3-[[2-(3-benzyloxyanilino)acetyl]amino]-N-(3-methoxypropyl)benzamide	1	
ZINC 66064646	3-[[2-(3-allyloxyanilino)acetyl]amino]-N-[[[(2S)-tetrahydrofuran-2-yl]methyl]benzamide	1	

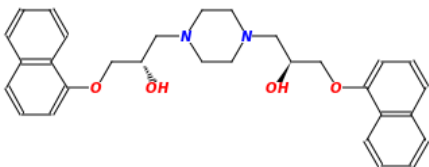
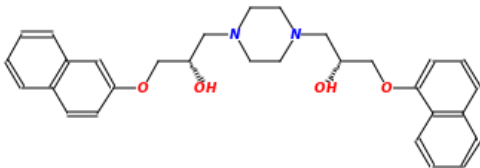
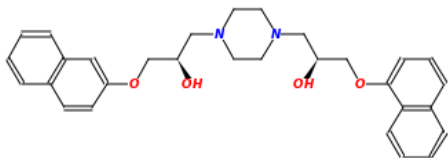
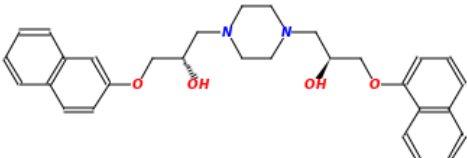
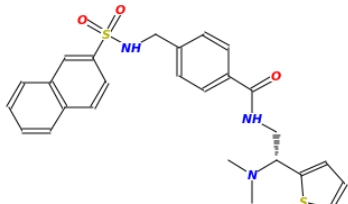
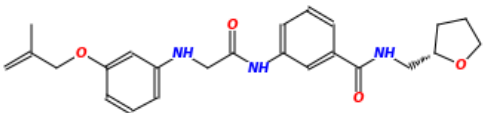
ZINC ID	NAME	Scaffold ID	2D Representation
ZINC 66065048	2-(2benzyloxyanilino)-N-[3-[[[(2R)-tetrahydrofuran-2-yl]methoxy]phenyl]acetamide	1	
ZINC 66065105	3-[[2-(2-benzyloxyanilino)acetyl]amino]-N-ethylbenzamide	1	
ZINC 66065561	N-benzyl-3-[[2-[2-(2-ethoxyethoxy)anilino]acetyl]amino]-N-methylbenzamide	1	
ZINC 66068768 (1 isomer)	N-benzyl-N-methyl-3-[[2-[3-[[[(2S)-tetrahydrofuran-2-yl]methoxy]anilino]acetyl]amino]benzamide	1	
ZINC 66068775	N-[3-(azepane-1-carbonyl)phenyl]-2-[3-[[[(2R)-tetrahydrofuran-2-yl]methoxy]anilino]acetamide	1	
ZINC 66070858	N-[3-(azepane-1-carbonyl)phenyl]-2-[2-[[[(2S)-tetrahydrofuran-2-yl]methoxy]anilino]acetamide	1	
ZINC 66070868	N-[3-(4-methylpiperidine-1-carbonyl)phenyl]-2-[2-[[[(2S)-tetrahydrofuran-2-yl]methoxy]anilino]acetamide	1	
ZINC 66092097	4-[[2-(3-ethoxyanilino)-2-oxo-ethyl]amino]-N-ethyl-N-phenylbenzamide	1	
ZINC 66093249	3-[[2-[3-(benzylcarbonyl)anilino]-2-oxo-ethyl]amino]-N,N-dimethylbenzamide	1	

ZINC ID	NAME	Scaffold ID	2D Representation
ZINC 66093361	3-[[2-(2-naphthylamino)acetyl]amino]-N-[[2-(2R)-tetrahydrofuran-2-yl]methyl]benzamide	1	
ZINC 66096875	3-[[2-(3-ethoxyanilino)-2-oxo-ethyl]amino]-N-[(1R)-1-phenylethyl]benzamide	1	
ZINC 66106842	N-(3-phenethoxyphenyl)-2-[4-piperidine-1-carbonyl]anilino]acetamide	1	
ZINC 66110260	N,N-dimethyl-3-[3-[[2-oxo-2-[3-(3-phenylpropoxy)anilino]ethyl]amino]phenyl]propanamide	1	
ZINC 66110539	3-[[2-[4-[3-(dimethylamino)-3-oxo-propyl]anilino]acetyl]amino]-N-propylbenzamide	1	
ZINC 66216270	3-[[2-(2-ethoxyanilino)acetyl]amino]-N-[(1R)-1-phenylethyl]benzamide	1	
ZINC 66226867	3-[[2-(4-benzyloxyanilino)acetyl]amino]-N-(2-methoxyethyl)benzamide	1	
ZINC 66226908	3-[3-[[2-(4-benzyloxyanilino)acetyl]amino]phenyl]-N,N-dimethylpropanamide	1	
ZINC 66227757	N[3-(4-methylpiperidine-1-carbonyl)phenyl]-2-[4-[[2-(2S)-tetrahydrofuran-2-yl]methoxy]anilino]acetamide	1	

ZINC ID	NAME	Scaffold ID	2D Representation
ZINC 66229400	N-[3-(morpholine-4-carbonyl)phenyl]-2-(4-phenethoxyanilino)acetamide	1	
ZINC 66482925	[(2S)-3-[(2S,6R)-2,6-dimethylmorpholin-4-yl]-2-hydroxypropyl]BLAHone	Unique	
ZINC 67674643	3-{{[7-(2,6-dimethoxyphenyl)-9-methoxy-2,3-dihydro-1,4-benzoxazepin-4(5H)-yl]methyl}-6-methyl-4H-chromen-4-one	Unique	

APPENDIX D

Representation of the remaining 7 potential drug candidates that are isomers of the molecules listed in Appendix C

ZINC ID	NAME	2D Representation
ZINC 19904800 (isomer of ZINC19904796)	1,1'-piperazine-1,4-diylbis[3-(naphthalen-1-yloxy)propan-2-ol]	
ZINC 22289086 (isomer of ZINC22289083)	1-{4-[2-hydroxy-3-(1-naphthyloxy)propyl]-1-piperazinyl}-3-(2-naphthyloxy)-2-propanol	
ZINC 22289089 (isomer of ZINC22289083)	1-{4-[2-hydroxy-3-(1-naphthyloxy)propyl]-1-piperazinyl}-3-(2-naphthyloxy)-2-propanol	
ZINC 22289092 (isomer of ZINC22289083)	1-{4-[2-hydroxy-3-(1-naphthyloxy)propyl]-1-piperazinyl}-3-(2-naphthyloxy)-2-propanol	
ZINC 32676046 (isomer of ZINC32676044)	N-[(2R)-2-(dimethylamino)-2-(2-thienyl)ethyl]-4-[(2-naphthylsulfonylamino)methyl]benzamide	
ZINC 66063110 (isomer of ZINC66062891)	3-[[2-[3-(2-methylallyloxy)anilino]acetyl]amino]-N-[[[(2S)-tetrahydrofuran-2-yl]methyl]benzamide	

ZINC ID	NAME	2D Representation
ZINC 66068769 (isomer of ZINC66068768)	N-benzyl-N-methyl-3-[[2- [3-[[[(2R)-tetrahydrofuran- 2- yl]methoxy]anilino]acetyl]amino]benzamide	

Cirriculum Vitae

Rüya Yakar was born on 12 January 1987, in Istanbul. She received her BS degree in Information Technologies in 2010 from Isik University.