Structure Based Drug Design

by

Bahar ÖNDÜL

A Thesis Submitted to the

Graduate School of Engineering

in Partial Fulfillment of the Requirements for

the Degree of

Master of Science

in

Computational Science and Engineering

Koç University

January, 2009

Koc University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Bahar ONDUL

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Prof. Burak Erman (Advisor)

_____

Assoc. Prof. Attila Gürsoy

_____

Assoc. Prof. Özlem Keskin

_____

_____

Date: _____16.01.2009_____

ABSTRACT

The aim of this study is to target two possible pathways of cancer which are (i) p65p50 protein heterodimer from NF-κB protein family and (ii) Fgfr2 from fibroblastic growth hormone receptors. A structure-based screening approach, based on a data set of 236 natural products for each protein is adopted. Docking analysis is performed for structure-based screening. Three scoring functions are used for docking. Docking results of both proteins and their ranking are compared using known scoring functions. It is shown that scoring functions have a tendency to rank specific structural groups selectively. AutoDock, which is based on a semi-empirical function, performs the rankings according to the properties of the ligands in the data set. Hence, AutoDock ranks via selecting a specific group from the data set. The two other scoring functions, Gold Score and Chem Score that come with the software GOLD (Genetic Optimization for Ligand Docking) show that specific groups of molecules are ranked selectively according to the properties of both the system and the ligand. Variation between selectivity of different scoring functions is based on the main equations used within each scoring function. On the other hand, as also reported in recent studies, although the ranking of a docked protein-ligand pair depends on the scoring function used, the orientation and interaction modes of docking performed with different scoring functions are similar.

ÖZET

Bu çalışmada kanser yol izi mekanizmaları içerisinde yer aldığı düşünülen iki protein hedef olarak belirlenmiştir, bunlar Nfκβ protein ailesiden p65p50 yapısı ve fibroblastik büyüme hormon reseptörlerinden Fgfr2 yapısıdır. Bilgisayar bazlı ilaç tasarım metodlarından yapı-etki ilişkisi odaklı tarama yapmaya ve efektör-hedef etkileşimininin yapısal tanımını yapmaya imkan veren "docking" metodu 237 doğal bileşikten oluşan bir moleküler veri bankasına iki ayrı sistem için de uygulanmıştır. "Docking" metodu için üç ayrı değerlendirme fonksiyonu kullanılmıştır. Bu değerlendirme fonksiyonlarından elde edilen sıralama sonuçları doğrultusunda hem sistemler için hem de fonksiyonların özelliklerine dair karşılaştırma yapılmıştır. Bu araştırmanın sonunda fonksiyonların bileşik grupları ayırdedebildikleri ve bu gruplar üzerinde seçici oldukları görülmüştür. Yarı ampirik kaynaklı değerlendirme fonksiyonundan ibaret olan AutoDock programının efektörün niteliklerine göre sıralama yaptığı ve özel bir grup bileşiğini bu doğrultuda seçtiği gözlenmiştir. Gold Score ve Chem Score olmak üzere iki ayrı değerlendirme fonksiyonu ile kullanılan Gold (Ligand "docking"i için Genetik optimizasyon) programının sıralama yaparken hedefin niteliklerine göre özel bir grup bileşiği seçtiği gözlenmiştir. Fonksiyonların bileşik grupları için gösterdikleri seçiciliğin nedeni her fonksiyonun ana denkleminde gizlidir. Ancak şu da görülmüştür ki, daha önceki çalışmalarda da olduğu gibi aynı hedef ve aynı ligant için yapılan analiz sonuçları, değerlendirme fonksiyonları sıralamaları farklı olmakla birlikte, analiz sonucu üretilen etkileşim özellikleri ve ligant-protein konumlandırmalarında yüksek benzerlik göstermektedir.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

vii

Chapter 1

LITERATURE REVIEW

In this chapter, we perform a literature survey for the protein NF-κB with specific emphasis on its relation to cancer [1], [2]. A similar but less detailed survey is performed for our second system, the Fgfr2 protein in Chapter 5.

1.1 NF-κB

Nuclear factor kappa B (NF-κB) is a group of proteins that are specified as transcription factors. When NF-κB is activated it is translocated to the nucleus. Transcription factors in general have the ability to bind to DNA in order to effect the transcription either in a positive or in a negative way [3], [4], [5], [6].
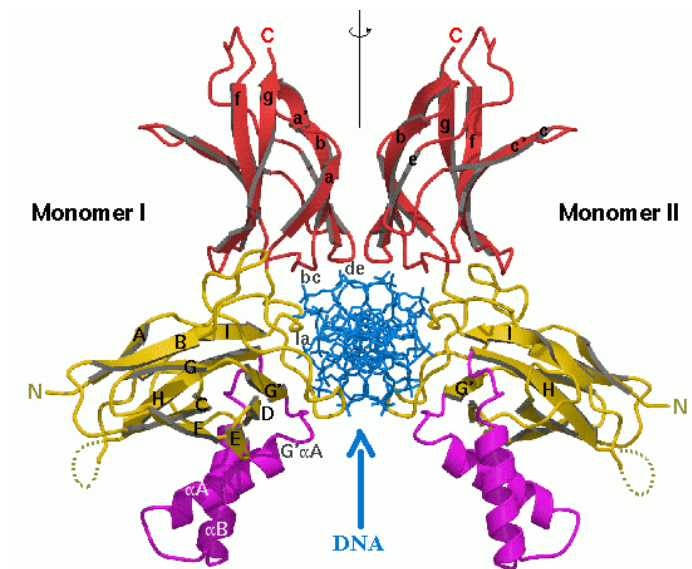
Figure 1 A general schematic representation of a transcription factor and how it binds to DNA [7].

NF-κB activation is mediatory for more than 180 genes transcriptions [8]. NF-κB regulates many of the genes that are mediatory for carcinogenesis and metastasis. In addition to this, chemotherapy and radiotherapy are inducers for NF-κB. NF-κB inhibition prevents tumor resistance to chemotherapeutic agents and increases the cytotoxic effects of the anticancer agents. Moreover, chronic inflammatory diseases can also promote cancer. NF-κB has a central role in promoting inflammation-associated cancer [9].
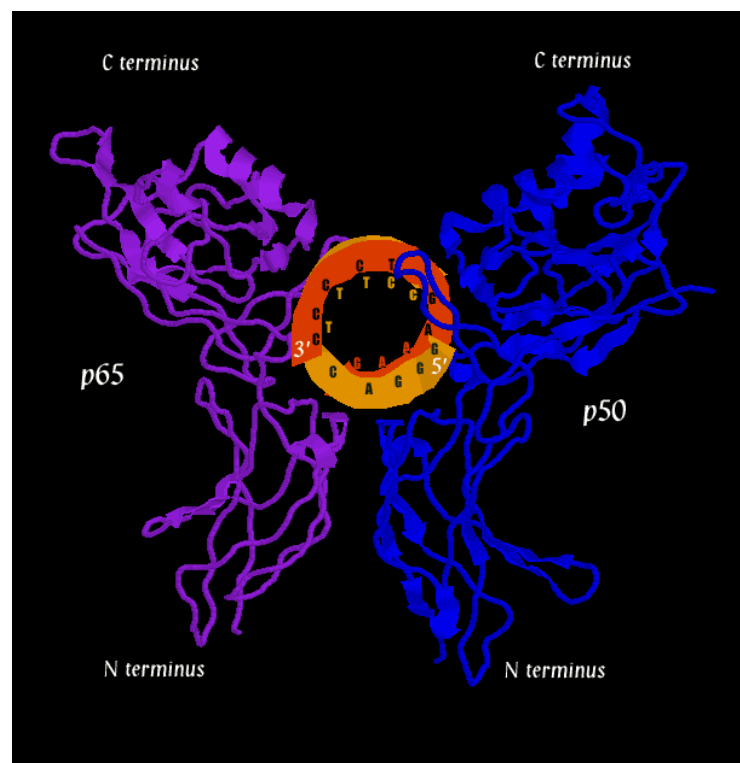
Figure 2 Representation of NF-κB (p65p50 heterodimer) [10].

NF-κB subunits have a conserved Rel homology with 300 amino acids that is required in DNA binding and so regulates transcription of many proteins. Subunits are classified into two main groups. First group of subunits contains NF-κB1 (p105/p50), NF-κB2 (p100/p52). The second group of subunits contains RelA (p65), RelB and c-Rel [11]. There are different combinations of subunits in the cytoplasm. The most common

combination is the heterodimer containing NF-κB1 subunit (p105/p50), RelA subunit (p65) and IκB-α inhibitory subunit [11]. P50-p65 heterodimer and p50 homodimer are crystallized with DNA and IκB-α; 3D structures are in Protein Data Bank (PDB). Encodings of protein structures of interest in PDB are: p50-p65 heterodimer bound to IκB-α as 1ikn, p50-p50 homodimer bound to DNA as 1nfk. Although our studies are focused on inhibitors targeted to the p50-p65 heterodimer computational studies on the homodimer p50-p50 are also performed.

1.1.1   NF-κB and Cancer

In the body, cancer appears when a group of cells stops functioning and starts to propagate uncontrollably. There are seven groups of procancer events that are identified as targets for anticancer strategies. The strategies for inhibition of cancer are: Reducing genetic instability, inhibiting abnormal expression of genes, inhibiting abnormal signal transduction, encouraging normal cell-to-cell communication, inhibiting tumor angiogenesis (generation of new blood vessels around tumor), inhibiting invasion and metastasis, increasing the immune response. In this thesis inhibiting abnormal expression of genes is chosen as the strategy. One way that gene expression can be normalized is through modifying the activity of the transcription factors. Transcription factors such as p53, AP-1(Activator Protein 1) and NF-κB play vital role in survival of cancer cells. Cancer cells produce insufficient amounts of p53 and excessive amounts of NF-κB and AP-1. In previous studies tumor regression is achieved in-vivo on mouse animal model with NF-κB inhibition [12]. Inhibiting NF-κB activity reduces the cell growth rate as indicated in Duffy D. C. et al. [13], [14]. Hence, the transcription factor NF-κB is chosen as the target of our studies against cancer. As mentioned above many genes are regulated by NF-κB : tumor necrosis factor (TNF), collagenases (collagen degrading enzymes), cell adhesion molecules and immunostimulating (increasing the response of immune system against deseases) proteins such as Interleukins. Having

excessive amounts of NF-κB in the cell provides proliferative and survival advantage to the cell itself. In other words activation of NF-κB can suppress apoptosis, by this way promoting chemo resistance and tumorogenesis [15].

1.1.2   NF-κB Inhibitors

An inactive state of NF-κB is present in the cytoplasm in a heterotrimeric form made up of p65, p50 and IκB-α monomers. IκB-α is phosphorylated and then degraded; afterwards p50-p65 heterodimer is transferred to the nucleus. After translocation to the nucleus the heterodimer binds to the promoter region of the DNA and activates the mediatory genes. Most of those mediatory genes are related to cell survival, cell adhesion, inflammation, differentiation and growth [15].

According to NF-κB/IκB-α crystal structure, the nuclear localization signal (NLS) polypeptide on p65, which triggers translocation of NF-κB from cytoplasm to nucleus, is effective on increasing the binding affinity of NF-κB's IκB-α. But the NLS polypeptide alone is not sufficient for NF-κB-IκB-α binding. Although Asparagine (202nd residue) and Serine (203rd residue) residues on p65 are not on the IkB-α binding site, they have a major role in IkB-α binding affinity. This is proven by mutation analysis. Mutation on 202nd residue of p65 Asparagine to Arginine caused a nine fold decrease on IkB-α binding affinity [16].

According to the data given above, an allosteric drug targeting approach is applied on our studies targeted on NF-κB and IκB-α binding. In docking studies defining the binding site as Asparagine 202nd and Serine 203rd centered on NF-κB complex will allow us to design allosteric inhibitors. Indeed nature uses allosteric interactions for regulations of transcription and protein activities. Conventional drug development is focused on direct active site inhibition, whereas allosteric drugs are a new class like p53 and nitric oxide synthase studies [17].

On the other hand, conventional approach on NF-κB activation is also studied as positive controls. Inhibitors that are previously known by their affinity to NF-κB specific binding site are the positive controls for our docking experiments. One of them inhibits binding of NF-κB on the DNA, known as NF-κB activation inhibitors; an enzyme inhibitor Aurine Tricarboxylic Acid (ATA), polyhydroxylate derivatives of phenolic compounds Gallic Acid (GA) and 5,7-dihydroxy-4-methylcuomarin (5,7 DHMC) are both experimentally and computationally studied by Pande et al.[18], Sharma R. K. 2000[19]. Dockings in the same systems as in Pande et al. [18, 19] are performed. The inhibitors are specified as inhibitors on NF-κB DNA binding, the DNA binding Region (DBR) on p50 subunit of the complex. DBR consists of 60-68th residues on crystallographic structure from PDB namely 1nfk. Docking results for three methods that we have used in our studies are applied to the same system explained in Pande et al.'s paper as described above. According to the results listed in the table in Appendix-3, ATA has the highest affinity to NF-κB DNA binding site in two of the three methods among three compounds. In addition to this, the results show that three of the molecules have affinity to NF-κB DNA binding site according to each three methods used. In conclusion, we have positive results for the tools used. Evaluations of the other molecules are carried out according to the data reached from Appendix-3.

Another approach of inhibiting NF-κB activation is inhibiting the translocation of the complex to the nucleus. Nuclear Localization signal (NLS) residues on NF-κB complexes play a key role. NF-κB has two NLS, one on p50 (sequence: VQRKRQKLM) and one on p65 (289-320th residues) protein. NLS on p65 is predicted to be bound as IκB-α and NF-κB bind, meaning that inhibiting NLS on p50 is not enough to stop the translocation of the complex to the nucleus [20]. Two inhibitors are reported for NLS of NF-κB: (i) a synthetic peptide, SN50, containing a hydrophobic membrane-translocating region and NLS of NF-κB p50 [11] and (ii) an aromatic diamine, 4-Methyl-N (3-phenyl-propyl)-benzene-1,2-diamine (JSH53) [21].

As mentioned in the previous paragraphs, one of the two main mechanisms to inhibit NF-κB activation is to reduce the affinity of NF-κB to DNA and thus decrease

the DNA binding. The other method is to inhibit the translocation of NF-κB to the nucleus. Drugs of the first strategy target NF-κB (RelA and p50 homodimer) itself. The inhibitors for NF-κB DNA binding are ATA, GA and 5,7 DHMC in addition to these in the literature (E)-3-(4-methylphenylsulfonyl)-2-propenenitrile (BAY 11-7082) , Ritonavr (HIV-1 protease inhibitor) , and Genistein (Isoflavone; Dietary agent) are also reported [10, 22, 23, 24].

However, many of the NF-κB inhibitor drugs are targeting IKK in order to inhibit the activation of NF-κB; there are IKK-independent pathways that trigger NF-κB activation [25]. For instance proteasome inhibitor is one of those.

Except IκB-degradation, there is another way to prevent NF-κB translocation to the nucleus. This inhibition method is performed by blocking the nuclear translocation factors to bind on NF-κB. Although SN50, a cell permeable peptide, does block the nuclear translocation of NF-κB [26] [27], does not inhibit IκB phosphorylation. SN50 has a part for membrane translocation and another part for nuclear localization signal NLS of NF-κB. JSH-23 an aromatic amide also inhibits NF-κB activity in the same manner as SN50 does [21].

Since IκB-α masks just one NLS of NF-κB on p65, IκB-α inhibition is not enough to keep NF-κB in cytoplasm, as described in previous papers. But IκB-β masks both NLS on NF-κB. The aim of inhibiting NF-κB activation by IκB-α degradation is not sufficient for a complete inhibition. Though IκB-α is not enough to prevent the translocation of NF-κB to the nucleus. In conclusion, a drug with a direct inhibition of NF-κB is more desirable [9] as in our focus of study.

1.2   Cancer

Healthy cells do communicate with each other and die when needed for the organism. Apoptosis is programmed cell death. This process is never processed properly by cancer cells. Thus, cancer cells need to be killed by endogenous forces such as special compounds called anticancer drugs. However, while killing the cells using drugs, it is hard to keep the healthy cells alive during the therapy. This is the challenge, because cancer cells have no difference in function, they are different just in activity. So the best way to inhibit cancer cells is not to destroy their morphology but to normalize the signaling inside the cell. Genetic instability, abnormal expression of genes, abnormal signal transduction and abnormal communication with neighboring cells are responsible for the bad signaling in cancer cells [13].

1.3   Drugs against Cancer

The three main classes of drugs used in the drug targeting strategies of cancer therapy are (i) immune stimulants, (ii) indirect-acting compounds, and (iii) direct-acting compounds [13].   Besides these, there are some challenges that are needed to be accomplished during drug design against cancer. They are described as follows:

A single "cure" has not been found for cancer, since there are 100 different types of cancer. Even after some suppression of cancer cells by some drugs, cells may develop resistance to them. As a result of these, chemotherapy is achieved by several drug combinations for different periods of time.

In addition to this, there are a few biochemical differences between cancerous cell and normal cell. Due to those biochemical similarities cancer drugs are limited by their toxicity to normal cells. Chemotherapeutic drugs on the market are not specific; consequently severe side effects are associated with the chemotherapy. For instance, as a result of chemotherapy targeting decrease in the rate of growth of the cancer cells, the side effects can be easily seen on the somatic tissues that naturally have a rapid turn

over like skin, hair, gastrointestinal and bone marrow. These healthy cells also end up with the damage of chemotherapy [28].

Some well known examples of anticancer drugs are: (i) directly DNA damaging agents in nucleus such as cisplatin, (ii) the antibiotics- daunoribicin, doxorubicin, and etoposide, (iii) mitotic spindle synthesis inhibitors such as vinblastine, vincristine, pacitaxel (Taxol), and (iv) pre-DNA building block synthesis inhibitors such as methatrecate, fluorouracil, hydroxyurea, and mercaptopurine [28].

1.4   Natural Products against Cancer

Traditional medicine is used mostly for inflammatory diseases and cancer. Plant derived molecules are nature's treasure box and are also candidate lead compounds for modern drug design. Aspirin and dioxin are examples of modern drugs originating from plant material [29].

The aim of choosing plant derived products is to decrease the severe effects of the chemotherapy. Natural products have less adverse effects than chemotherapy drugs [13]. As described in the literature, natural compounds on the average are 21-fold less toxic than most chemotherapy drugs.

1.5   NF-κB drug candidate Dataset Used in this study

Since NF-κB is described as a protein complex that is capable of binding to the enhancer sequence of immunoglobulin's heavy chain and kappa light chain by Sen and Baltimore, this protein is a drug target for inflammatory diseases and cancer in drug design as described above [30]. Natural products against NF-κB are the focus of this study. Plant secondary metabolites that were already tested for their anticancer properties are chosen from literature. A small group of screening data set against NF-κB pathway is prepared. The compounds screened in our data set are classified into 27

major structural group of molecules; alkaloids, benzenoids, amides, benzofurans, cardiac glycosides, coumarins, cyclopenta benzofuran, diarylheptanoids, diterpenoids, fatty acids, flavonoids, harmol alkaloids, lignans, monoterpenes, naphtopyran, notwithanolides, rotenoids, phenylphenalones, pramanicin, sesquiterpenes, simaroubolides, stilbenoids, triterpenes, withanolides, β-Ionone, wide group of cycloartene triterpenoids, a wide group of resveratrol derivatives, other compounds (NF-κB inhibitors.) [31][32][33].

Positive control of the data set consists of three compounds that are previously studied both experimentally and computationally Pande et al.[18], Sharma R. K. 2000 [19]. Overall, the data set consists of 236 compounds.

Chapter 2

COMPUTATIONAL METHODS

In this chapter, the computational methods used in our study are described in detail. Docking methods and scoring functions are explained.

2.1 Introduction (Docking)

In the living cell there are a variety of interactions of molecule and ligand pairs. Predicting the quality and quantity of the interaction pairs is always of interest because prediction sheds light on many metabolic pathways and mechanism of diseases. Hence, predicting target interactions boils down to novel approaches to drug design. Predicting the binding orientation, conformation and interaction energy is theoretically possible by computational methods [34]. For this purpose, the 3D crystal structure of the molecule of interest is needed. The first crystal structure was myoglobin performed at 1958 (Kendrew, J.C. et al). The structure was solved by X-ray analysis. 3D structures of proteins are predicted by two methods: (i) X-ray crystallography and (ii) NMR (Nuclear Magnetic Resonance). The collection of structural information of proteins triggered research on novel computational methods. These 3D structures have been used for computational interaction research since 1980s starting with QSAR (Quantitative Structure-Activity Relationship) studies. Structure based functions are developed over the years and docking is one of the tools for implementing the computational technology. In docking analysis, a receptor of known structure and a small molecule (ligand) are searched for a favorable binding. During docking analysis, the translation (location of binding of the ligand on the receptor), the orientation of the ligand and the final conformation of the ligand is optimized according to a given set of assumptions.

In-silico methods save money and time, because screening drug molecules *in-vivo* or *in-vitro* with laboratory work is costly. *In-silico* analysis narrows down the number of candidate drug molecules to be tested in the laboratory. This means efficient use of time and money for industry. The best ligand for a target molecule predicted is referred to as 'a lead drug molecule'. The quality of the lead molecules can be predicted by docking studies.

We mentioned that the assumptions in docking are the parameters of the scoring function and the search algorithm. An ideal scoring function calculates the binding free energy. But none of the scoring functions achieved that goal, because scoring functions make assumptions on some independent variables that affect binding affinity [35]. Nevertheless, scoring functions determine binding modes that are quite similar to the native binding structure. There are many tools and scoring functions of docking. In this thesis GOLD (Genetic Optimization for Ligand Docking) and AutoDock software are used for docking experiments. Scoring functions can be force field based, empirical, knowledge-based, or hybrid of these. Force field methods are based on molecular mechanics force fields which model the interaction between atoms through a combination of bonded and non-bonded two-body interactions. Empirical methods use experimentally calculated data components rather than theoretically derived components. Knowledge-based methods are based on the probability of a certain interaction for a given pair of atoms [36]. In our studies three scoring functions are used: Gold Score, Chem Score and AutoDock. Gold Score is a force field based function. Chem Score is an empirical scoring function [37]. AutoDock is a hybrid scoring function [37]. Free-energy scoring functions of AutoDock uses linear regression analysis for AMBER force field and empirical data. The empirical data consists of a set of diverse protein-ligand complexes with known inhibition constants [3], [38]. Molecular mechanics methods are designed to study bound conformations and free energy perturbation calculations between molecules with single atom change. Empirical methods are needed for rapid ranking binding free energies of compounds that are dissimilar. There are a variety of choices for searching algorithm methods also. For

instance, systematic search, molecular dynamics, simulated annealing, incremental construction, rotamer libraries and Genetic algorithms are used [39]. Simulated annealing and Genetic algorithms are stochastic sampling techniques [40]. The docking tools used here are all genetic algorithm search based methods.

Genetic algorithm search method is a probabilistic search method using a distributed set of samples (individuals) from the space (a population) to generate a new set of samples (offspring) [41]. The fitness function decides whether individuals survive or not and generates offsprings among the survivors for the next iteration (run) of the search.

Docking experiments resulted in many success stories such as nelfinavir; (Viracept) by Pfizer drug against HIV at 1994 [43], amprenavir (Agenerase) from Vertex&GSK; a HIV protease inhibitor at 1995 [44], zanamivir (Relenze) by GSK; Influenza neuraminidase inhibitor at 2000 [45].

The main drawback of docking tools is that none of the scoring functions result in the exact binding affinity compared to binding affinity measured in the laboratory. The native binding mode of a ligand with a receptor may not be ranked as the first one according to the scoring function [5]. Hence, a detailed evaluation of the lead compounds is needed after docking.

2.2 Tools

2.2.1   AutoDock

AutoDock is a docking tool that processes ligands according to a set of grids obtained from the properties of the protein. AutoDock is the first model for modeling the ligand with full conformational flexibility. The first version was written in Arthur J. Olson's laboratory in 1990. AutoDock uses GA as a global optimizer combined with energy minimization as a local search method (Morris et al. 1998) [38]. LGA

(Lamarckian Genetic Algorithm) algorithm, a combination of Global search and local search methods, is used in AutoDock.

In AutoDock 4.0's model free energy of binding equation consists of molecular mechanics-based and empirical terms multiplied with weight coefficients. These weight coefficients refer to the terms: van der Waals forces, Hydrogen Bonding, coulombic potential, desolvation (Release of water electrostatically bond to a particle in solution), and torsional parameters shown in Equation 2.1 [38]. The first four parameters are fundamental molecular mechanics terms and the other two are entropic terms. Desolvation upon binding and hydrophobic effect term models the solvent entropy changes at solute-solvent interfaces. The torsion term models the restriction of internal rotation and global rotation and translation [3]. Complete equation of AutoDock 4.0 that models free energy of binding is shown in Equation 2.1.

$$\Delta G = \Delta G_{vdw} \sum_{i,j} \left(\frac{A_{i,j}}{r_{ij}^{12}} - \frac{B_{i,j}}{r_{ij}^{6}} \right) + \Delta G_{hbond} \sum_{i,j} \left(\frac{C_{i,j}}{r_{ij}^{12}} - \frac{D_{i,j}}{r_{ij}^{10}} + E_{hbond} \right) \qquad (2.2.1)$$

$$+ \Delta G_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + \Delta G_{tor} N_{tor} + \Delta G_{sol} \sum_{i_c,j} S_i V_j e^{\left(-r_{ij}^2 / 2\sigma^2 \right)}$$

Calculations for binding free energy are done in four portions. First, the final intermolecular energy is calculated as a sum of van der Waals, hydrogen bond, desolvation and electrostatic energies. The other three components are: final total internal energy of the ligand, torsional free energy of the ligand and the energy of the unbound system.

2.2.2   GOLD

GOLD is an automated docking optimization tool that uses Genetic Algorithm [3], [4]. The scoring function that ranks the binding modes consists of terms of hydrogen bonding, pairwise dispersion potential that is able to describe a significant contribution

to hydrophobic energy of binding and a molecular mechanics term of internal energy of the ligand [50]. In this thesis, the fitness functions used in GOLD are Gold Score and Chem Score.

The Gold Score fitness function is force field based and it consists of four components: Protein-ligand hydrogen bond energy, Protein-ligand van der Waals energy, ligand internal van der Waals energy and ligand intramolecular hydrogen bond energy. The equation for Gold Score fitness score is shown in Equation 2.2 with ext (protein-ligand energy), int(ligand internal term) terms indicated. Score (int) includes both ligand van der Waals energy terms and ligand torsional terms.

$$GoldScore = Score(hb\_ext) + Score(vdw\_ext) + Score(hb\_int) + Score(int) \quad (2.2)$$

The Chem Score fitness function is empirical and calculates the free energy of ligand binding. Chem Score consists of contact terms as follows: lipophilic and metal-ligand binding contributions, hydrogen bonding, internal torsion terms, covalent and constraint scores as shown in Equation 2.3 [50]. The complete equation for binding free energy calculated by Chem Score is shown in Equation 2.4.

$$ChemScore = \Delta G_{bingding} + P_{clash} + c_{internal}P_{internal} + \left(c_{covalent}P_{covalent} + P_{constraint}\right) \quad (2.3)$$

$$\Delta G_{binding} = \Delta G_0 + \Delta G_{hbond} + \Delta G_{metal} + \Delta G_{lipo} + \Delta G_{rot} \quad (2.2.4)$$

Parameters in the equations above consist of binding free energy change, protein-ligand clash penalty, protein-ligand hydrogen bond contribution value, internal ligand torsional penalty, covalent contribution value and additional constraint contribution. Chem Score hydrogen bond contribution is calculated over all donor-acceptor pairs of protein and ligand. Protein-ligand lipophilic contribution is done by Chem Score lipo term.

2.3 Docking Preparation and Parameters

The data set of ligand 3D structures are drawn in ChemBio3D Ultra 11.0 and presented in Chapter 4 and 5. Energetically the relaxed conformation is needed for the ligand. The geometric minimization and energy minimization with MM molecular dynamics module of ChemBio3D Ultra 11.0 are performed on each ligand.

In AutoDock, the input ligand and protein structures are kept in pdbqt files that include the coordinates of the molecule and also the newly applied Gastieger charges (electrostatic model of each molecule [51] and just polar hydrogen demonstrated on the each atom). Subsequent process performed by Auto Grid builds the model representing the protein structure as in grids. AutoDock is performed with the LGA (Lamarckian Genetic algorithm) with the following parameters: population size of 250, number of evaluations 2 500 000, number of generations of 50 000, mutation rate of 0.02, crossover rate of 0.8. Optimization is preformed for 100 runs.

In GOLD, ligand and protein input files are mol2 files. The compounds are changed into GOLD specific atom type format. After preparation of ligand and protein mol2 files, the next step is to run the docking optimization program with the parameters inserted by the user. The parameters are as follows: population size of 100, selection pressure of 1.1, number of islands of 5, niche size of 2, crosswt of 95, allele mutatewt of 95, migratewt of 10. The optimizations are performed for 10 runs. All three tools are performed for optimizing each ligand from the data set to NF-κB protein with the specified binding site centered on Arg202- Ser203 residues with a sphere of 20 Å.
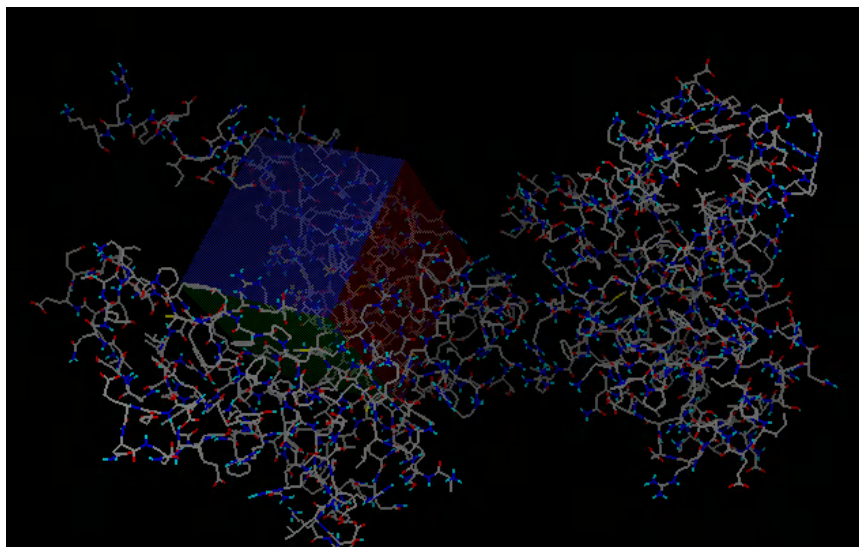
Figure 2.1 NF-κB binding site represented in 3D box in AutoDock tools 3.0.

In AutoDock the binding site is defined as a box in space shown in Figure 2.1. In GOLD, binding site is defined as a sphere in space with the same boundary residues.

Scheme of Work Flow

Dataset Preparation

(Small molecule structures with all hydrogen,

Minimization with Chem Draw 3D v.08)

↓

Protein 3D structure from PDB

(All hydrogen in GOLD, polar hydrogen only in AutoDock)

↓

Automated Docking input file preparation

↓

Docking

↓

Analysis of output files

↓

Choice of lead molecules

↓

Lead Optimization
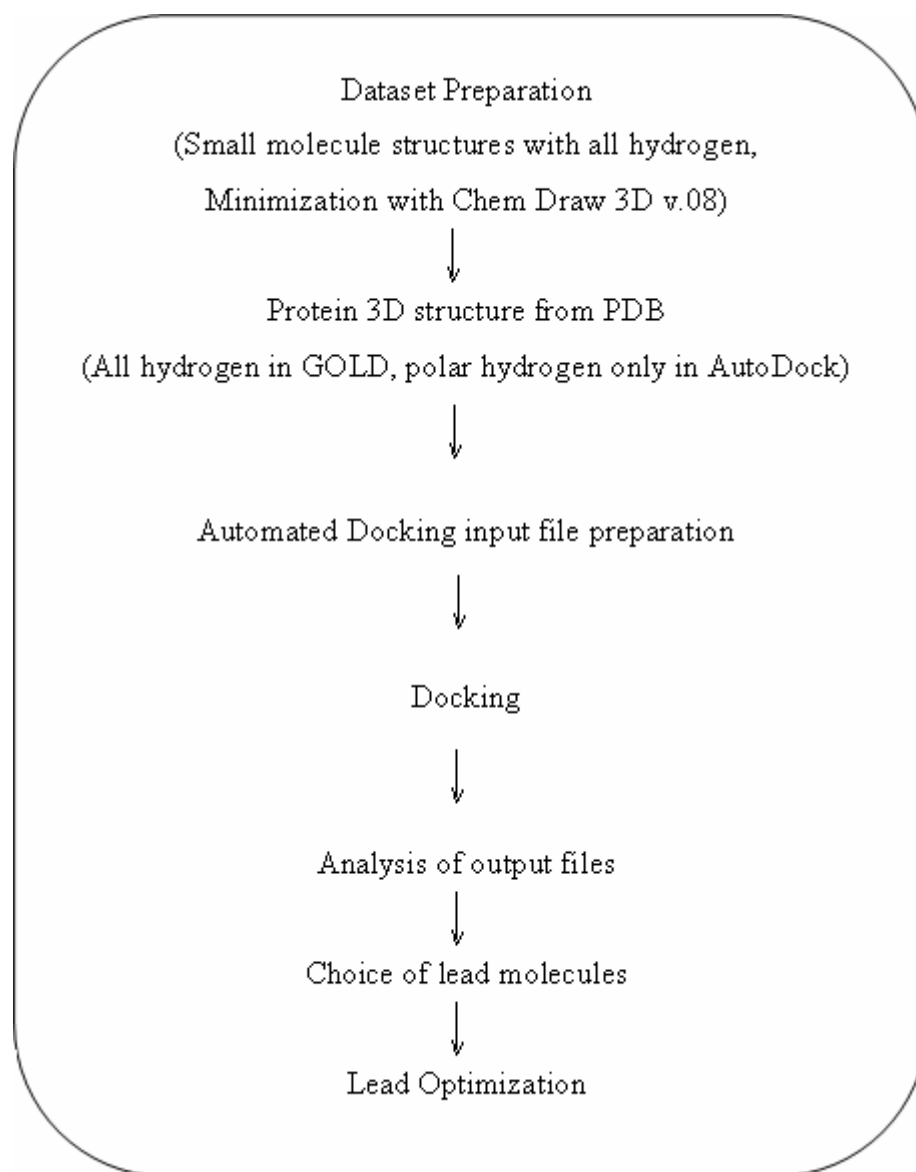
2.3.1 Scheme of work flow for drug discovery**.**

Chapter 3

DRUG DATA SET

In this chapter compounds in data set are observed in detail. Major structural groups in the data set are described. Appendix 4 is supplementary for this chapter.

3.1 The Drug Data Set

As mentioned in the introduction chapter, natural products that are used in this study can be classified into 27 major structural group of molecules; alkaloids, benzenoids, cardiac glycosides, diarylheptanoids, diterpenoids, flavonoids, lignans, notwithanolides, rotenoids, sesquiterpene lactones, stilbenoids, triterpenes, withanolides, amides, benzofurans, coumarins, cyclopenta benzofuran, fatty acids, monoterpenes, naphtopyran, phenylphenalones, sesquiterpenes, simaroubolides, stilbenolignan, wide group of cycloartene triterpenoids, harmol derivatives, a wide group of resveratrol derivatives and other molecules active to Nfκβ. In this chapter all this compounds are observed in detail. Properties of the compounds and IUPAC name are generated via MarvinSketch, and detailed list can be seen in Appendix-4.

3.2 Natural Products

Organic compounds that are produced by plants are classified into two groups: primary and secondary metabolites. In contrast to primary metabolites, secondary metabolites are not essential in plant metabolism. Secondary metabolites are often referred as natural products and are used as drugs against several diseases. While primary metabolites are involved in nutrition of plant, secondary metabolites influence ecological interactions between plant and its environment. Ecologically, natural products are competitive among plants, protective against herbivores and microbial infection, attractive for pollinators and seed-dispersing animals. This phytochemicals

are of interest in several industries like dyes, polymers, glues, oils, fibers, waxes, flavoring agents, perfumes, insecticide, herbicides, antibiotics and several other drugs.

Plant natural products are divided into three major groups: terpenoids, alkaloids, phenolic compounds. Terpenoids are derived from the 5-carbon precursor isopentenyl diphosphate (IPP). Terpenoids come in a variety more than 25 000 compounds identified as secondary metabolites. Alkaloids are nitrogen containing compounds whose biosynthesis is derived from amino acids. 12 000 alkaloids are specified from plants. Phenolic compounds are biosynthesized in shikimic acid pathway or malonate/acetate pathway. 8 000 phenolic compounds are specified from plant.

3.2.1 Terpenoids

Terpenoids are derived by repetitive fusion of branched 5-carbon units based on the isopentane skeleton. Terpen decomposition yields to isoprene gas.
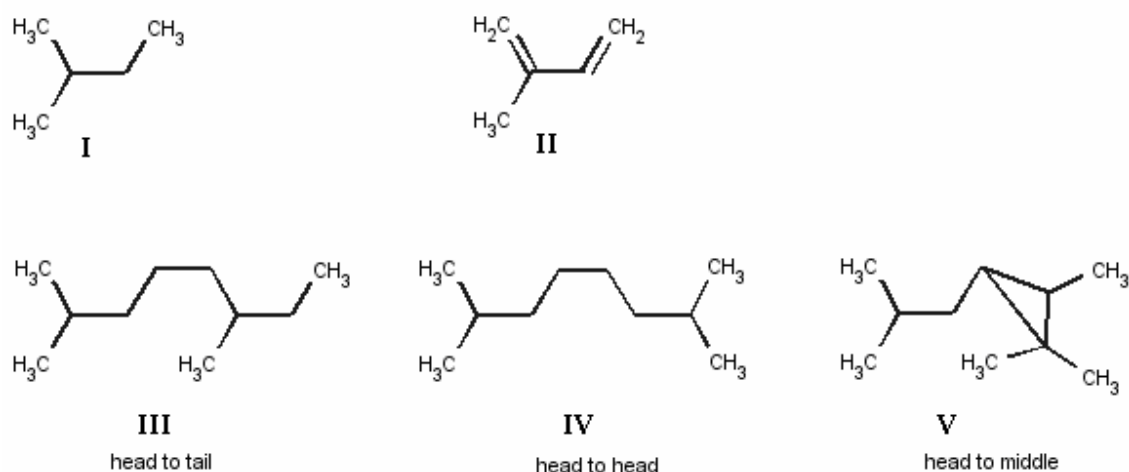
Figure 3.1 Isopentane molecule shown as structure I, Isoprene gas is shown as structure II, Structures III, IV and V are three different structures that two isopentane units form a monoterpene.

Isoprene units and their elongation types are shown in Figure 3.1. The isoterpene unit is classified as hemiterpene with its five-carbon structure. Terpenoids are named as; monoterpenes (10-carbon structure (C_{10}); two isoprene units), sesquiterpenes(15-carbon structure), diterpenes (C_{20}), norditerpenes (C_{19}), triterpenes (C_{30}), tetraterpenes (C_{40}), polyterpenes (structurally containing more than 8 isoprene units), meroterpenoids (natural products with mixed biosynthetic origins that contain terpenoid derived units). For instance, flavor and perfume giving essential oils, antibiotics, giberellin hormones, anticancer agents like taxol, toxins, long polymers like rubber and cytokinins (plant growth substances) are terpenoid structure in nature.
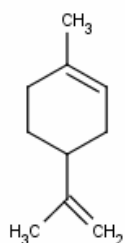
Figure 3.2 Limonene, a cyclic monoterpene that is an essential oil in citrus gives the main odor.

3.2.2 Alkaloids

Alkaloids are nitrogen containing chemical defense systems of plants. 20 % of flavoring plants produce alkaloids. Strictosidine is the precursor of many alkaloids. Strictosidine is synthesized from L-tryptophan derived Tryptamine and secolognine.
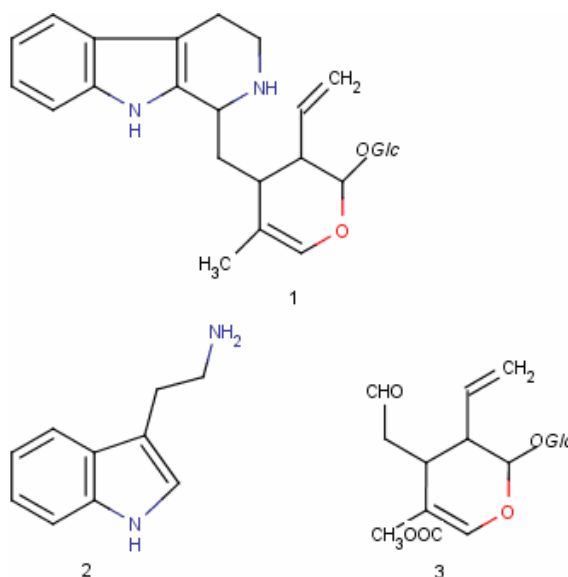
Figure 3.3 1-Strictosidine, precursor of many alkoloids, 2-Tryptomine derived from L-tryptophan, 3-Secolognine.

Ecochemically alkaloids are divided into two groups (i) pyrolizidine for the toxic compounds from roots and (ii) quinolizidine for insecticides from leafs. Examples of alkoloids are a variety of compounds such as: Nicotine a pyridine alkaloid, quinine derived from L-tryptophan, active principle of the ergot fungus (agroclavine) an indole alkaloid, Purine alkaloids (caffeine and theobromine) derived from aspartate, glycine and glutamine, camptothecin anticancer agent.
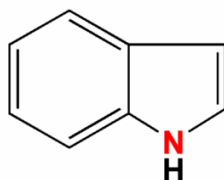
Figure 3.4 Indole alkoloid [52].

3.2.3 Phenolic Compounds

Phenolic compounds derived from phenylproponoid and phenylpropanoid-acetate. Natural products that are aromatic with hydroxyl groups attached.
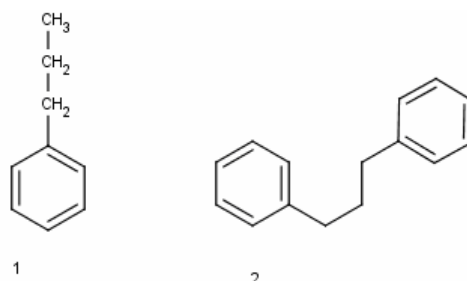
Figure 3.5 Structure 1: Phenylpropanoid (C_6C_3), structure 2: Phenylpropanoid-acetate (C_6C_3-C_6)

Phenolic compounds are involved in defending plants, establishing flower color, flavor components like taste and odor and plant growth. Defending is by protecting against bacterial and fungal pathogens, discouraging herbivores and inhibiting seed germination. Examples are flavonoids, lignans as antioxidants and defend against pathogens, coumarins, furanocoumarins, stilbenes.
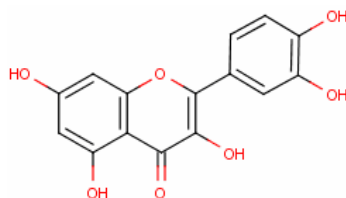
Figure 3.6 Quercetin, a flavonoid (C6C3-C6).

3.3 Data Set of Major Structures

Data set with 236 molecules are described below under 27 major structural groups. Drug properties used for describing molecules below are Logarithmic distribution constant (logD: octanol-water distribution constant, referred as pH dependent partition coefficient for ionizable compounds), accessible surface area in $Å^2$ (ASA), hydrophobic accessible area / hydrophilic accessible area ratio (Pho/phi_ASA ratio), Number of H-bond acceptor, donor sites [53]. Positive values of logD refer to lipophilic structures while negative values of logD are hydrophilic structures [54]. The more lipophilic a drug, the more penetrating capability to membrane it has [55, 53]. Pho/phi_ASA ratio over one means that compound has lipophilic sites mostly and below one means hydrophilicity. Leads with less molecular complexity has less drug like structures this can be refered as less total solvent accessible area [56].

3.3.1 Alkaloids

Nitrogen containing phythochemicals, i.e., the alkaloids in our data set are shown in Figure 3.7.
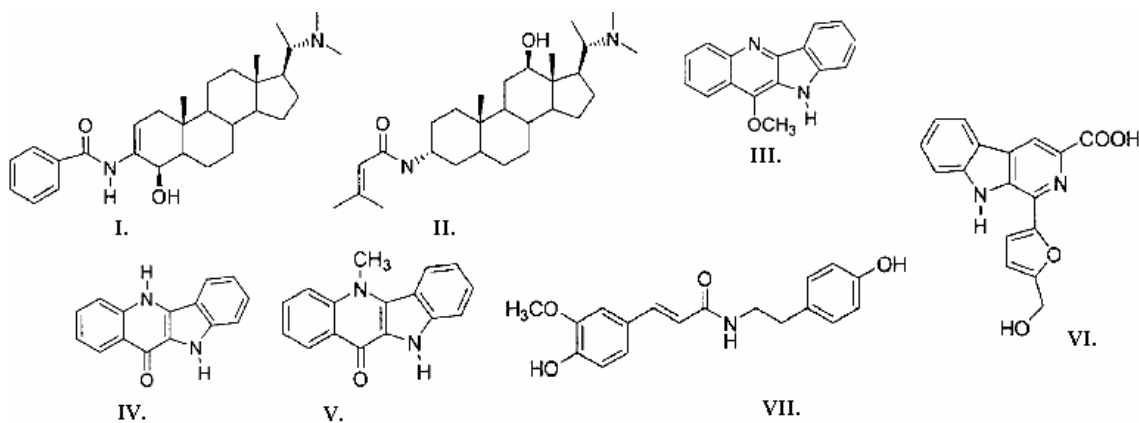
Figure 3.7 Alkaloids in data set are shown above as I: compound m01, II: compound m02, III: compound m50, IV: compound m51, V: compound m150, VI: compound m151 and VII: compoun152.

All alkaloids in the data set are lipophilic with their positive logD values. Most hydrophobic compounds are m02 and m50 with Pho/phi accessible area ratios of 12.89 and 11.95 in sequential order.

3.3.2 Amides

The general amide structure of the amides in data set is shown in Figure 3.8.
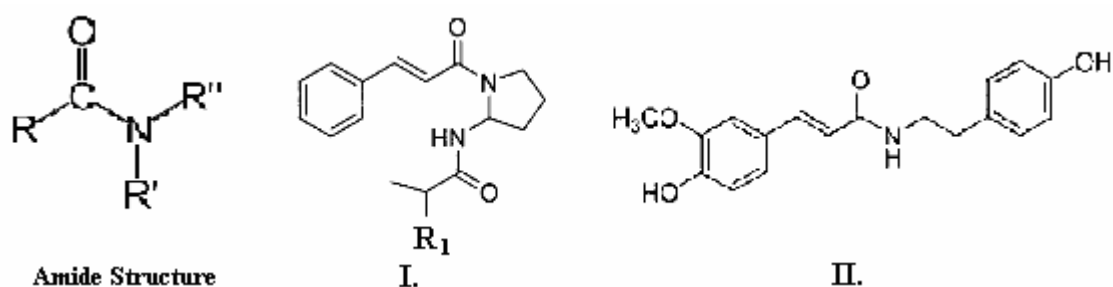
Figure 3.8 The general amide structure is shown on the left hand side. Compound m52 and m53 are represented as I. I with R1:CH2CH3 for m52, I with R1:CH3 for m53, II: m152.

Pho/phi_ASA ratios are high for m52 and m53 with values as 14.69, 13.07 in the same order. Alhough logD values of 3.12, 2.72 can be interpreted as acceptable lipophilicity, they are in contrast with Pho/phi_ASA ratios achieved. Similarly compound m152 is 2.92, which is an acceptable level of lipophilicity. Both compounds have 1 donor, 2 acceptor sites, but m152 has 3 donor, and 4 acceptors. The charged structure of m152 has lower Pho/phi_ASA ratio than other two amide structures in data set.

3.3.3 Benzenoids

Benzene like structures are named as benzenoids. For instance, phenyl and benzyl are substituent groups with benzene ring.
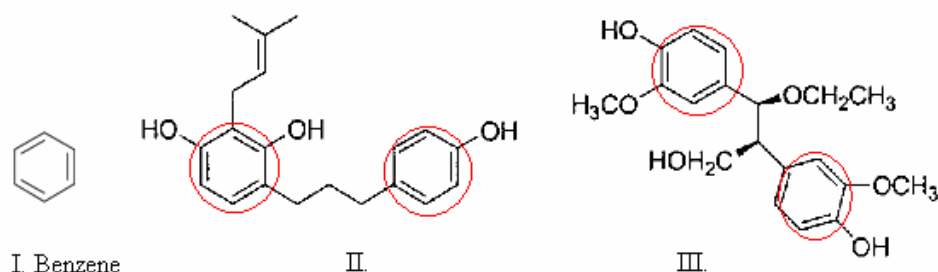
Figure 3.9 Benzene ring shown in I, Benzenoid structures in dataset II: m03, III:m04.

Compound m03 is more lipophilic than m04. The Pho/phi_ASA accessible area ratios of m03 and m04 are 4.96 and 3.39, respectively. LogD values at pH 7.4 are respectively 5.6, 2.15. Numbers of donor and acceptor atoms of the above compounds are the same. Each has 4 donor and 2 acceptor sites.
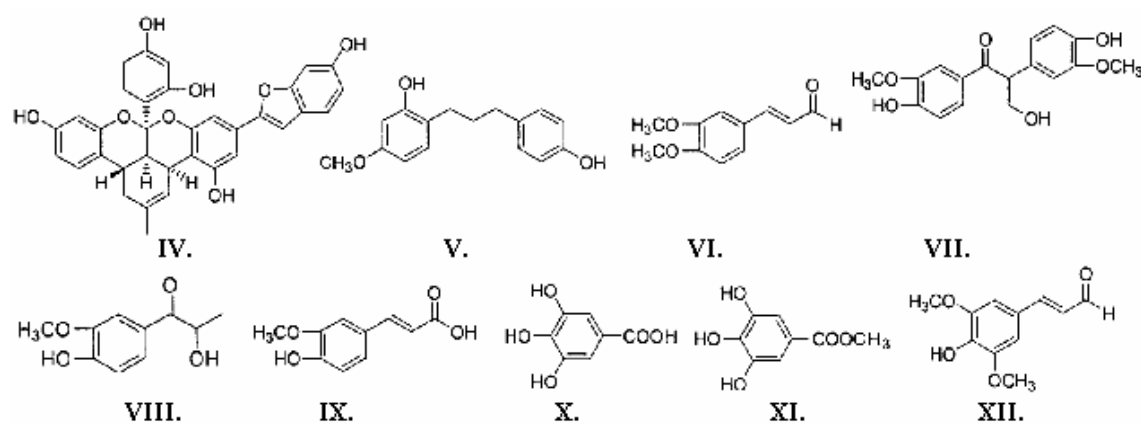
Figure 3.10 Other bezonoids in data set are IV: m54, V: m55, VI: m57, VII: m59, VIII: m58, IX: m60, X: m61, XI: m62, XII: m63.

Two compounds among eleven are hydrophobic and all other nine shown in Figures 3.9 and 3.10 are lipophilic. Compound m61 is the most lipophilic with logD value of -

2.56 and Pho/phi_ASA ratio of 0.53 values. Compound m54, m55, m57 has Pho/phi_ASA ratios greater then four with highly positive logD values. Other than these compounds, in Figure 10 compounds have hydrogen bond acceptor values are varying three to seven.

3.3.4 Benzofurans

Benzofurans are compounds that contain benzene and furan rings.
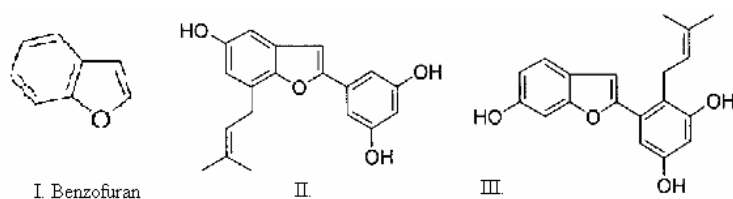
Figure 3.11 Benzofuran structures in data set, II: m05, III: m06.

Both m05 and m06 has the same logD value of 4.42 and same number of donor acceptor sites. But total ASA values and Pho/phi_ASA ratios are different. While m05 has ratio of 2.92 m06 has ratio of 3.26. Compound m06 is more lipophilic than m05.
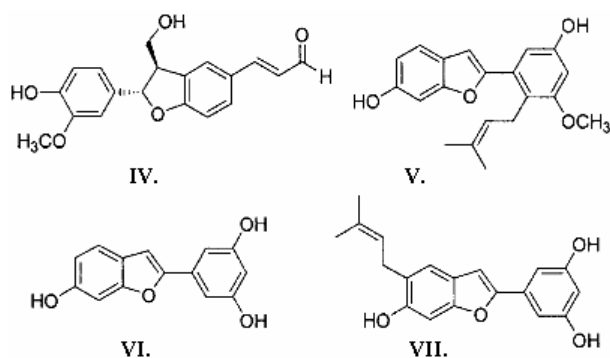
Figure 3.12 Four benzofuran compounds in data set are shown above. IV represents compound m64, V: m65, VI: m66, VII: m67.

All six benzofuran compounds are lipophilic with positive logD values with similar values around three. Compound m64 has the lowest logD values with two hydrogen bond acceptor sites and five acceptor sites.

3.3.5 Cyclopenta Benzofuran

An antifungal compound Flavagline is an example of cyclopenta benzofuran derivatives [57]. The cyclopenta benzofuran structure in our data set is shown in Figure 3.13.
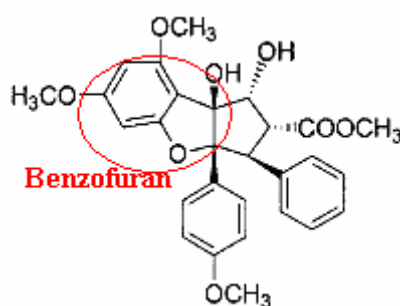
Figure 3.13 Compound m71 is shown.

Compound m71 has Pho/phi solvent accessible surface area ratio of 6.85 and logD 2.19, hence has a lipophilic structure. Compound m71 has 2 donor 7 acceptor sites.

3.3.6 Cardiac glycosides

Compounds act either beneficially or toxically on heart with sugar and non sugar moiety. Both m07 and m08 has the same logD value of 1.62 and same number of donor acceptor sites.
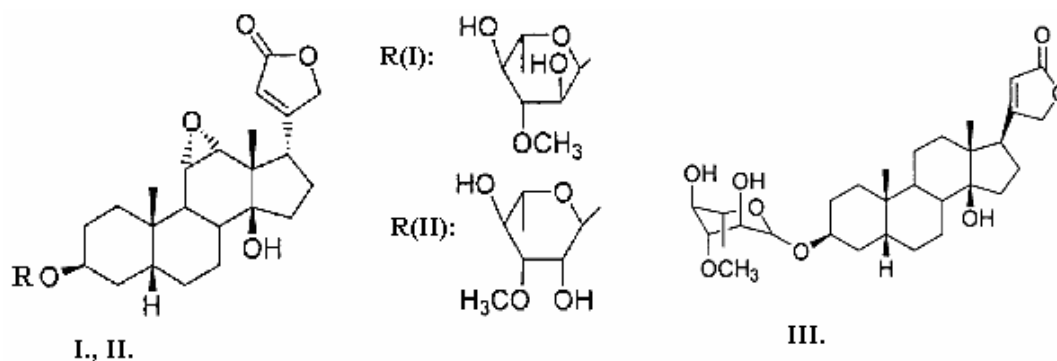
Figure 3.14 Cardiac glycosides in data set are represented as a general structure on the left. R represents the sugar moity of the molecules. I and II represent the sugar moieties of m07 and m08 subsequently. III represents compound m68.

All three compounds are lypophylic. Pho/phi_ASA ratios are also quite similar as 3.57 and 3.82 in descending order while compound m68 has Pho/phi_ASA ratio of 4.14. Even though these compounds have similar properties, their total ASA values varies 621.9 to 575.4 $Å^2$.

3.3.7 Coumarins

Coumarins have blood thinning, antifungicidal, and antitumor activity. Coumarin is present in several plants, for instance in strawberries, apricot, cherries, cinnamon, licorice and lavender [pythochemicals] [58]. Coumarins in our data set are shown in Figures 3.15 and 3.16.
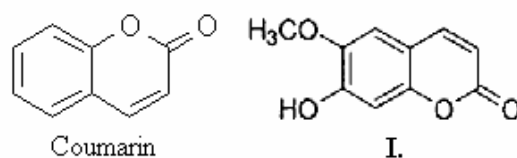
Figure 3.15 General structure of coumarin molecule is represented on the left. Compound m70 is shown with I having carbonyl and hydroxyl functional groups on the main coumarin skeleton.

Compound m70 has more hydrophilic property than the other molecules in the data set with a logD value of 0.98, although its Pho/phi accessible area ratio is 1.97, with a small surface area with one donor and three acceptor sites. As compound DHMC5 also has a hydrophilic structure with 12 hydrogen bonding sites and negative logD value (logD = -0.20)
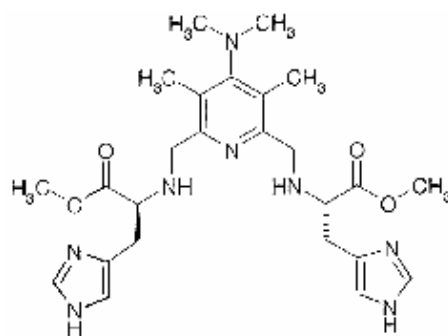
Figure 3.16 5,7-dihydroxy-4-methylcuomarin as DHMC5 in data set is shown above in the group of coumarins.

3.3.8 Diarylheptanoids

Diarylheptanoids consist of two aromatic rings separated by heptanes (seven carbons). Curcumin-like compounds are good examples that have medical use.
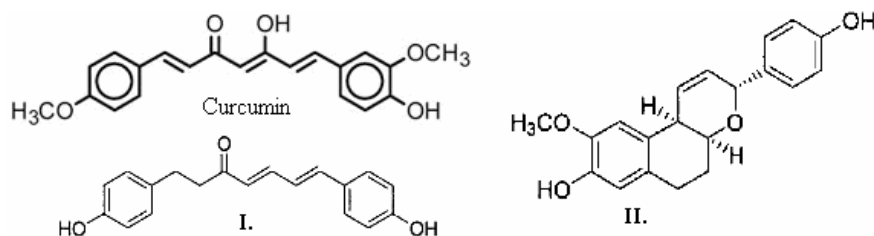
Figure 3.17 Diarylheptanoid example Curcuminis shown with our data set compounds; I referred as compound m72 and II above referred m10, [59].

Both compounds have lipophilic character. Compound m72 has more lipophilic structure than compound m10 with logD values of 5.14 and 3.71 in the same order. Pho/phi_ASA ratios are similar of 4.16 and 4.13. Both compounds have same donor sites and three to four acceptor sites.

3.3.9 Diterpenoids

Diterpene derived molecules in our data set are shown below Figure 3.18.
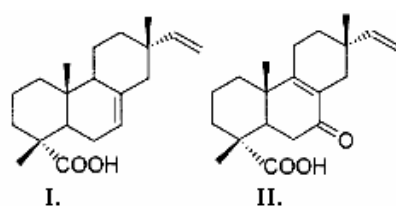
Figure 3.18 Diterpenoids m73 is shown with I and m74 is shown with II.

Both of the compounds are lipophilic with positive logD and high ratio of Pho/phi_ASA as 9.03 and 6.09 in sequential order. The high difference in this ratio is because m74 has one more acceptor site while both having no donor sites with structure of small volume.

3.3.10 Fatty acids

Fatty acid derivatives have a long aliphatic tail with carboxylic acid. For example, caprylic acid used in perfumery is a fatty acid.
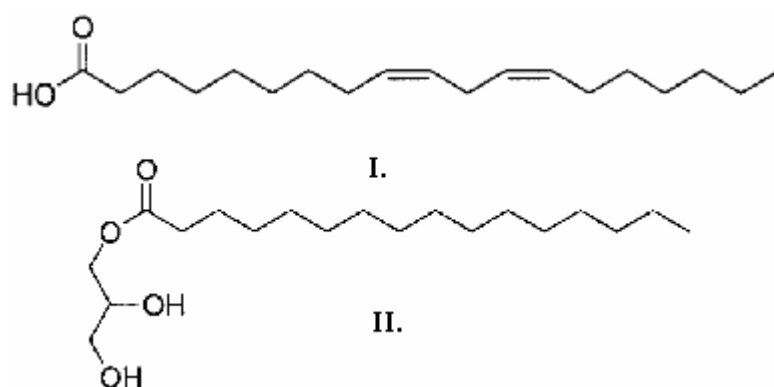
Figure 3.19 Two fatty acids in data set are shown. Compound m75 and m120 are shown as I and II.

Compound m75 is a highly lipophilic compound with Pho/phi_ASA value of 7.87, logD of 3.89. This compound has acceptors sites and no donor site.

3.3.11 Flavonoids

Flavoids have a phenolic structure. An example of flavonoids is their presence in green tee, wine and cacao. Some flavonoids have antioxidant or anticancer effects. They are also used in medication of heart diseases. Figures 3.20, 3.21, 3.22 and 3.23 show related compounds in our data set. Figure 3.20 shows fourteen flavonoids among fifty-three in the data set.
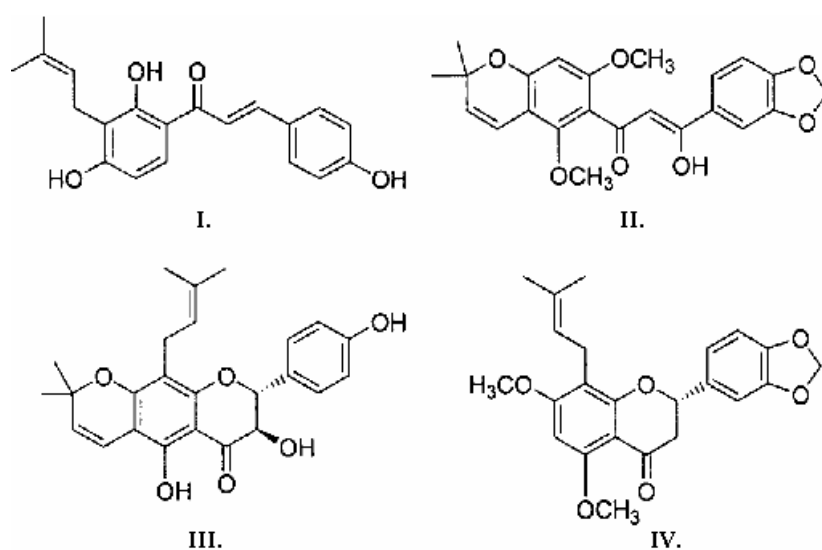
Figure 3.20 Four flavonoid compounds in data set are shown. Compound m96, m105, m100 and m106 are shown as I, II, III and IV in sequential order.
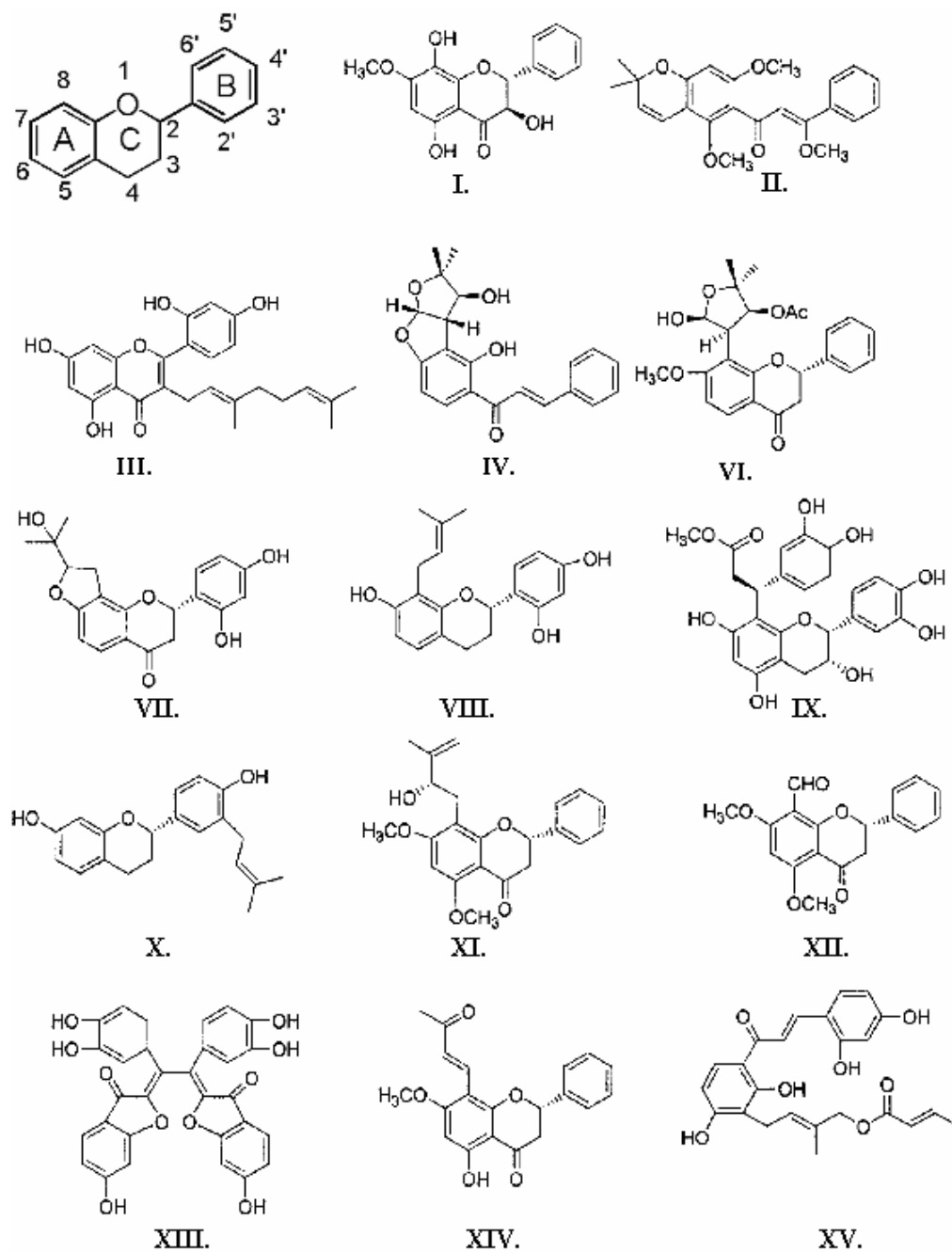
Figure 3.21 Structural representation of a flavonoid is demonstrated on the top left. Modifications are made at the numbered sites shown. I: m22, II: m23, III: m24, IV: m25, VI: m26, VII: m12, VIII: m13, IX: m14, X: m15, XI: m17, XII: m16, XIII: m18, XIV: m19, XV: m21, [60].
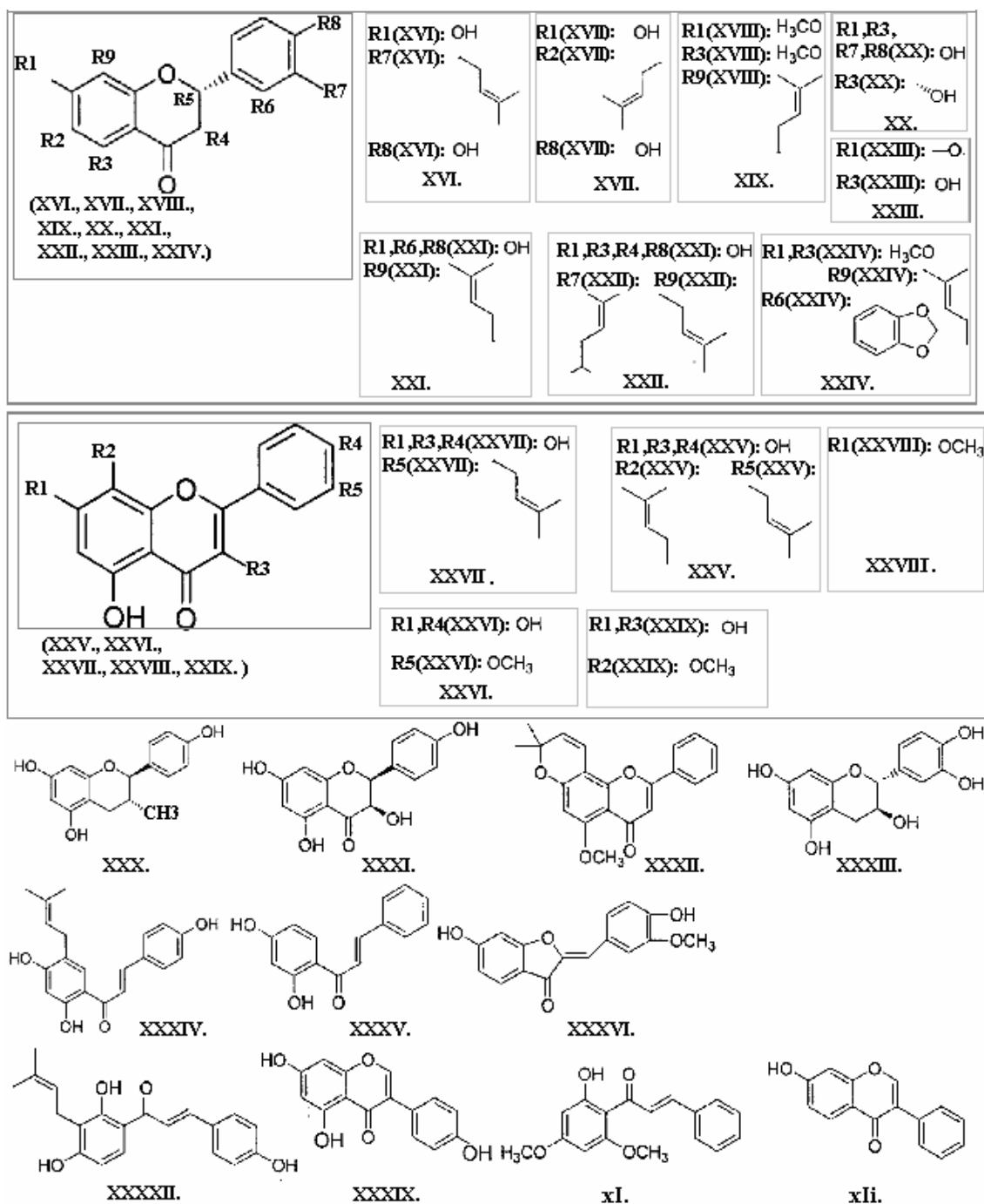
Figure 3.22 Flavanoid compounds in data set. XVI: m76, XVII: m78, XIX: m82, XX: m89, XXI: m90, XXII: m99, XXIV: m106, XXVII: m98, XXV: m80, XXVI:m86, XXVIII: m95, XXIX: m103, XXX: m77, XXXI: m78, XXXII: m81, XXXIII: m83, XXXIV: m79, XXXV: m84, XXXVI: m85, XXXVIII: m88, XXXIX: m92, xI: m93, xIi: m94.
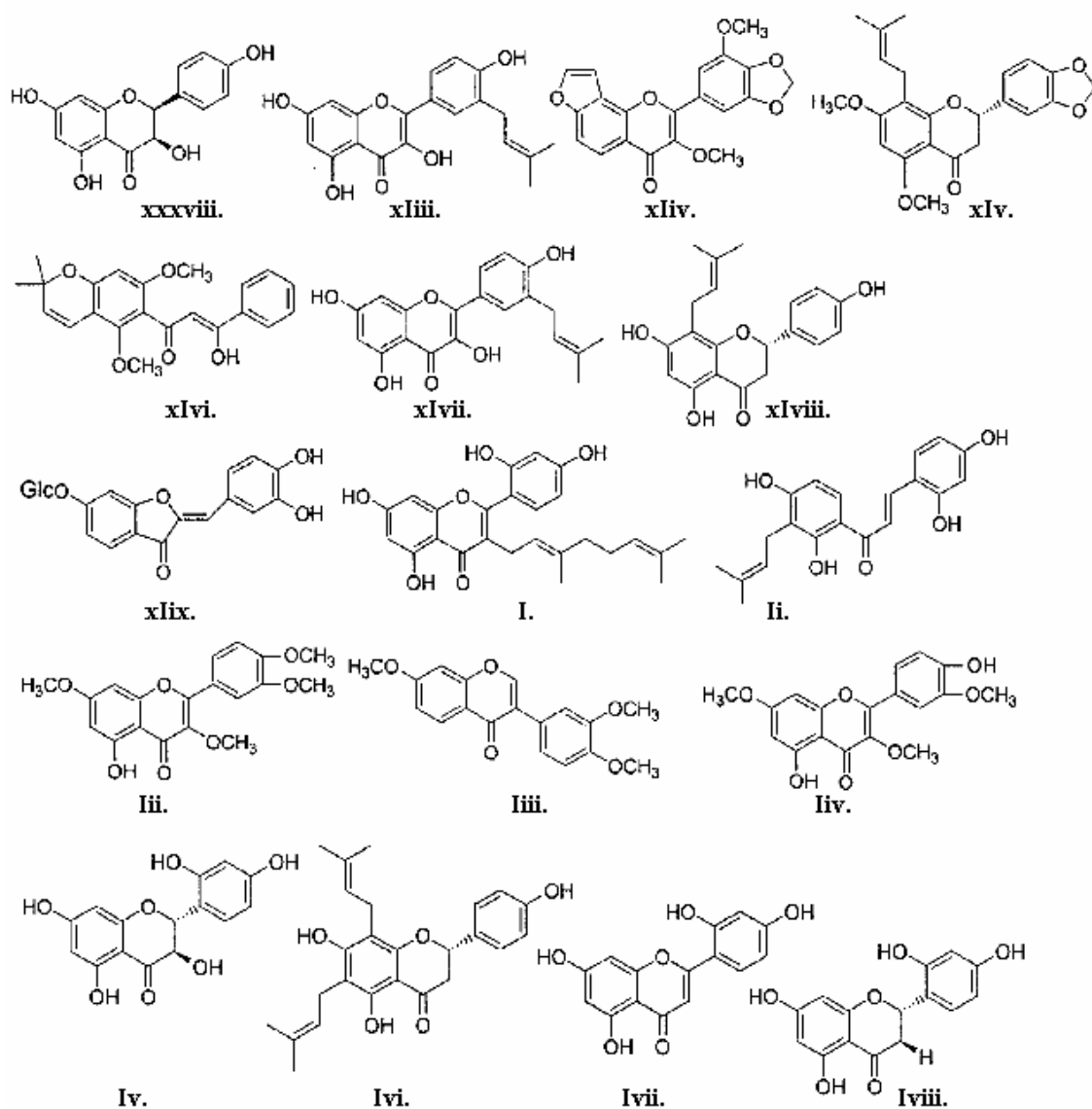
Figure 3.23 Flavonoids in data set. xIiii: m98, xIiv: m102, xIv: m106, xIvi. m107, xIvii: m108, xIviii: m109, xIix: m111, I: m112, Ii: m113, Iii: m114, Iiii: m115, Iiv: m116, Iv: m153, Ivi: m155, Ivii: m156, Iviii: m157.

In Figure 3.21 the most lipophilic structure is m23 with logD value of 10.56 and m13 and m18 are both the most polar in this group of molecules in compounds. Compounds m13 and m18 are with the same logD of 1.54 values. Donor sites vary

from 0 to 7 in 14 compounds. Compound m18 has the highest number of acceptor sites as 10 while m13 has just four. When we observe all 53 compounds in Figure 3.20, 3.21, 3.22 and 3.23, we see that all structures are lipophilic except two molecules. Compound m89 and m103 are hydrophilic with negative logD values -0.63, -0.7 in sequential order. Compound m89 is also containing twelve hydrogen bonding sites which is the highest number of all data set although having Pho/phi_ASA of 1.21. Compound m82 has Pho/phi_ASA ratio of 12.09 and logD value of 3.88 which indicates high lipophilicity.

3.3.12 Lignans

Lignans are present mostly in plant seeds and are referred to as antioxidant effects. They are derived from cinnamyl units. Most Phytoestrogens are lignans with their polyphenollic structures. Enterodiol is one example of phytoesrogens as lignans shown with data set compounds in Figure 3.24.
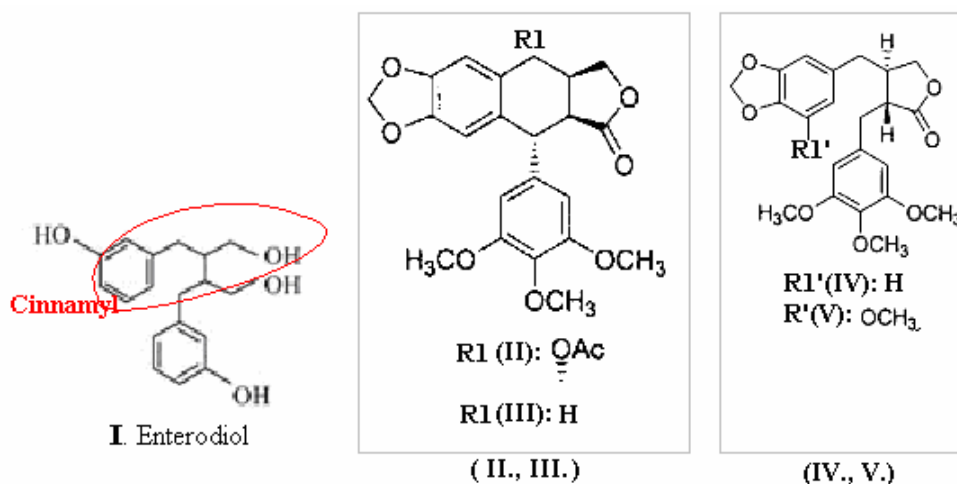
Figure 3.24 Enterodial is an example shown as I. Lignan structures in data set. II: m27, III:m28, IV: m123, V: m125, [61].
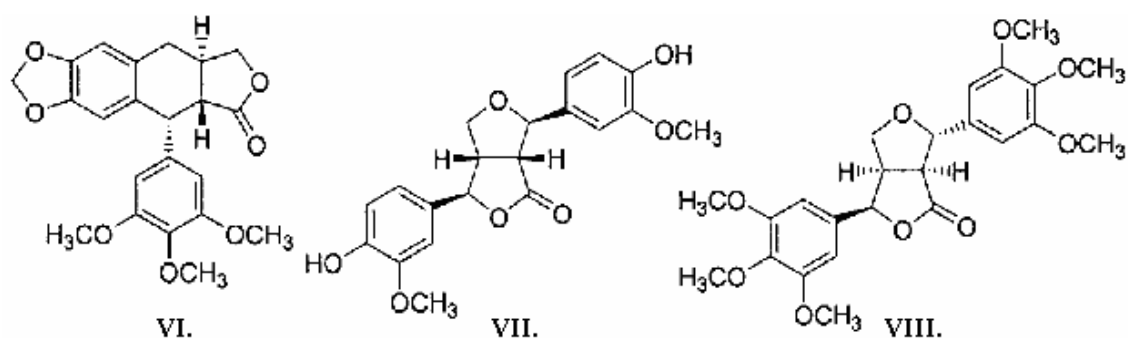
Figure 3.25 Other lignan structures in data set, VI: m122, VII: m124, VIII: m126.

All liganans in the data set shown in Figure 3.25 are lipophilic. Compound m126 has the highest Pho/phi_ASA ratio of 5.71 with eight acceptors. Compound m124 has two donors while the other six compounds have none.

3.3.13 Monoterpenes

Monoterpenes consist of two isoprene units ($C10H16$). They may be in acyclic, monocyclic or bicyclic forms. The compound in our data set is a bicyclic molecule shown in Figure 3.26. A well known example for monoterpenes is menthol; used as topical pain reliever, antipuretic.
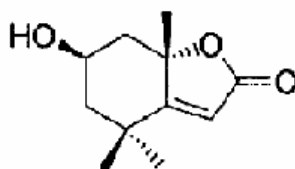
Figure 3.26 Compound m127 in data set.

Pho/phi accessible area ratio shows that m127 has a hydrophobicity of 3.25 and an acceptable degree of LogD with a value of 1.29. One donor and two acceptor sites are present at pH7.4.

3.3.14 Naphthopyran

Examples of naphthopyran derivatives are used in lens and dye industry for their photochromic properties.
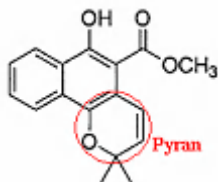
Figure 3.27 Compound m128 from data set.

Compound m128 above has a highly lipophilic structure with Pho/phi_ASA ratio of 5.25 and logD value of 4.31. It has one donor and three acceptor sites at pH 7.4.

3.3.15 Norwithanolides

Norwithanolides are withanolides containing 27 carbons [62]. Seven different compounds in our data set are norwithanolides as shown Figure 3.28.
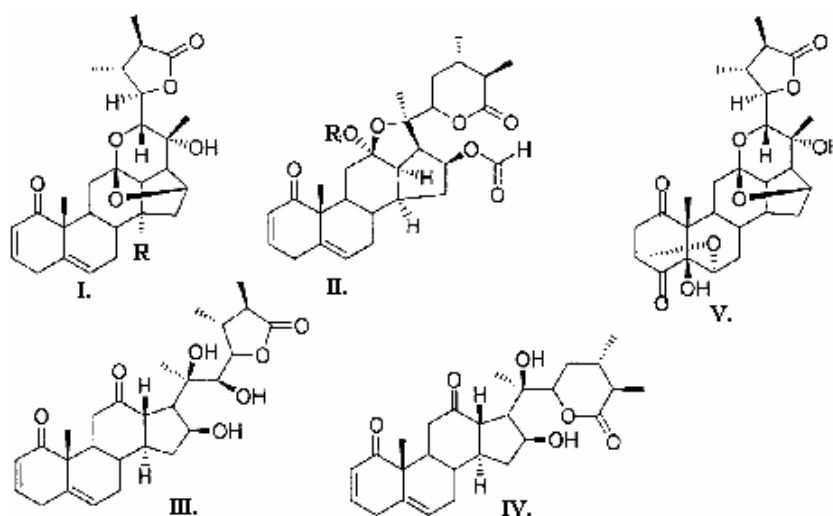
Figure 3.28 Seven norwithanolide structures in data set. Structure I represents m29 with hydrogen as radical group, m30 with hydroxyl as radical group. Compound m33 and m34 are represented with structure II with R as CH_2CH_3 and CH_3 in the same order. III: m32, IV: m31, V: m35.

Compound's LogD values at pH 7.4 are around two, ideal lipophilicity for a drug candidate but m33 and m34 are lipophilic and have 4.41, 4.07 degrees of logD in the same order. Data set norwithanolide compounds have high acceptor site counts up to 8 as in m35 although having low donor sites even none as in m27, m28, m33, m34.

3.3.16 Phenylphenalenone

Phenylphenalenone C19H12O compounds are derived from the phytoalexins family as antibiotics produced by the plant as a defensive [63].
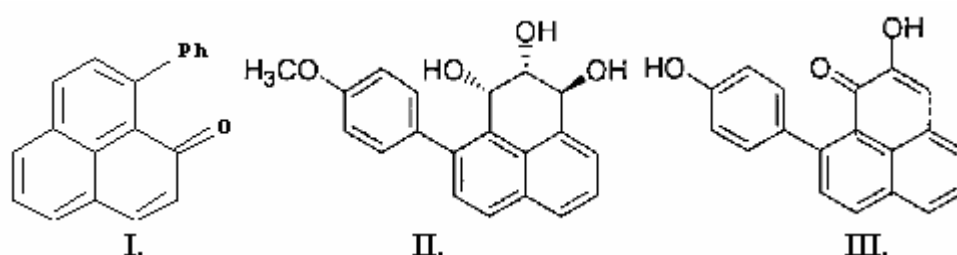
Figure 3.29 Phenylphenalenone skeleton is represented in I. Two compounds from same group of molecules are shown as II: m129, III: m130 in data set.

Although two compounds have different donor and acceptor sites like donors 3 to 1, acceptors 4 to 3 they have similar logD and Pho/phi_ASA ratio values. Pho/phi accessible area ratio is calculated as approximately 4 and logD is around acceptable value of 2 for each compound.

3.3.17 Rotenoids

Rotenones are known to be highly toxic biologically active forms of isoflavonoids. They inhibit NADH:ubiquinone oxidoreductase activity as chemo preventive agents from Derris root [64], [42].
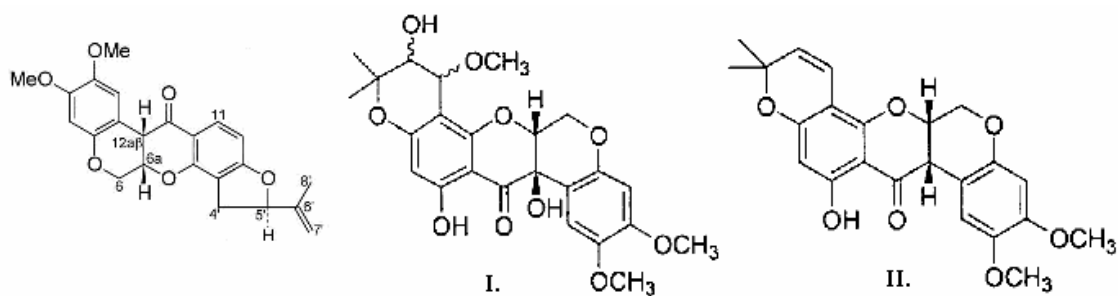
Figure 3.30 General rotenoid structure is demonstrated on the left. Compound m36 and m161 are demonstrated in sequential order.

LogD of 0.6 degree indicates polarity but Pho/phi accessible area ratio is 3.58 that mean controversially high hydrophobicity. Three donor sites and ten acceptor sites are present at pH 7.4.

3.3.18 Sesquiterpenes

Data set sesquiterpenes have three isoprene units with cyclic structures ($C_{15}H_{24}$).
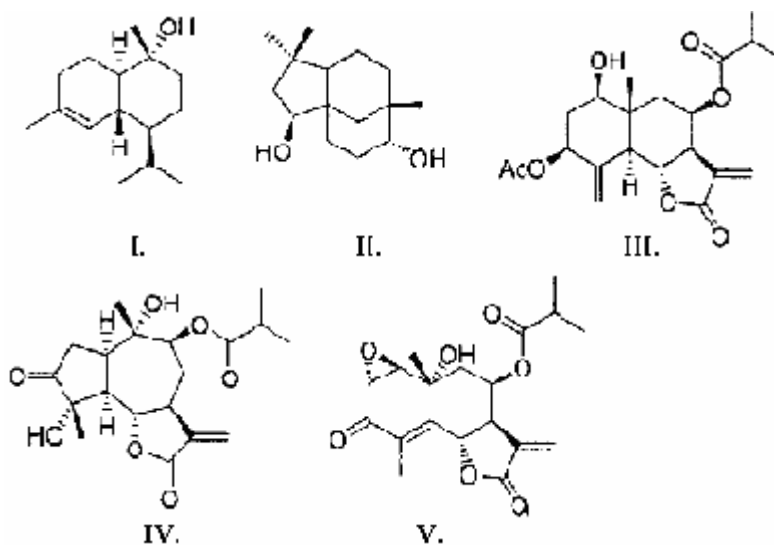
Figure 3.31 Sesquiterpenes in data set are 5 compounds as I: m131, II: m132 , III:m133, IV: m134 and V: m135.

Among five sesquiterpenes, m131 has the highest Pho/phi_ASA ratio of 19.29 and the highest logD of 3.07 even though other four compounds have acceptable logD

values for lipophilicity. With five hydrogen bond acceptors highest acceptor sites are on m134 and m135. Each compound has one or two donor sites.

Sesquiterpene lactones

Another Sesquiterpene with a lactone (cyclic ester) moiety in our data set is shown below.
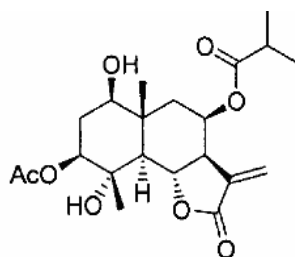
Figure 3.32 The structure of Sesquiterpene lactone compound m37 is shown.

Compound m37 has two donor and 5 acceptor sites at pH 7.4. LogD of 0.9 indicates polarity but Pho/phi accessible area ratio is 3.69 which, as for m36, is a significantly high value for hydrophobicity.

3.3.19 Simaroubolides

Simaroubolides are quassin like compounds which consists of triterpenoid and lactone.
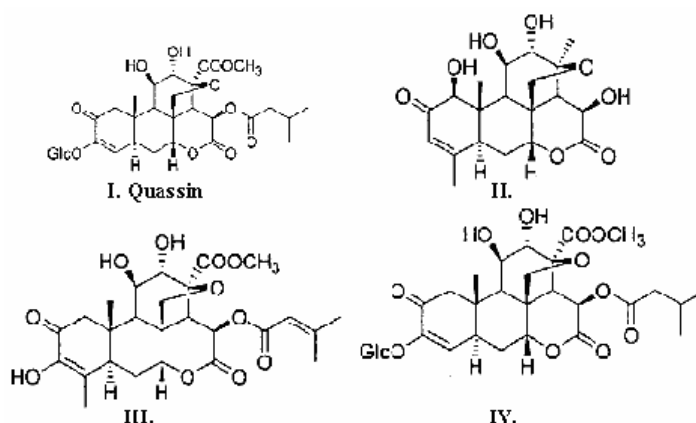
Figure 3.33 An example to simiroubolide shown in I; and quassin in II: m137, III: m138, and IV: m139.

Simiroubolide compounds in the data set have hydrophobic property with 0.27 to -3.26 values of logD and Pho/phi accessible area ratios of 2.05 to 3.09. Most donor and acceptor sites are found in m139 in this group with 6 donor sites and 13 acceptor sites.

3.3.20 Stilbenoids

Hydroxylated stilbene, are referred as antioxidant compounds. Stilbenolignan structure, consists of stilbene and lignan moieties. Structures are shown in Figure 3.34 and 3.35.
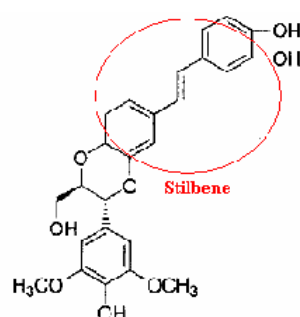
Figure 3.34 Stilbenolignan in data set, Compound m42;.
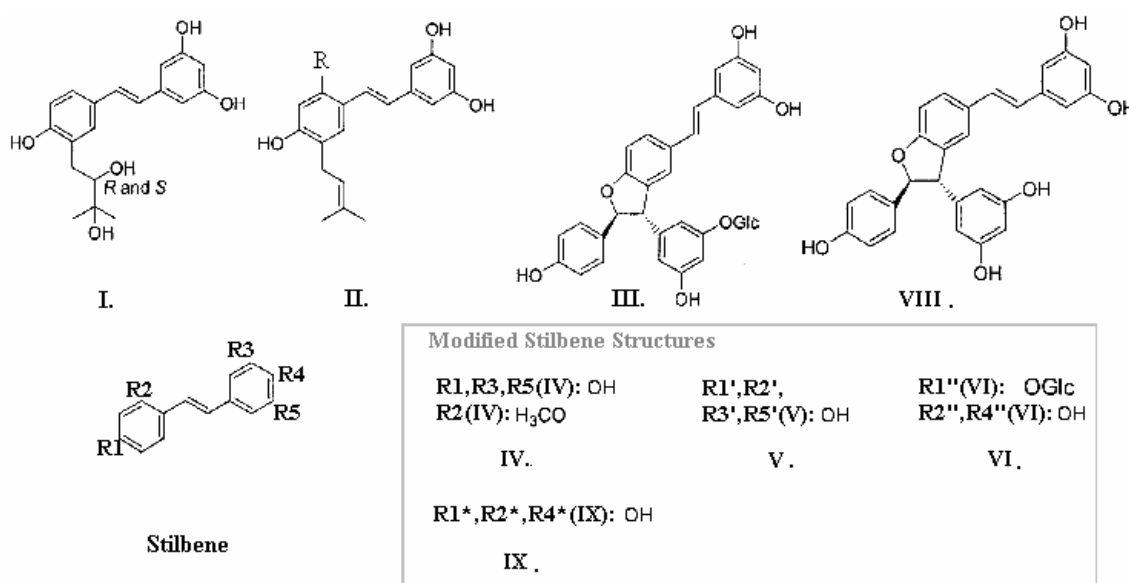
Figure 3.35 Stilbene skeleton is shown as a general structure and six other stilbenes shown in square as, IV: m140, V::m141, VI: m142, VIII:m144, IX:m145. Compounds m39 and m40 are represented as II. Related radical groups are hyroxyl and hydrogen in the same order for each. I: m38, III: m41.

The simplest group of diarylethene is a derivative of stilbene. Diarylethene consists of aromatic groups attached to each end of the carbon-carbon double bond. Resveratrol is one of the examples. Stilbene with four stilbenoid structures in the data set is shown in Figure 3.35. Compounds m144, m40 and m39 have highest logD 6.5, 5.43, 5.13 in the same order. All stilbenoids in data set are lipophilic with positive logD values. Pho/phi_ASA ratio for each stilbenoid structure in the data set is around 2. The most lipophilic compound m40 in this group has a Pho/phi_ASA ratio of 3.78. Compound m41 has the highest hydrogen bonding sites with 8 acceptor and 11 donor sites. In Figure 3.34 Compound m42 has Pho/phi_ASA ratio, 3.16 logD 3.88, 4 donor and 8 acceptor sites.

3.3.21 Triterpenes

Saponins are plant triterpenes with sugar groups. Bruceantin, a triterpene, has anti-neoplastic activity (decrease in abnormal cell's proliferation). Sequalene is a linear form of triterpene derived from shark liver oil, precursor of cholesterol [46].
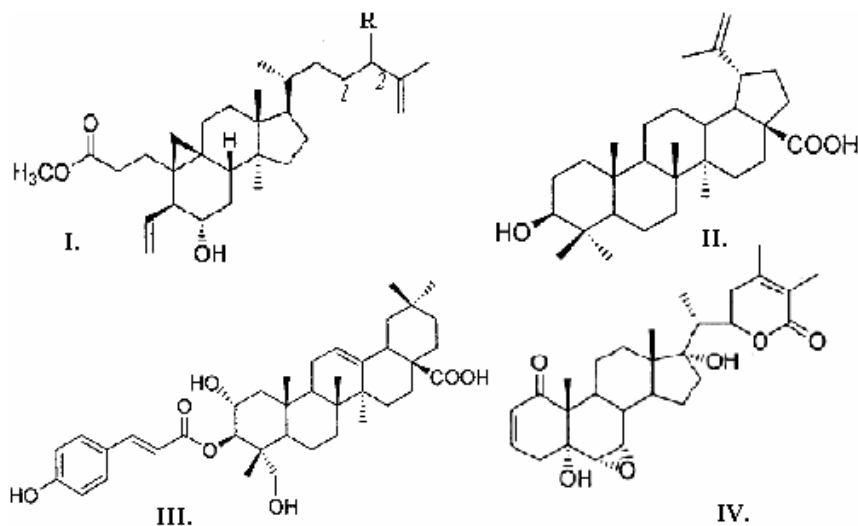
Figure 3.36 Two triterpene structures m43 and m44, represented as I. Compound m43 has hydrogen in place of R, a double bond between carbons numbered 1 and 2. Compound m44 has OOH as a radical group. II refers to compound m147, III refers to m162 and IV refers m165.

Triterpenes contain 30 carbons and 6 isoprene units. Steroids are modified triterpenes. Compounds shown in Figure 3.36 are hydrophobic with logD values around 4. While compound m147 has the highest Pho/phi_ASA ratio of 8.39 with highest number of hydrogen binding sites in this group, with 3 donor and 6 acceptors.

3.3.22 Withanolides

Withanolides are suppressers of cancer cell proliferation and inhibit NFκβ activity, hence promoting apoptosis.
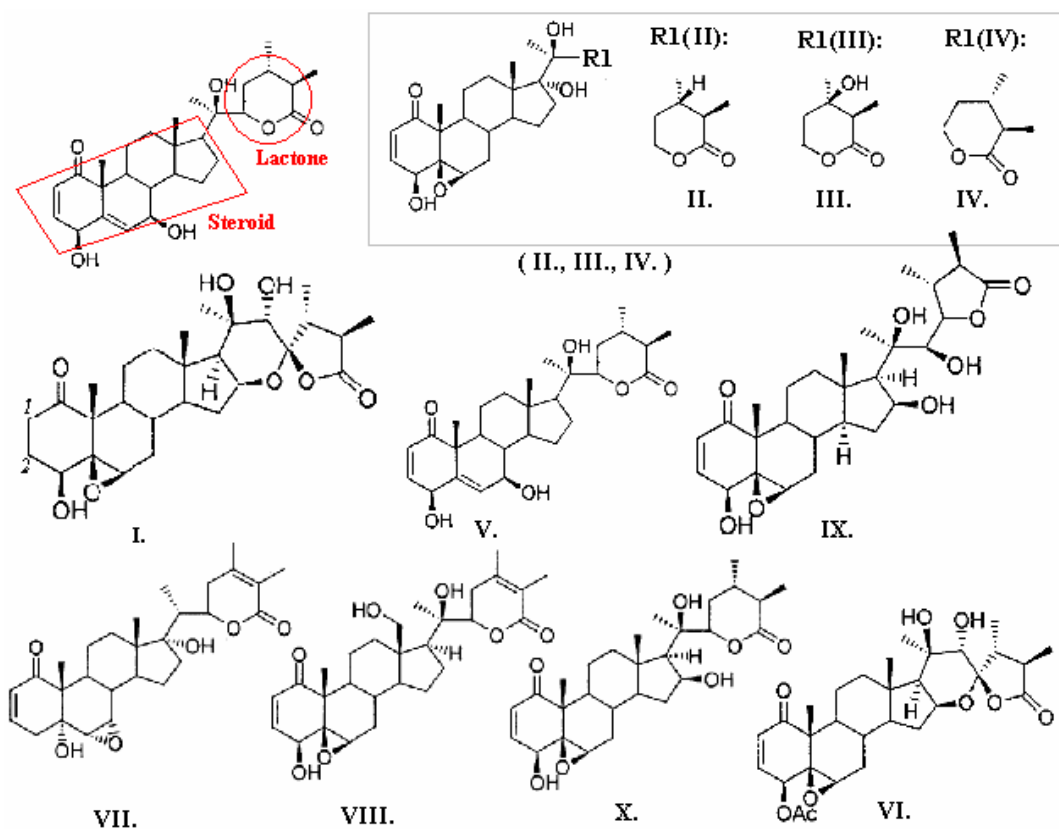
Figure 3.37 Five withanolides can be demonstrated. Compound m45 has the structure as shown in I . Compound m46 has the structure as I but with double bonds with 1st and 2nd carbon as shown. Compound m47, m48 and m149 are shown as II, III, IV in the same order. Compound m49 is as shown as V, VI: m148, VII: m160, VIII: m163, IX: m164, X: m166.

Ergostane type steroidals structurally C22 (structure contains 22 carbons) and C26 are oxidized to form lactone. Steroidal lactones have the chemical formula $C_{28}H_{46}O_2$. Compounds induce neural regeneration. Reported activities are: (i) anti-inflammatory, (ii) anticonvulsive, (iii) antitumorogenic, (iv) immunosuppressive, (v) antioxidant. One source is root of Ashwagandha an Indian herbal [47]. Other namings are *Withania Somnifera,* Winter Cherry, and Indian Ginseng. Plant family *Solanaceae* produces withanolides. Compounds shown in Figure 3.37 are lipophilic. Compound m166 has the highest Pho/phi_ASA ratio of 6.48. LogD values 0.74 and 0.7, respectively, of the two compounds m48 and m164 are more neutral than the other compounds in the group.

3.3.23 Cycloartane triterpenoids

Cycloartane triterpenoids consist of cycloartene and triterpene moieties. Modified structures are shown on the general structure of **Cycloartegenol** in Figure 3.38 and 3.39**.**
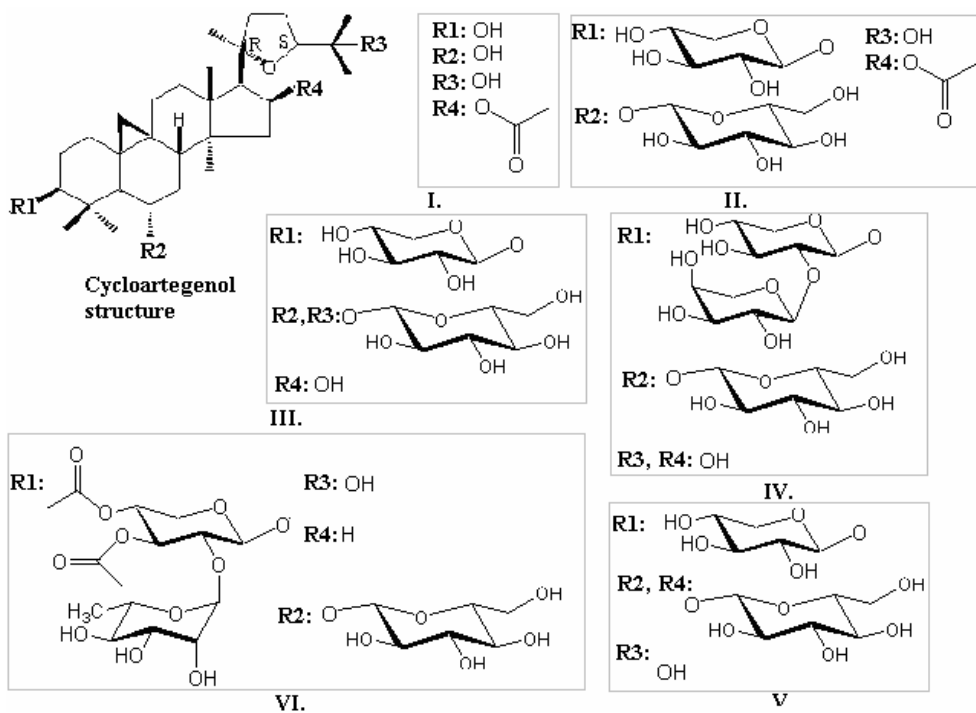
Figure 3.38 Cycloartagenol derivatives; I: Cyclosieversigenin, II: TrojanosideA, III: TrojanosideB, IV: TrojanosideH, V: TrojanosideK, VI:TrojanosideJ.
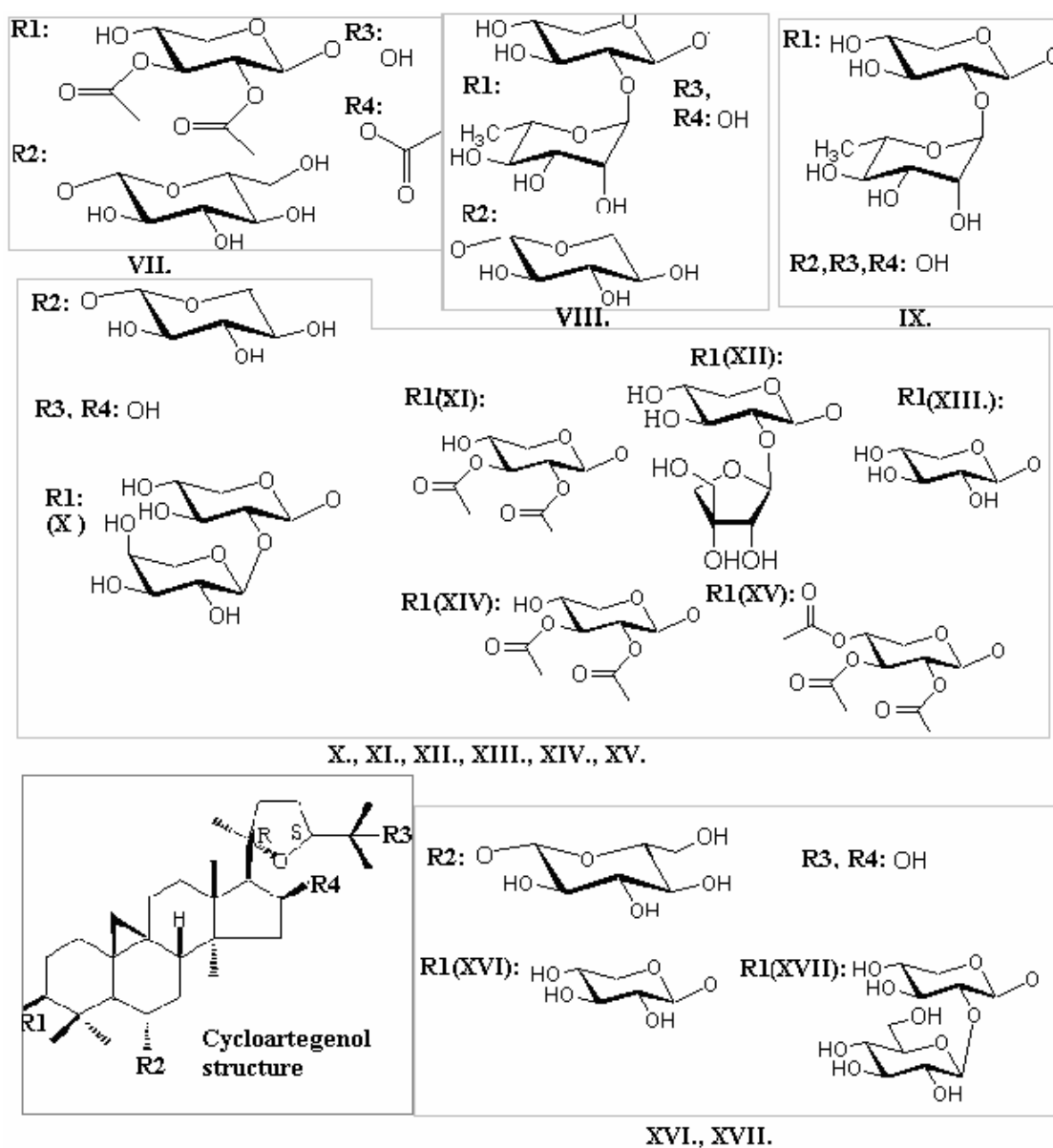
Figure 3.39 Cycloartagenol structures in data set are described with radical groups R1, R2, R3 and R4 in groups. VII: TrojanosideI , VIII: AstraservisianinXV , IX: Elongatoside , X: AskendosideD , XI:AstraservisianinII , XII: Baibutoside , XIII: AstraservisianinX , XIV: AstragalosideI , XV: Acetylastragaloside, XVI: AstragalosideIV , XVII: AstragalosideVI .

Among Cycloartegenols in data set maximum Pho/phi accessible area ratio is 6.93 with compound Cycloartagenol (Cyclosieversigenin). Cycloartagenol is the most hydrophobic with logD of 2.37 and less donor site number of 3 and 5 acceptor sites.

All the others are having logD of negative or neutral values. The most hydrophilic compound is TrojanodeK with a logD value of -3.24, with 12 donor sites and 19 acceptor sites which is the same for TorjanosideB. Both compounds have the same radical groups but in different places in R3 and R4 radical groups, hence having different volume and Phi/pho_ASA values. Other cycloartane derivatives are shown below in Figures 3.40, 3.41, 3.42 and 3.43.
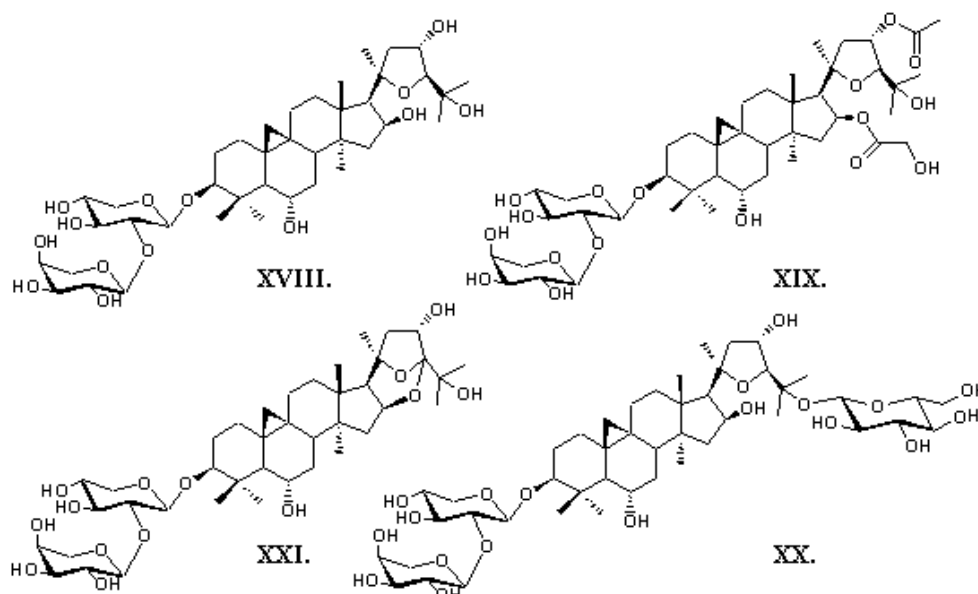
Figure 3.40 Cycloartane derivatives in data set are represented with XVIII: compound IC-1, XIX: IC-2, XX: IC-3, XXI: IC-4.

IC-4 has 12 donor and 19 acceptor sites and is the most hydrophobic compound among four in Figure 3.40. Compound IC-4 has a LogD value of -3.69 and Pho/phi_ASA of 2.38. The other three compounds are also of hydrophobic character.

Cyclocanthogenin and derivatives in data set are shown with radical groups in Figure 3.41.
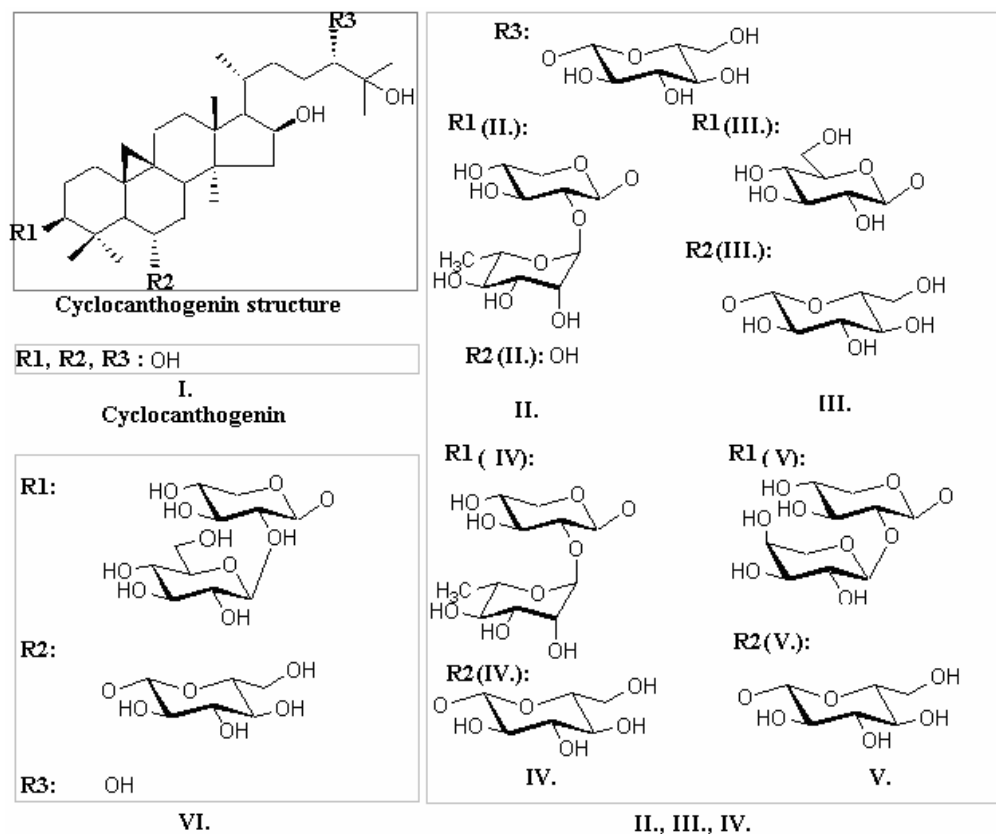
Figure 3.41 Cyclocanthogenin general structure is represented on the left corner. Cyclocanthogenin and its derivatives in data set are represented. I: Cyclocanthogenin, II:TrojanosideC, III: TrojanosideD, IV: TrojanosideE, V: TrojanosideF, VI: CyclocanthosideG.

The most hydrophilic compound is TrojanodeF with -4.26 logD values with 15 donor sites and 23 acceptor sites.
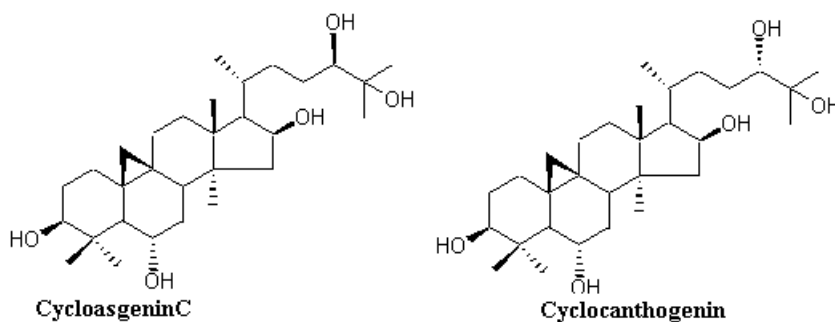
Figure 3.42 CycloasgeninC has the same structure as Cyclocanthogenin; it is R isoform in 24th carbon rather than S. Both compounds has same logD of 2.86 and 5 donor, 5 acceptor sites. But the volume of the structure hence Pho/phi_ASA are slightly different; 5.72, 5.97 in sequential order.

Cyclocephalogenin and Macrophyllogenin structure and derivatives in data set are shown with radical groups.
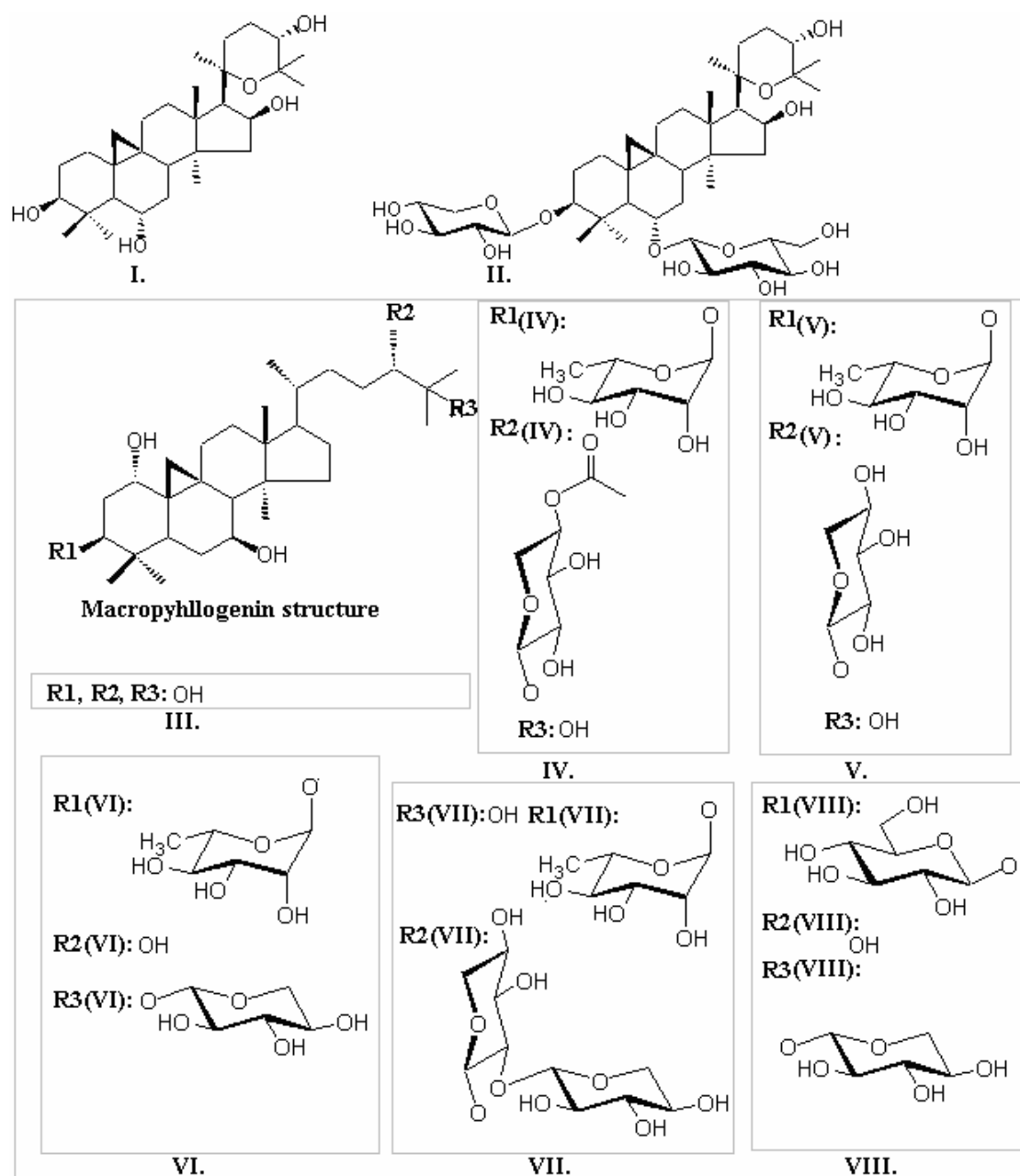
Figure 3.43 Cyclocephalogenin and macrophllogenin derivatives are shown in figure. I:Cyclocephalogenin, II:CyclocephalosideI, III: Macrophyllogenin, IV: MacrophyllosaponinA, V:MacrophyllosaponinB, VI: MacrophyllosaponinC, VII: MacrophyllosaponinD, VIII:MacrophyllosaponinE.

Most hydrophilic among eight compounds with 11 donor and 15 aceptor sites and logD of −1.35 and Pho/phi_ASA of the lowest 2.71, compound is MacrophyllosaponinE. The most hydrophobics are Macrophyllogenin and Cyclocephalogenin with 2.88, 2.44 logD values in sequential order.

3.3.24 HARMOL ALKOLOIDS

Harmol alkoloids are effective on the nervous system and has cytotoxic effects.
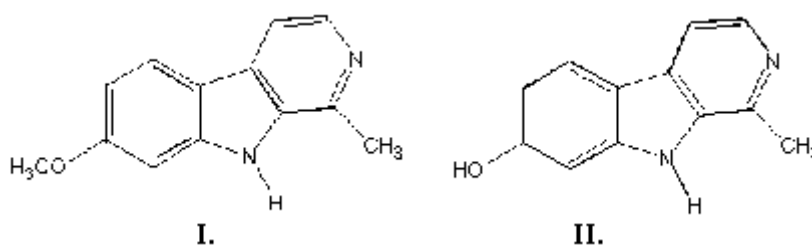
Figure 3.44 Harmol alkoloids in data set are shown above. I: Harmin, II: Harmol.

Both compounds are hydrophobic. Compound harmin has Pho/phi_ASA value of 8.82 and has one donor site and two acceptor sites. Harmol has a total of four hydrogen bonding sites and as a result, a Pho/ohi_ASA value of 4.36.

3.3.25 Resveratrol derivatives

Resveratrol molecules are stilbene derivatives. Structures are shown in Figure 3.45.
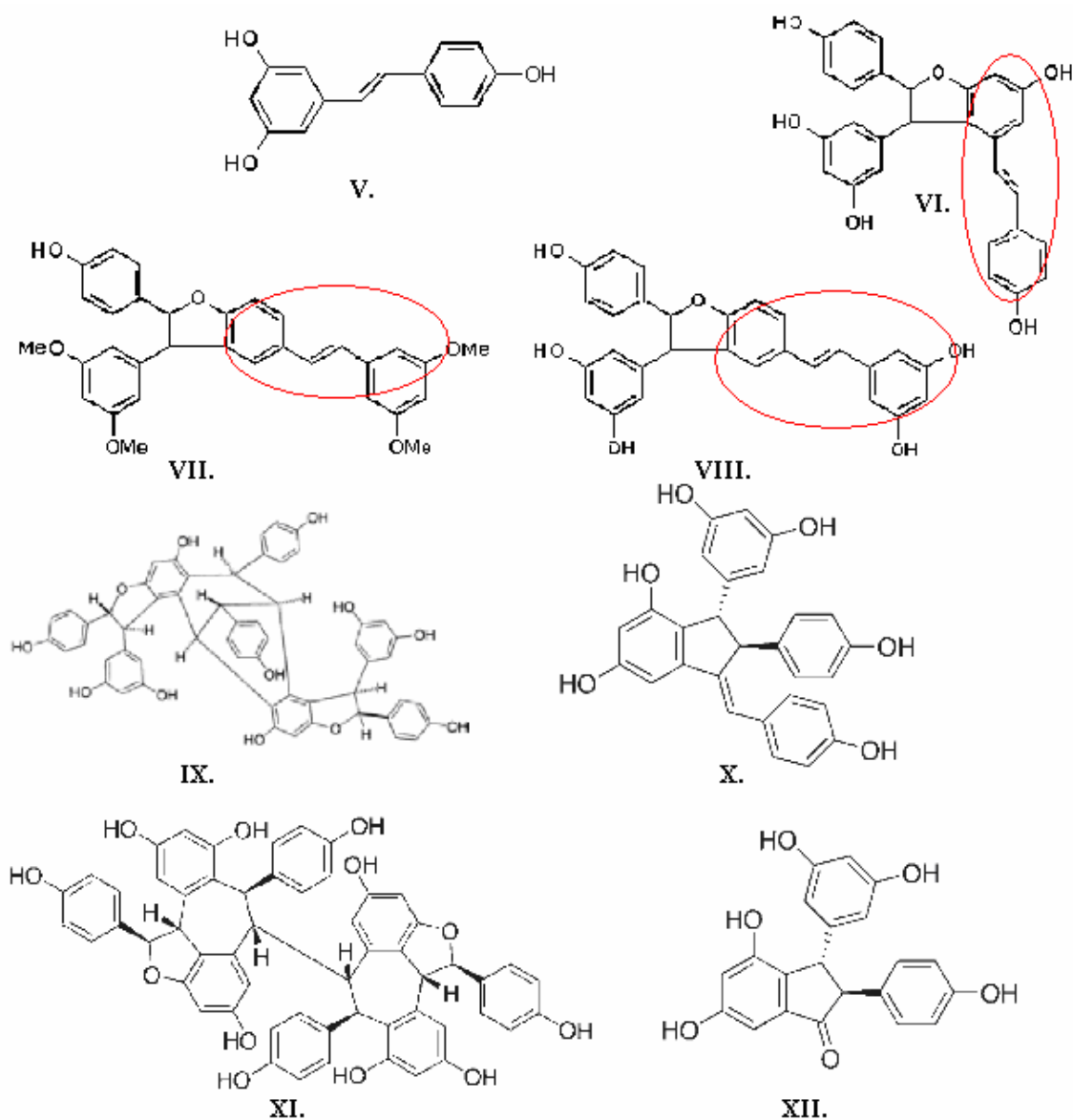
Figure 3.45 Resveratrol is demonstrated at V. Resveratrol derivatives in data set are shown above. VI: 12-viniferin, VII: Pterostilbene-dehydrodimer, VIII: Resveratrol-trans-dehydrodimer, IX: vaticanolC, XI: ampelopsinD, XII: paucifloralF. Some of the stilbene structures are shown in circle.
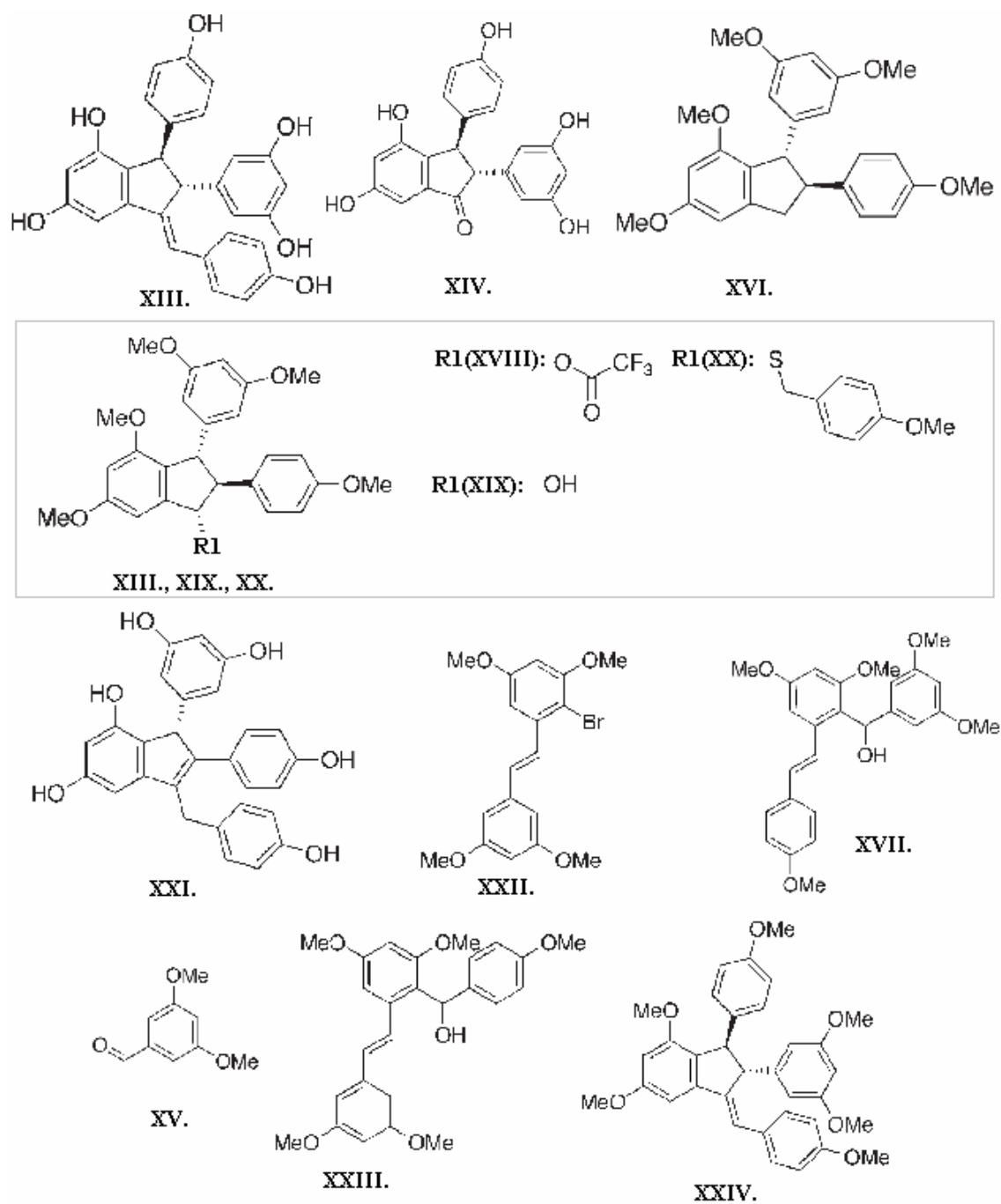
Figure 3.46 Resveratrol derivatives in data set are shown in figure. XIII: quadrangularinA, XIV: IsipausifloralF, XV: res10, XVI: res11, XVII: res23, XVIII: res14, XIX: res15, XX: res16, XXI: res17, XXII: res18, XXIII: res20, XXIV: res23, XXV: res24. Similar structures are shown on one representative structure in gray are quadrangularinA, res15, res16.
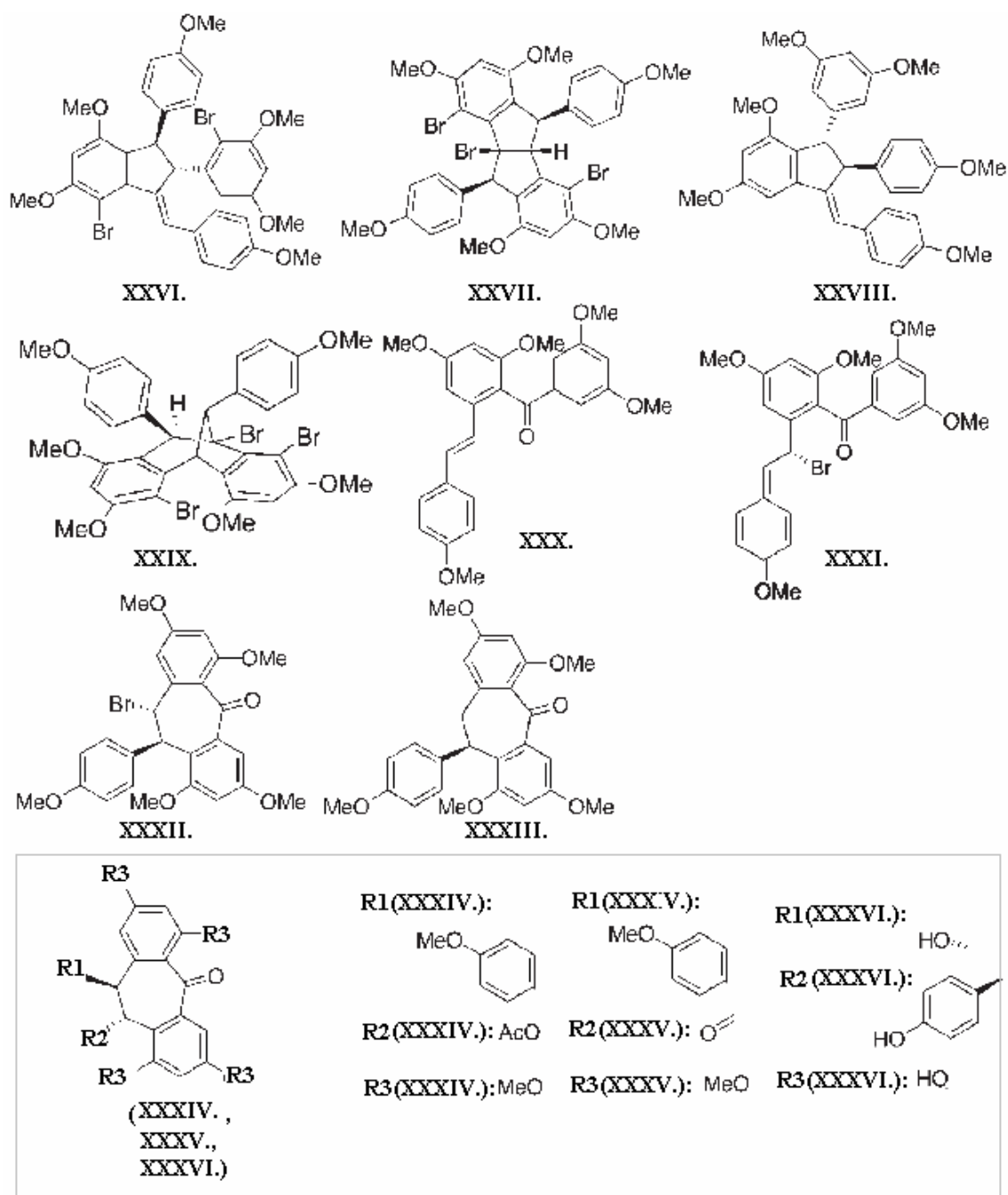
Figure 3.47 Resveratrol derivatives in data set are shown in figure. XXVI: res25, XXVII: res27, XXVIII: res28, XXIX: res30, XXX: res31, XXXI: res32, XXXII: res33, XXXIII: res34, XXXIV: res36, XXXV: res37, XXXVI: res38. Similar structures are shown in gray which res36, res37, res38.

Forty compounds are listed in this group. All compounds are lipophilic with highly positive logD values. Compound Res27 has the highest logD value of 8.25.

3.3.26 Pramanicin

Praminicin molecule is an anti-fungal drug. It is shown that it induces cell death and induces apoptosis in Jurkat T leukemia cells [65]. Pramicin and ten derivatives are used in our data set. Structures are shown in Figure 3.48.
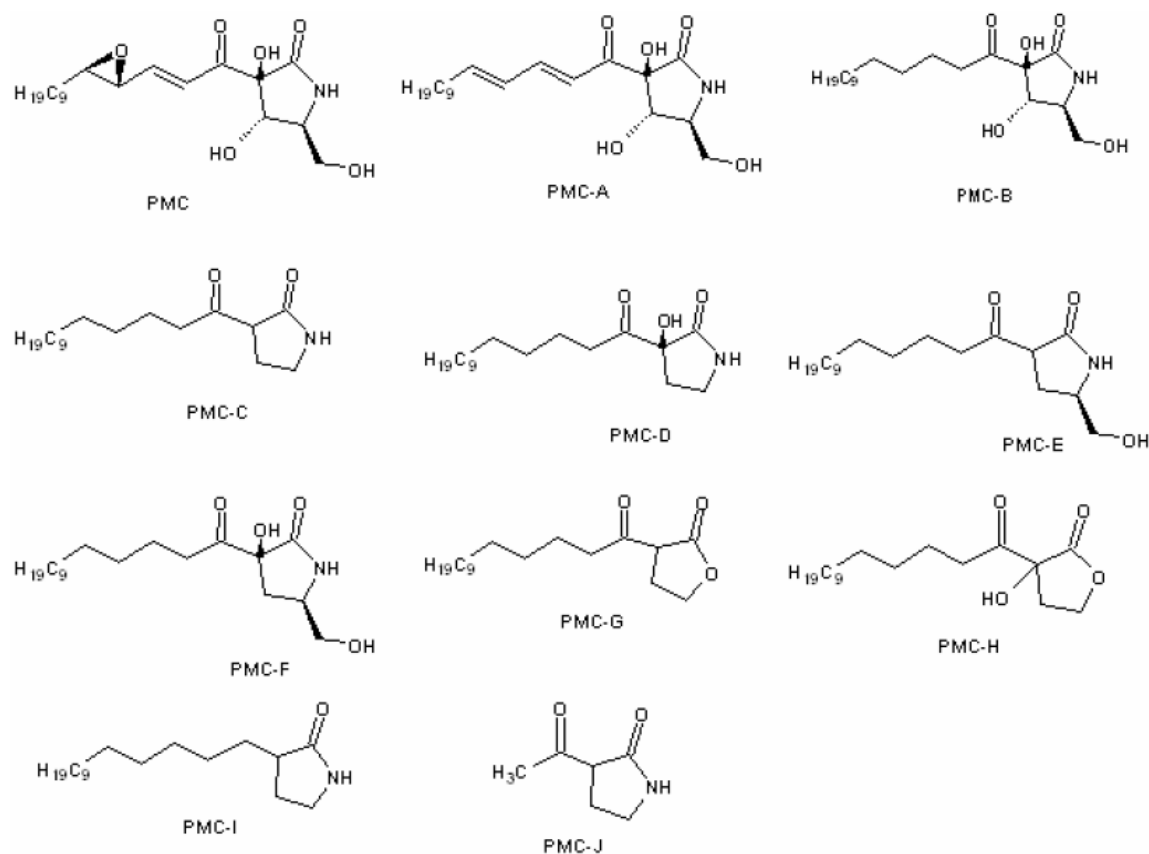
Figure 3.48 Pramanicin molecule and its ten analogs are shown.

Pramanicin and analogs are lipophilic except compound Pmc-J. Pmc has logD value of 1.16 while Pmc-J has a hydrophobic structure with logD value of -0.30. Pmc has four donor and six acceptor hydrogen binding sites while Pmc-J has one donor and two acceptor hydrogen binding sites. Pmc-G is the most lipophilic of eleven structures logD value of 6.03. Other analogs are lipophilic structures with logD value of around two.

3.3.27 β-Ionone

β-Ionones are ketones which are used as food additives with emulsifier properties. An example is used in the data set and structure is shown in Figure 3.49.
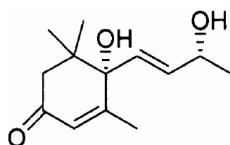
Figure 3.49 β-Ionones as compound m121in data set is shown. With logD value of 1.39, hydrogen binding donors of two and acceptors of three.

3.3.28 Other compounds; NF-κB Inhibitors
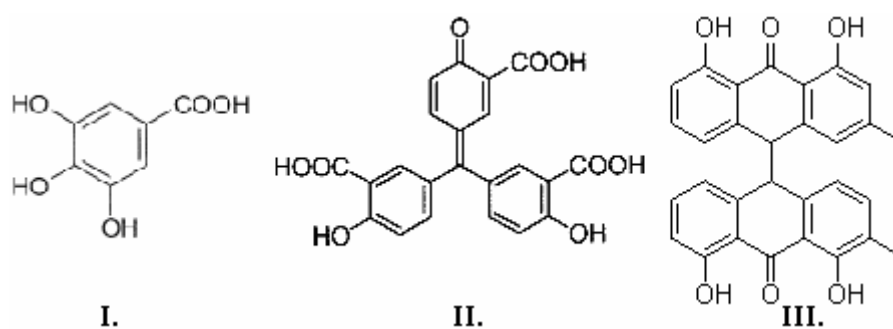
Figure 3.50 NF-κB inhitors are shown. Polyhydroxylate derivative – GA (gallic acid), Aurine Tricarboxylic Acid (ATA) and MMD are shown as I, II and III in sequential order.

Gallic acid in data set is a hydrophilic structure with a logD value -2.56 and with three donor and five acceptor hydrogen bonding sites. Aurine Tricarboxylic Acid, (ATA) a hydrophilic structure with a negative logD value of -6.23, which is the lowest value of all data set, is shown in Figure 3.50. ATA has two donor and nine acceptor hydrogen binding sites. Compound MMD is a lipophilic structure with a logD value of 8.36 and Pho/phi_ASA ratio of 2.91. MMD has four donor and six acceptor hydrogen binding sites.
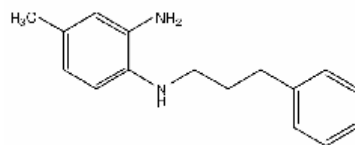
Figure 3.51 Sjh53

Sjh53 has lipophilic properties; logD value of 3.72, Pho/phi_ASA ratio of 8.15.

Chapter 4

RESULTS OF NF-κB CALCULATIONS

The data set of 236 natural compounds described in Chapter 2 are docked to NF-κB by three different docking algorithms; Gold Score, Chem Score and AutoDock. Gold Score and Chem Score calculations are carried out on GOLD 4.0. And AutoDock 4.0 is used with LGA (Lamarckian genetic algorithm) algorithm.

4.1   Results

The top ten ranked molecules and scoring functions which are used to rank them are listed in Table 4.1 (with best binding modes). These compounds listed are flavonoids, benzenoids, alkaloids, withanolides, cycloartane triterpenoids, lignans and stilbenoids. They are selected from twenty seven compound groups and a majority of these groups are classified as phenolic natural compounds.

Docking with target NF-κB by Chem Score does not rank compounds by considering their specific groups among this data set. However, AutoDock and Gold Score show a tendency for specific structural groups while ranking compounds. More specifically, eight of the ten Gold Score hits are in the stilbenoids group, and seven of the ten AutoDock hits are in the cycloartane tirterpenoids as can be seen from Table 4.1. Thus, Docking analysis with target NF-κB show that; the two scoring functions, Gold Score and AutoDock, have a tendency to select specific molecular structural groups among this data set.

Chem Score (lowest binding free energy)	Gold Score Fitness (highest fitness results)	AutoDock (lowest binding free energy)
m21 *(flavonoid)*	m144 *(stilbenoid)*	TrojanosideJ *(cycloartane triterpenoid)*
m54 *(benzenoid)*	m14 *(falavonoid)*	AskendosideD *(cycloartane triterpenoid)*
m26 *(flavonoid)*	Resrevatrol-trans-dehydrodime *(stilbeneoid)*	m163 *(withanolide)*
m02 *(alkaloid)*	vaticanolC *(stilbenoid)*	CyclocephalosideI *(cycloartane triterpenoid)*
m165 *(withanolide)*	m24 *(flavonoid)*	Elongatoside *(cycloartane triterpenoid)*
CycloasgeninC *(cycloartane triterpenoid)*	res16 *(stilbenoid)*	AstrasieversianinXV *(cycloartane triterpenoid)*
MMD	res25 *(stilbenoid)*	m29 *(norwithanolide)*
m48 *(withanolide)*	Hopeaphenol *(stiilbenoid)*	TrojanosideB *(cycloartane triterpenoid)*
m124 *(lignan)*	res24 *(stilbenoid)*	TorjanosideH *(cycloartane triterpenoid)*
m151 *(alkaloid)*	m41 *(stilbenoid)*	m160 *(withanolide)*

Table 4.1 Chem Score top ten compounds are ranked with their binding free energy results, Gold Score top ten compounds are ranked for their fitness score, AutoDock top ten compounds are ranked with their binding free energies calculated. Gold Score hits are among stilbenoids and AutoDock hits are among cycloartane triterpenods by majority.

The top ten ranked molecules and scoring functions are listed in Table 4.2 along with average results of binding modes. As can be seen from Table 4.2 average result hits of Gold Score show the same trend for stilbenoids as observed in Table 4.1. AutoDock shows two different trends for the best result hits and the averaged result hits. AutoDock averaged results hits are withanolides and norwithanolides (which are also derived from withanolide) structures in a majority. While the best result hits are cycloartane triterpenoid structures in majority.

Chem Score Average	Gold Score Average	AutoDock Average
m26 *(flavonoid)*	m14 *(flavonoid)*	m163 *(withanolide)*
m30 *(norwithanolide)*	res16 *(flavonoid)*	m29 *(norwithanolide)*
m21 *(flavonoid*	Resrevatrol-trans-dehydrodime *(stilbeneoid)*	m81 *(flavonoid)*
m72 *(diarylheptanoid)*	vaticanolC *(stilbenoid)*	m149 *(withanolide)*
m39 *(stilbenoid)*	paucifloralF *(stilbenoid)*	m30 *(norwithanolide)*
Pterostilbene-dehydrodimer *(stilbenoid)*	m155 *(flavonoid)*	m31G *(norwithanolide)*
m148*(withanolide)*	m144 *(stilbenoid)*	m47 *(withanolide)*
res13 *(flavonoid)*	m67 *(benzofuran)*	m35 *(norwithanolide)*
m141 *(stilbenoid)*	m41 *(stilbenoid)*	m160 (Rotenoid)
m88 *(flavonoid)*	m76*(flavonoid)*	m45 *(withanolide)*

Table 4.2 Compounds are ranked with average values of each algorithm used.

Average values are calculated from all runs done for each compound.

Best results obtained from three scoring functions are different from each other. There is no correlation between the best results of the three scoring functions. In Figures 4.1, 4.2, and 4.3 we search for an overall correlation on the full data set. Charts are non-scaled and no significant correlation between scoring function results is observed. Gold Score vs Chem Score graph shown in Figure 4.1 has better linear regression (R^2=0.18) but not significant. Results of each binding modes are averaged and molecule hit lists are rearranged. Averaged results are shown in Figures 5.4, 5.5 and Figure 5.6 has much more significant linear regression. R^2 tables of averaged results are shown in Appendix 5. Linear Regression for Gold Score vs. Chem Score (R^2=0.38) and AutoDock vs. Chem Score (R^2=0.28) are significant.
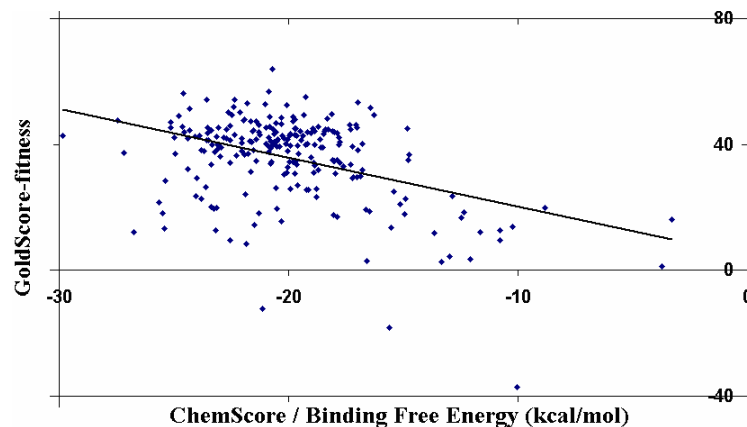
Figure 4.1 Results in Chem Score binding free energy and Gold Score fitness are compared in graph. Linear regression is not significant; R2:0.1758.
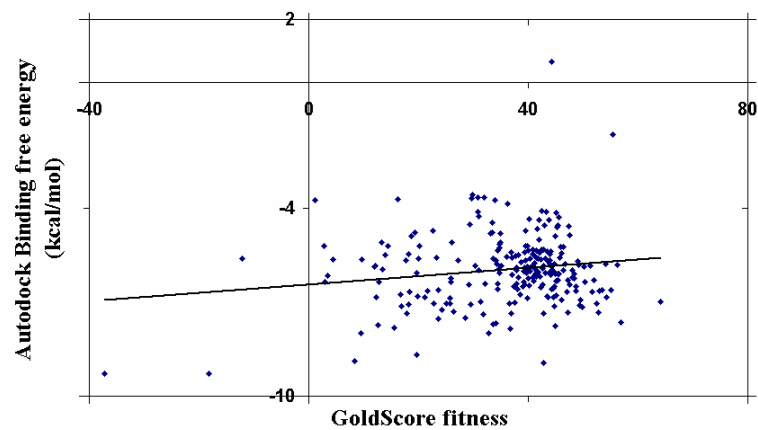
Figure 4.2 Results in Gold Score fitness and AutoDock binding free energy are compared in graph. Linear regression is not significant; R2:0.0249.
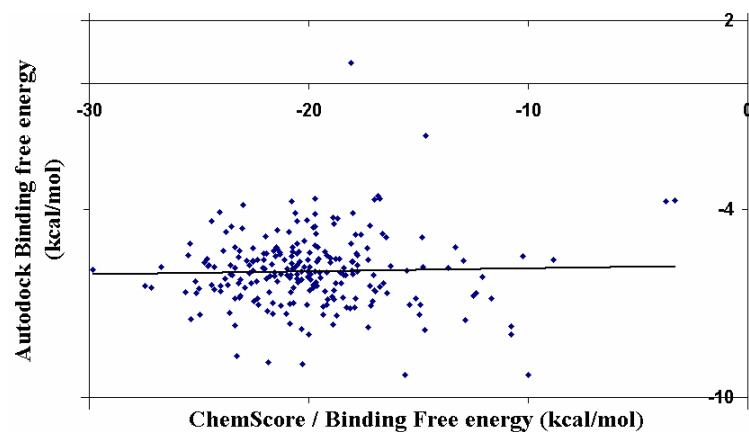
Figure 4.3 Results in Chem Score binding free energy and AutoDock binding free energy are compared in graph. Linear regression is not significant (R^2: 0.001).
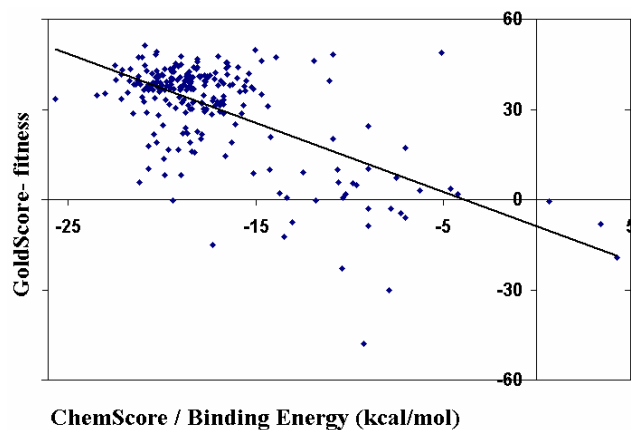
Figure 4.4 Average results in Chem Score binding free energy and Gold Score fitness are compared in graph. Linear regression is significant; R^2: 0.366.
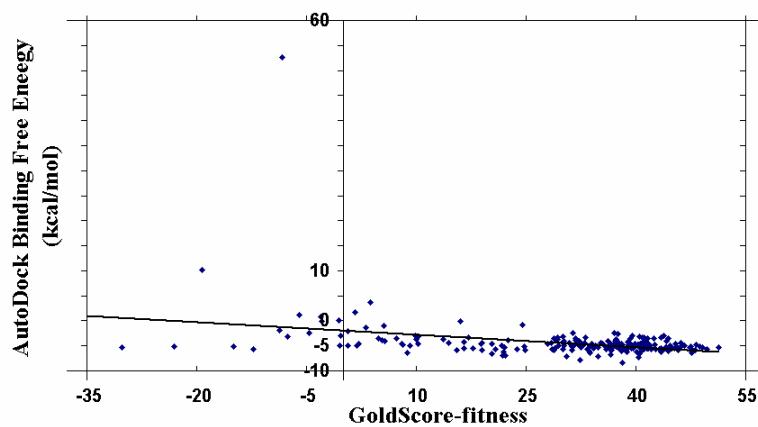
Figure 4.5 Average results in Gold Score fitness and AutoDock binding free energy are compared in graph. Linear regression is not significant; R2: 0.1167.
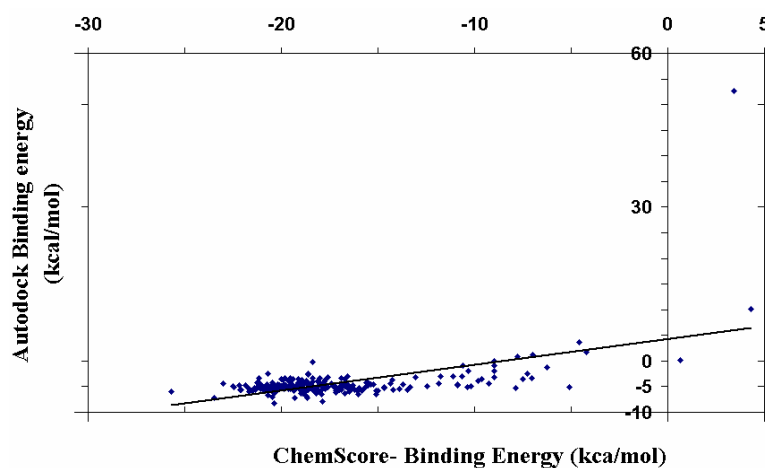
Figure 4.6 Average results in Chem Score and AutoDock binding free energies are compared in graph. Linear regression is not significant, R^2: 0.288.

As mentioned above the best results obtained from three scoring functions are different from each other. Here we have a closer look at the best two results of each scoring function in order to have a better understanding. Investigation of the best results individually will give us an insight for determining which tool and scoring function is the most suitable for our data set. Detailed results of selected compounds are demonstrated in the rest of this chapter.

4.2 Detailed Results of Selected Compounds

Top two ranking compounds are selected for each scoring function. Compounds that are common in the best ten list of the three scoring functions are selected only. For example, the first two ranked compounds for AutoDock have no affinity according to Gold Score fitness, hence they are not selected. Compound m163 (3rd ranked in AutoDock), compound CyclochepalosideI (4th ranked in AutoDock), compound m144 (1st ranked in Gold Score), compound m14 (2nd ranked in Gold Score, 1st ranked in Average Gold Score), compound m21(1st ranked in Chem Score) and compound m26(3rd ranked in Chem Score DG) are selected and details are shown in Table 4.3.

Molecule Name	AutoDock binding free enegy (kcal/mol)	(AutoDock) Average b. energy	Gold Score	(Gold Score) Average fitness	Chem Score DG b,f,e, (kcal/mol)	(Chem Score) Average b. energy
m163	**-8,94**	-8,33	42,79	38,19	-20,27	-20,37
CyclochepalosideI	**-8,87**	-5,03	8,39	0,66	-21,85	-13,34
m144	-5,27	-4,84	**64,05**	32,62	-21,43	-20,98
m14	-7,05	-4,56	**56,85**	20,29	-15,4	-10,88
m21	-5,92	-4,34	42,63	35,24	**-29,84**	-23,01
m26	-6,5	-6,04	37,4	33,58	**-27,18**	**-25,71**

Table 4.3 Two hits of each docking tool is shown.

Molecular orbital theory states that frontier orbital LUMO and HOMO are the reaction centres of most possible electron acceptor and most possible electron donor, respectively. Specified orbital of each molecule may show us the general trends for hydrogen binding between ligands and proteins. Molecular orbital calculations are done in order to understand the general trend in hydrogen bonding between NF-κB and the compounds. Lowest unoccupied molecular orbital (LUMO) and highest occupied molecular orbital (HOMO) of eleven residues in the binding site of NF-κB are shown in Figure 4.7. All figures shown below are drawn by Chem3Dpro11.0 [48], WebLabViewer Pro [49] and Hermes-based GOLD interface [50].
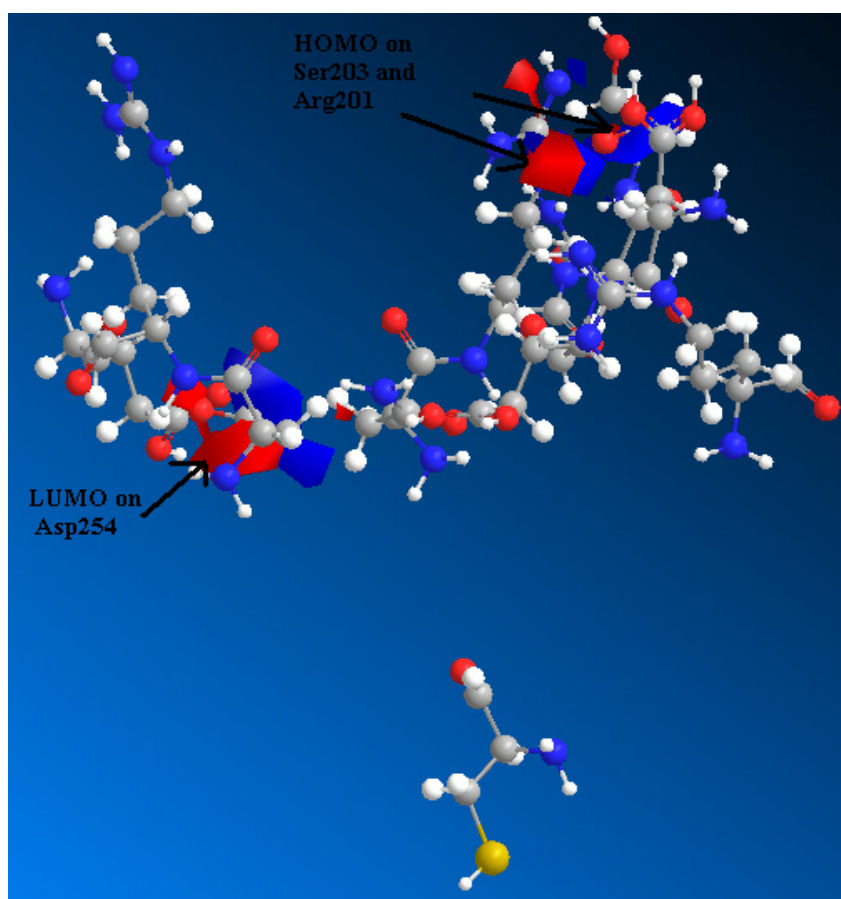
Figure 4.7 LUMO and HOMO of NF-κB are calculated with extended Hückel calculation. HOMO is on Serine [203rd] and Arginine[201st] with -10.403 eV, LUMO is on Aspartic acid [254th] with 0.985 eV.

4.2.1   m163 (3rd ranked in AutoDock)

Docking shows no hydrogen bond for this binding mode. On the left corner possible hydrogen bonds within the 2.5Å distance cut-off are shown in Figure 8. One hydrogen bond donor of m163 is shown together with Asparagine[200th] on NF-κB.
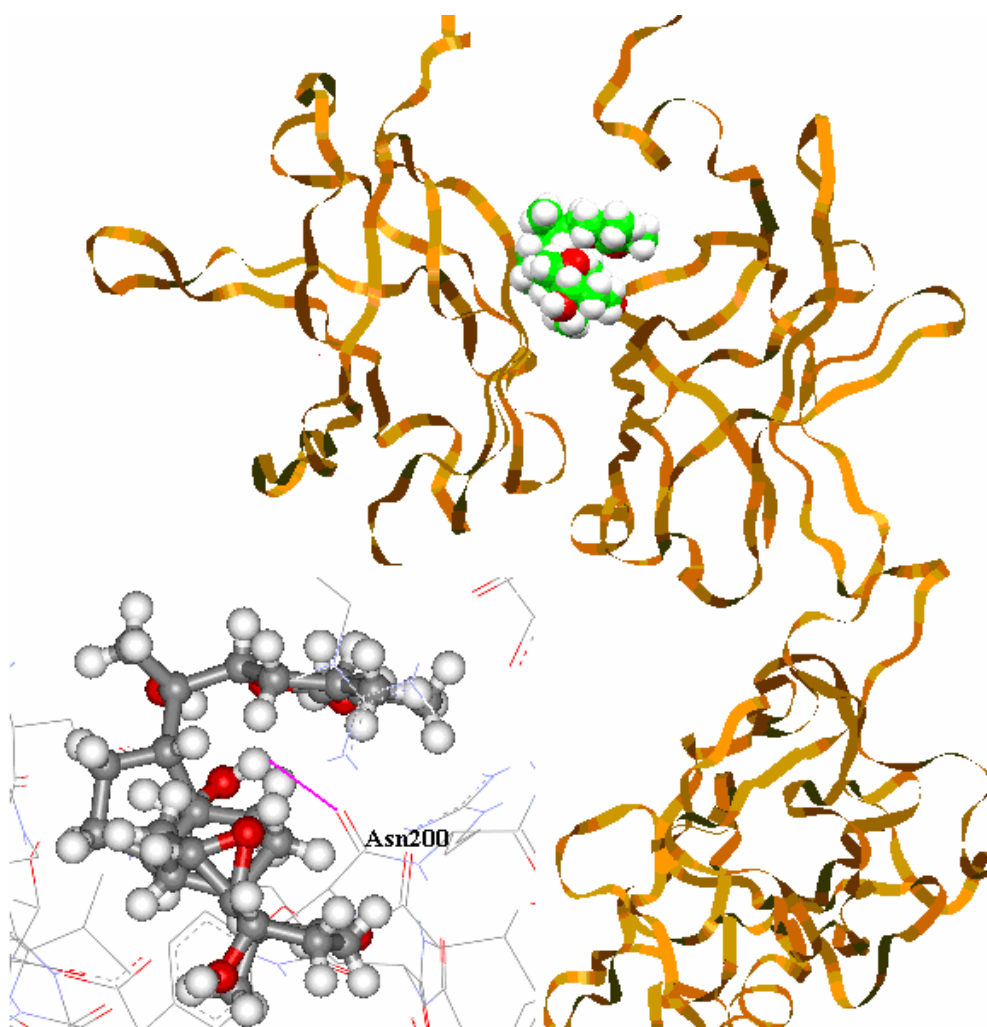
Figure 4.8: Compound m163 docked with NF-κB with AutoDock. On the left corner hydrogen bonds in 2.5Å distance are shown. Hydrogen bond donor of m163 is shown with Asparagine[200th] on NF-κB .

Compound m163 is a withanolide derivative which is named also as steroidal lactone. Compound has three donor and six acceptor hydrogen binding sites but in this binding

mode there is only one donor that is available for hydrogen bonding with the protein. Compound m163 is lipophilic with logD value of 2.46. Compound has Pho/phi_ASA ratio of 4.79 with total solvent accessible area of 536 $\mathring{A}^2$. Compound m163 has lipophilicity because logD value is positive and Pho/phi_ASA ratio is over unity. But logD around two is an acceptable lipophilicity for specific binding. This lipophilicity refers to a good drug like structure. On the other hand molecule has the lowest total solvent accessible area among other five hit compounds, meaning that less complex structure, less drug like structure [56]. In Figure 4.8 below HOMO (Highest occupied molecular orbital) of compound calculated with extended Hückel is shown. As Figure 4.8 indicates that hydrogen bond is built between the HOMO of compound m163 and the protein. Binding mode of the molecule is supported by molecular orbital theory.
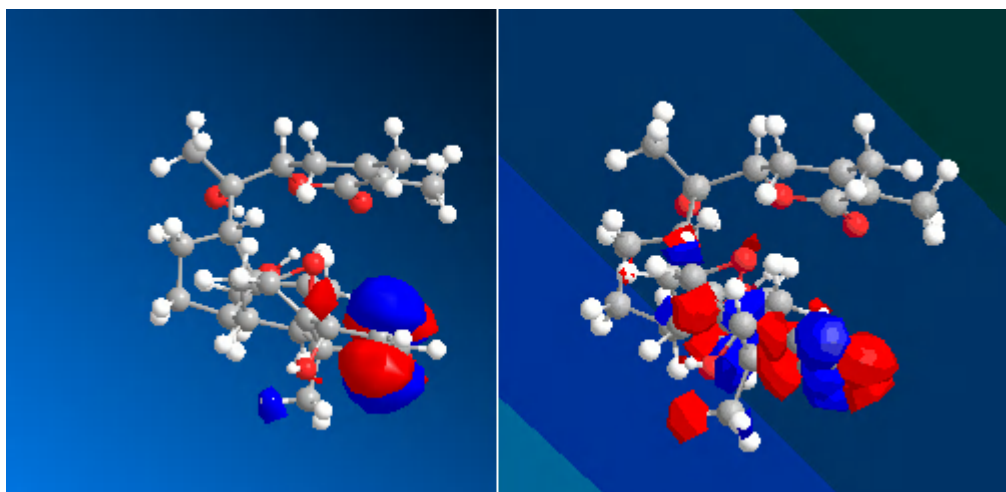
Figure 4.8 Compound m163 docked to NF-κB with AutoDock. On the left compound LUMO and HOMO on the right with extended Hünkel orbital surfaces shown. Red as positive, blue as negative orbital surface.

4.2.2   CyclochepalosideI(4th ranked in AutoDock)

The best ranked mode for docking of cyclocephalosideI to NF-κB with AutoDock along with Lamarcian genetic algorithm function represented as in Figure 4.9.
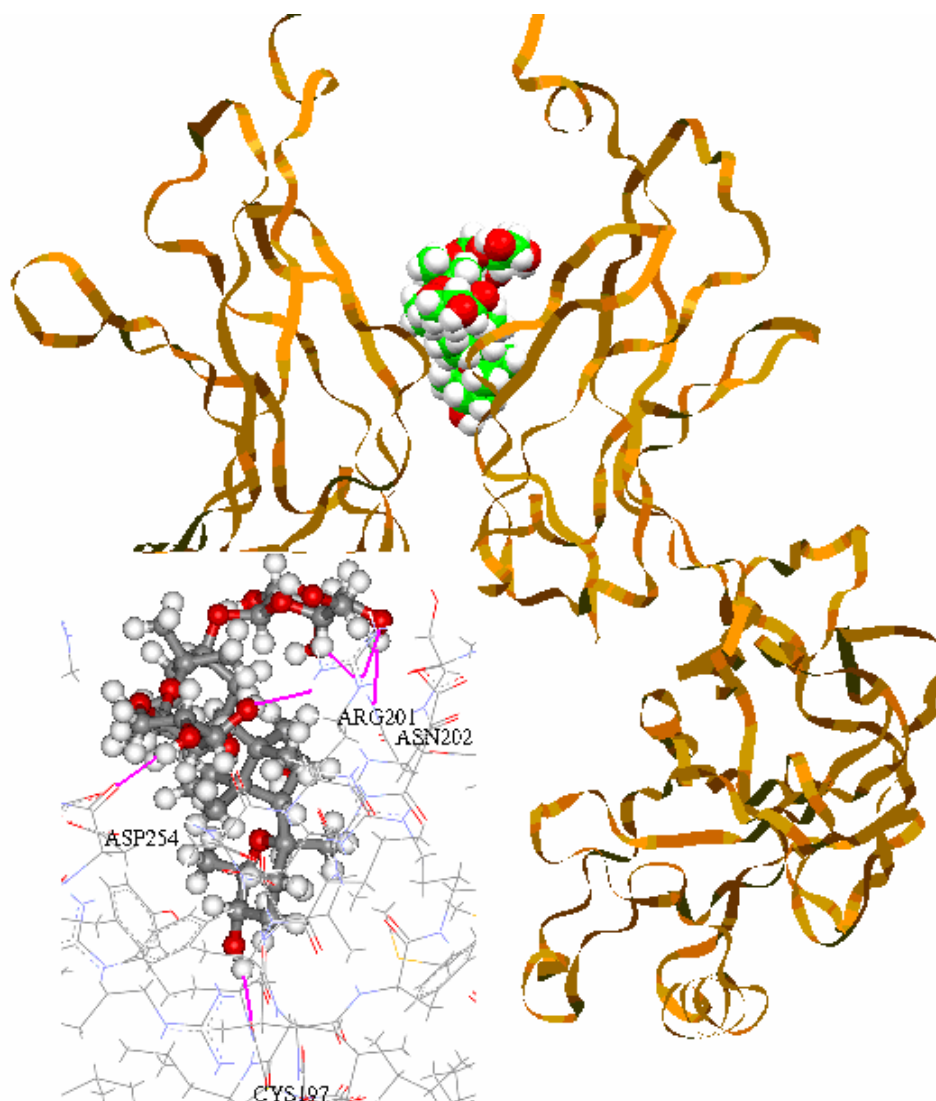
Figure 4.9 Docked NF-κB and cyclocephaloside best binding mode. On the left corner; six H-bonds within distance of 2.5A; four residues on the NF-κB as shown in purple contact points. Contact points are on residues; Arginine[201st], Asparagine[202nd], Aspartic acid[254th], Cysteine[197th] .

Docking free energy calculation shows that three hydrogen bonds in this binding mode including residues 201st, 202nd and 197th on the protein. However there are possible six hydrogen bonds within 2.5Å cut-off distance as shown in Figure 4.10. Two hydrogen bond donors and one acceptor of the compound is shown with Asparagines[202nd residue], a hydrogen bond acceptor on the compound is shown with Arginine[201st], other two donors on the compound are shown in contact with Aspargine[254th] and Cysteine[197th] on NF-κB .

As can be seen in Figure 4.10, although there are nine donor and fourteen acceptor hydrogen binding sites, only five of them are used for short contacts with the protein. Cycloartane triterpenoid group Cyclocephalogenin derived molecule CyclocephalosideI is hydrophilic with logD value of -1.12, which is not good for a drug permeability. This compound has solvent accessible area of 738 Å^2 while Pho/phi_ASA ratio is 3.03. In Figure 4.10 below HOMO (Highest occupied molecular orbital) of that compound calculated with extended Hückel is shown. Homo of cyclocephalosideI has contact with 201st residue as a hydrogen bond shown in Figure 4.10. Hence, binding mode of CyclocephalosideI is supported by molecular orbital theory.
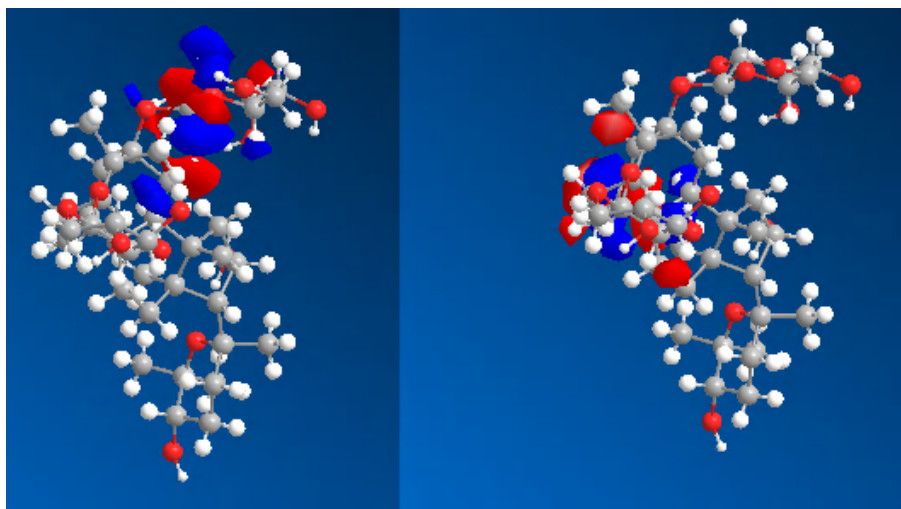
Figure 4.10 Compound cyclocephalosideI docked to NF-κB with AutoDock. On the left compound on the right with Hückel orbital surfaces are shown. Red as positive, blue as negative orbital surface.

4.2.3 m144(1st ranked in Gold Score)

Best ranked mode for m144 with Gold Score function represented as below in Figure 4.11.
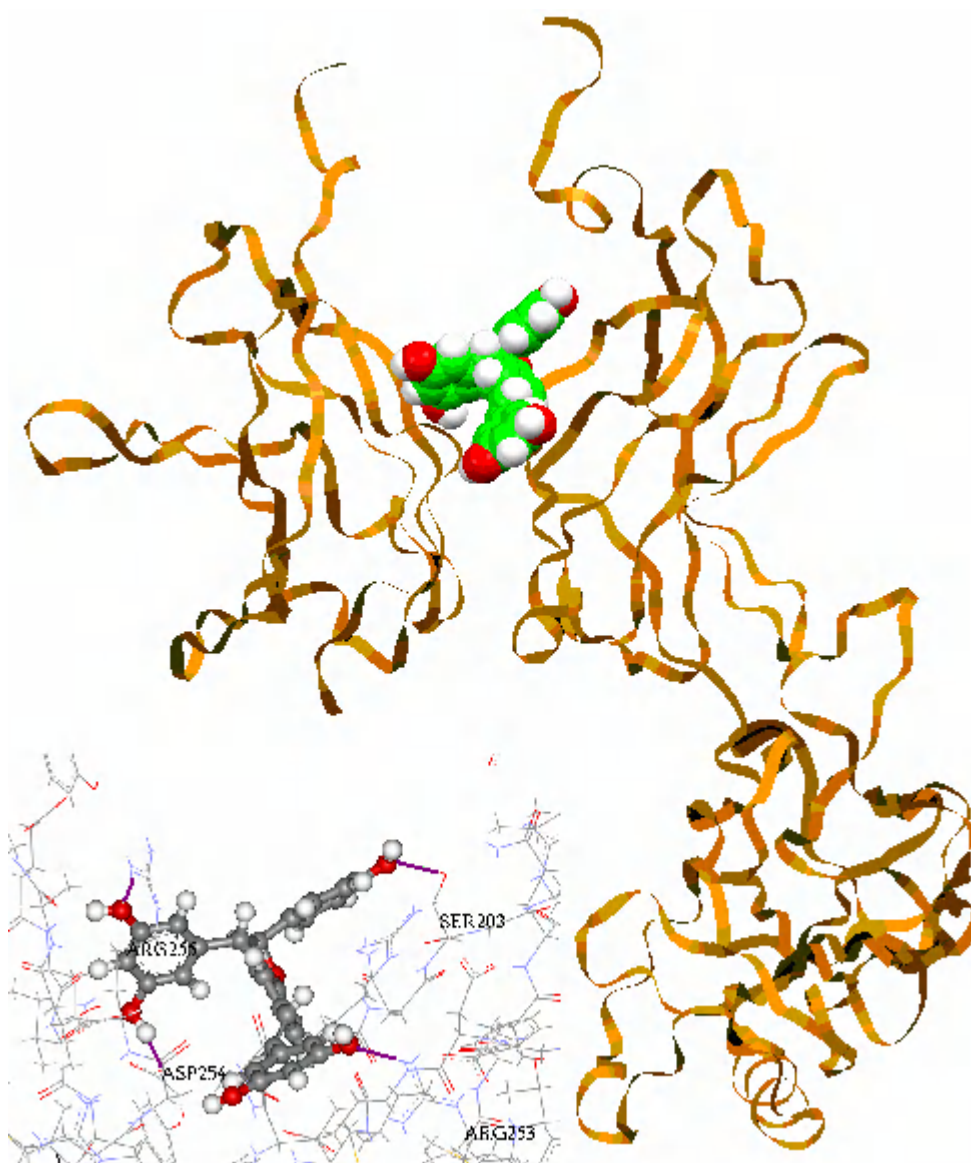
Figure 4.11 Docked NF-κB and m144 best binding mode. On the left corner; H-bonds within distance of 2.5A; four atoms on the NF-κB are shown as purple contact points. Contact points are on residues; Serine [203rd], Arginine[253rd], Aspartic acid[254th], Arginine[255th] and compound atoms namely O[27], O[55], H[2], O[10].

In this binding mode the compound m144 has four possible hydrogen bonds in the protein as shown in Figure13. Compound m144 is examined as one of the lipophilic stilbenoids in Chapter 3 and Appendix 4. From this group, m144 has the highest logD at pH 7.4 value of 6.5. Another lipophilicity indicating property is a Pho/phi_ASA value which is 2.12. Compound is highly lipophilic with highest second logD value of six lead compounds. Total solvent accessible area of the compound is 701 Å^2. This compound has five donor and six acceptor hydrogen bonding sites while in this binding mode with NF-κB three acceptors and one donor are near to hydrogen bonding.

Hückel charges on compound m144 is along with its frontier orbital calculated by extended Hückel. HOMO (Highest occupied molecular orbitals) are calculated with extended Hückel method in Figure 4.12. None of the specific hydrogen bonding shown in Figure 4.12 is originated from frontier orbitals of compound m144. Binding mode of m144 is inconsistent with molecular orbital theory.
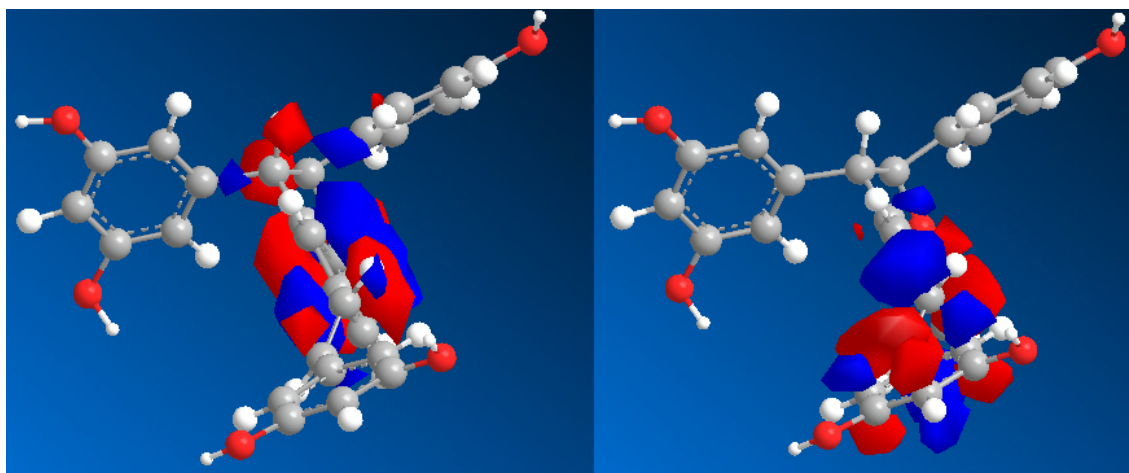
Figure 4.12 Compound m144 docked to NF-κB with Gold Score. On the left HOMO and LUMO of the compound on the rigth with Hückel orbital surfaces shown. Red as positive, blue as negative orbital surface.

4.2.4   m14(2nd ranked in Gold Score)

Best ranked mode for compound m14 with Gold Score function is given in Figure 4.13.
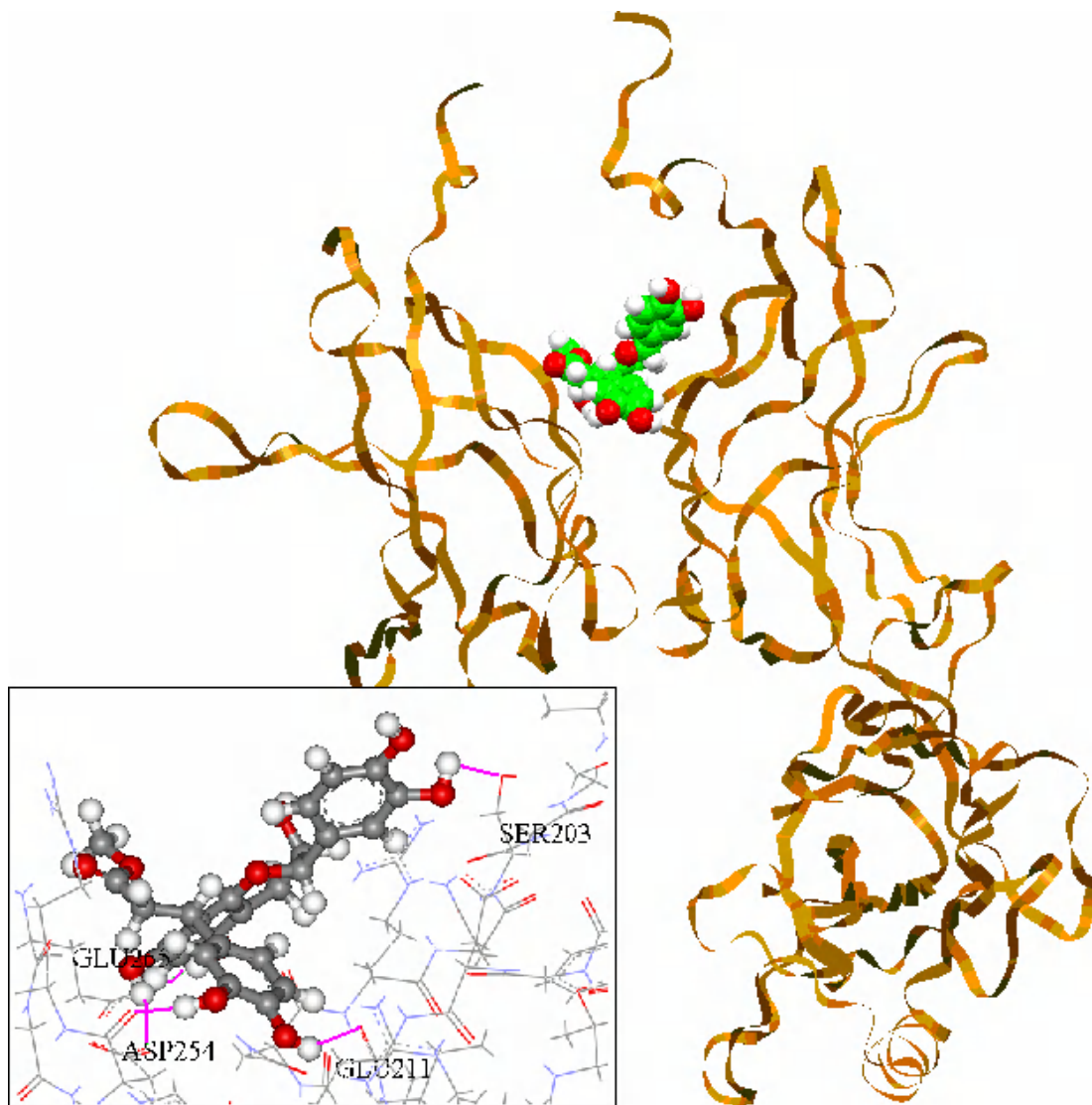
Figure 4.13 Docked NF-κB and m14 best binding mode. A close look is on the left corner; H-bonds within distance of 2.5A; four residues on the NF-κB as shown in purple contact points. Contact points are on residues; Serine [203rd], Glutamic acid[211th], two contacts on Aspartic acid[254th], Glutamic acid[265th].

The compound m14 has five possible hydrogen bonds with four different residues on NF-κB protein. All four hydrogen bonding sites are donors on compound m14 in Figure 4.14.

This flavonoid, m14, is lipophilic with a logD of 2.81at pH 7.4. It has a total solvent accessible area of 657 Å^2 while Pho/phi_ASA ratio is 1.54. This compound has lipophilic structure with positive logD value. The compound m14 has seven donor and nine acceptor hydrogen bonding sites while in this docked pose four donors are active.

Hückel charges on compound m14 are observed after docking. HOMO (Highest occupied molecular orbitals) and LUMO are calculated with extended Hückel method in Figure 4.14. As can be seen with the conformational hydrogen bond formed on LUMO of the compound with aspartic acid [254th]. Binding mode of the compound is consistent with molecular orbital theory.
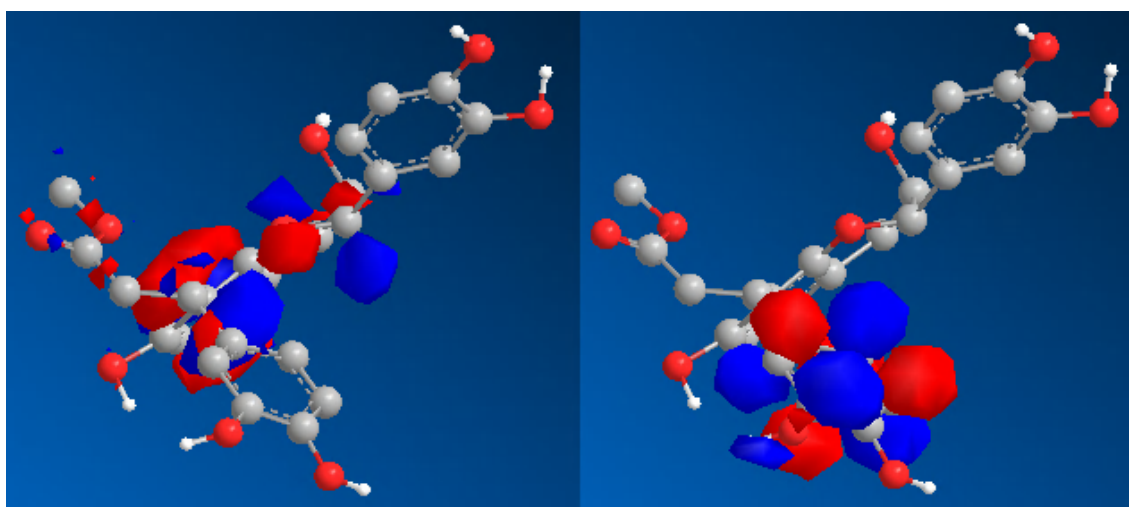
Figure 4.14 Compound m14 docked to NF-κB with Gold Score. On the left compound HOMO and LUMO of the compound on the rigth with Hünkel orbital sufaces are shown. Red as positive, blue as negative orbital surface.

4.2.5   m21(1st ranked in Chem Score)

Best ranked mode for compound m21 with Gold Score function is represented in Figure 4.15.
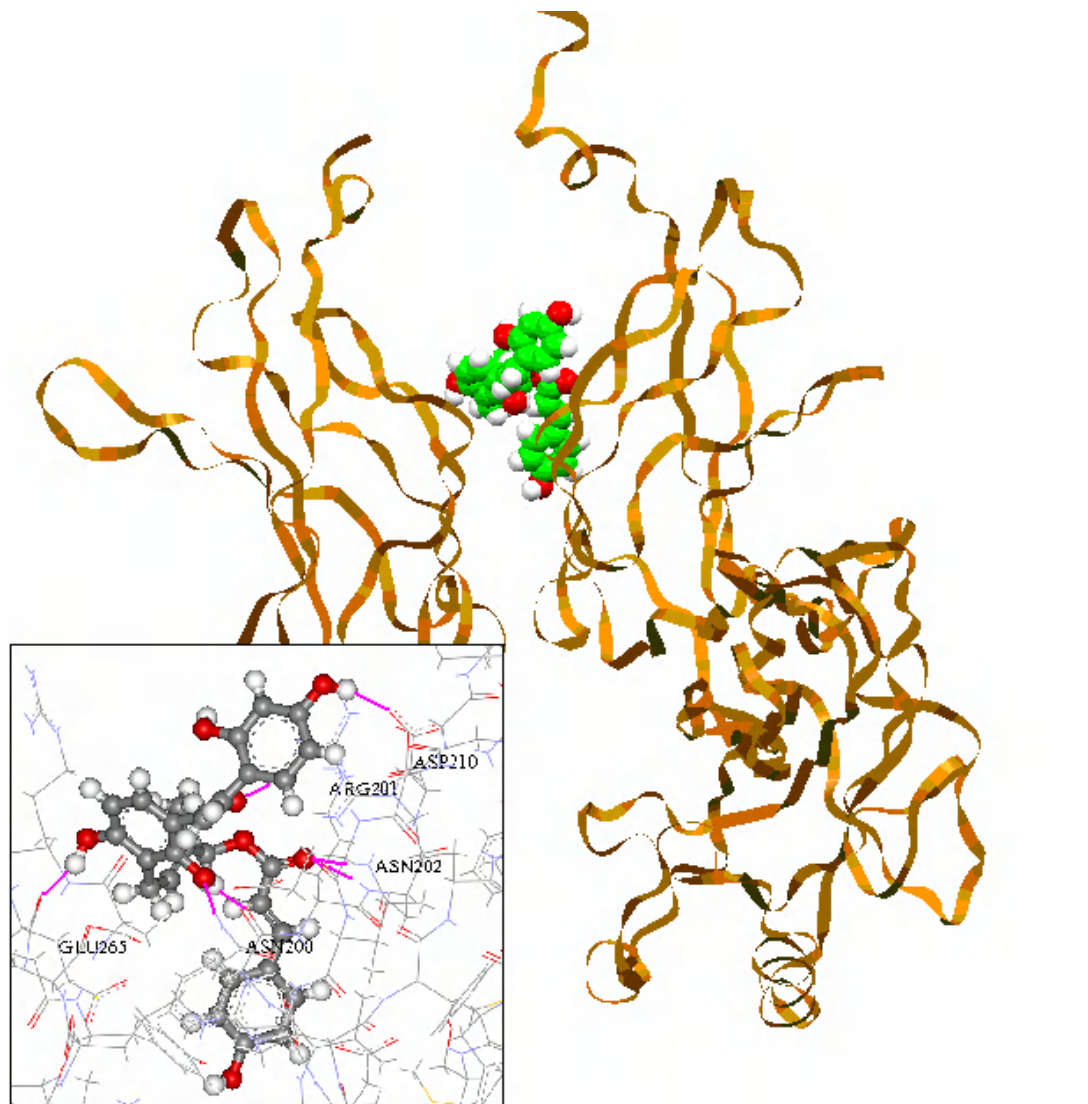
Figure 4.15 Compound m21 docked to NF-κB   with Chem Score and best binding mode achieved is shown above. On the left corner; H-bonds within distance of 2.5 Å; five residues on the NF-κB as shown in purple contact points. Contact points are on residues; Asparagine[200th], Arginine[201st], Asparagine[202nd], Aspartic acid[210th], Glutamic acid[265th]. Compound makes seven hydrogen bonds as shown with purple.

Compound m21 is a highly lipophilic flavonoid with a logD value of 6.65 at pH 7.4. And the compound has Pho/phi_ASA ratio of 2.62 with total solvent accessible area of 757 Å^2. The compound has five donor and seven acceptor sites in the native structure. Docked complex with m21 has seven hydrogen bonds built, one with Aspartic acid[210th], one with Arginine[201st], two with Asparagine[202nd], two with Asparagine[200th], and one with Glutamic acid[265th]. Compound m21 used three donor and three acceptor sites with this five contact sites on the protein. Figure 4.16 show the HOMO and LUMO calculated with extended Hükel method. LUMO of the compound is observed with hydrogen bonding with Aspartic acid[202nd]. HOMO of the compound is observed with hydrogen bonding with Arginine[201st]. Compound m21 has the highest total ASA as a result of having a complexity stucture. Because it is known that the more complex structure the beter drug candidate it is. Morevover, seven hydrogen bonds are built in this binding mode. The compound m21 has the highest lipophilicity of all six leads. Lipophilicity is needed for premeability of a drug, but high lipophilicity may cause non-specific bindings.
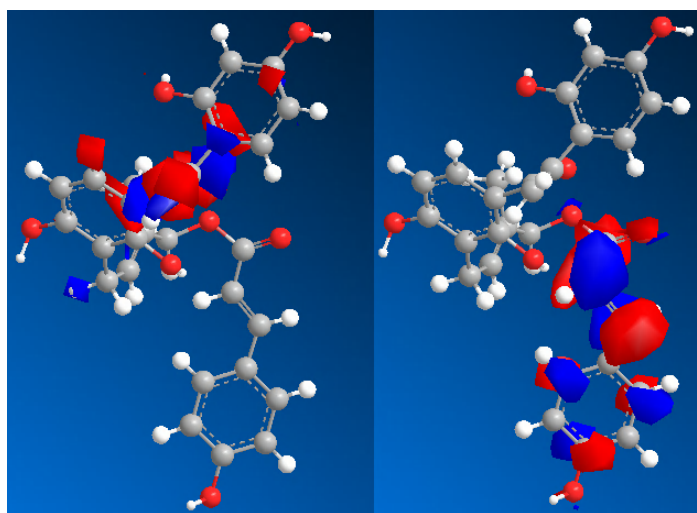
Figure 4.16 Compound m21 docked to NF-κB with Chem Score. On the left compound HOMO and LUMO on the rigth with Hünkel orbital sufaces are shown. Red as positive, blue as negative orbital surface.

4.2.6   m26(2nd ranked in Chem Score)

Best ranked mode for m26 via Chem Score function is given in Figure 4.17.
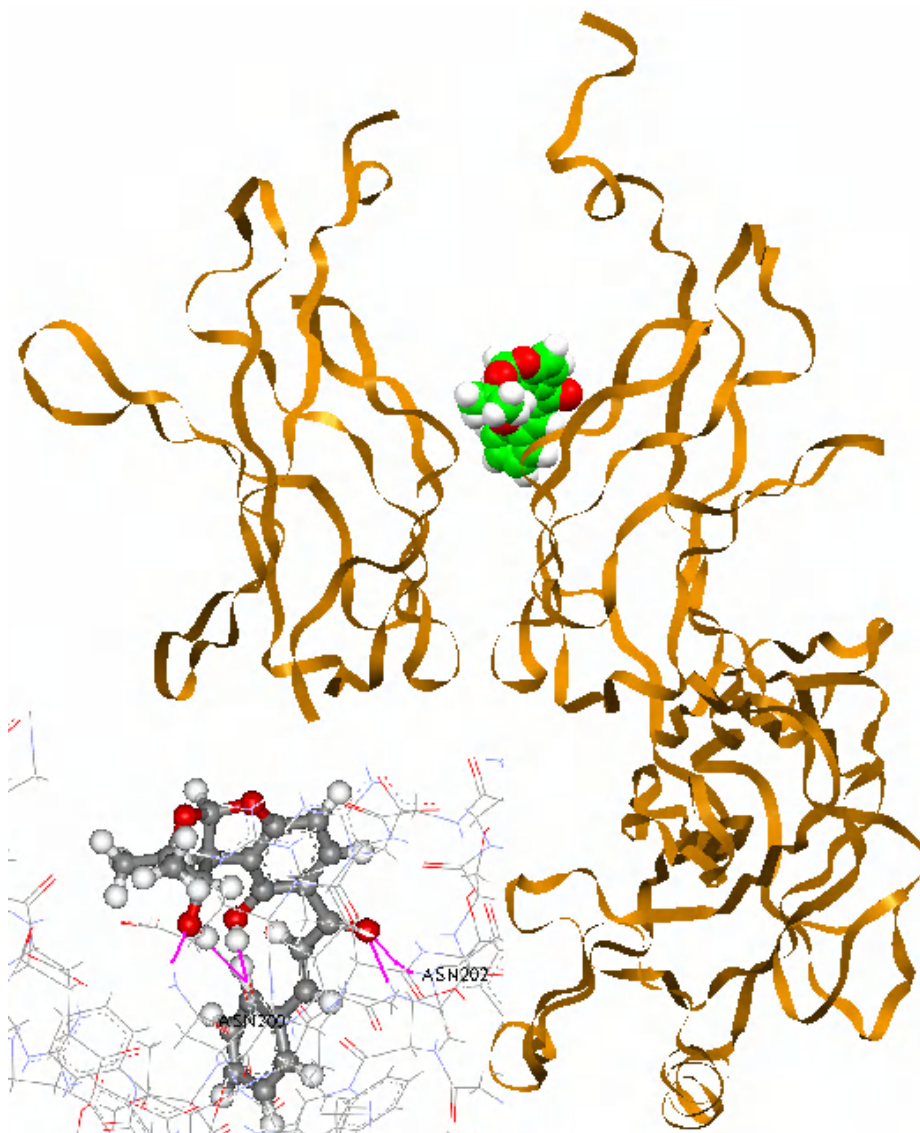
Figure 4.17 Compound m26 docked to NF-κB   with Chem Score and best binding mode achieved is shown above. On the left corner; five H-bonds within distance of 2.5A; two residues on the NF-κB as shown in purple contact points. Contact residues are both asparagines 200th and 202nd residues on NF-κB

.

Compound m26 is also a flavonoid lipophilic with a logD value of 3.55. Compound m26 has Pho/phi_ASA ratio of 5.26 with total accessible surface of 539 $Å^2$. In the native structure the compound has two donor and five acceptor sites while in the docked conformation two donors and two acceptors sites are used in hydrogen bonding. The frontier orbitals calculated with extended Hükel method is shownin Figure 4.18. There are hydrogen bonds formed between HOMO and LUMO of the compound and the NF-κB .
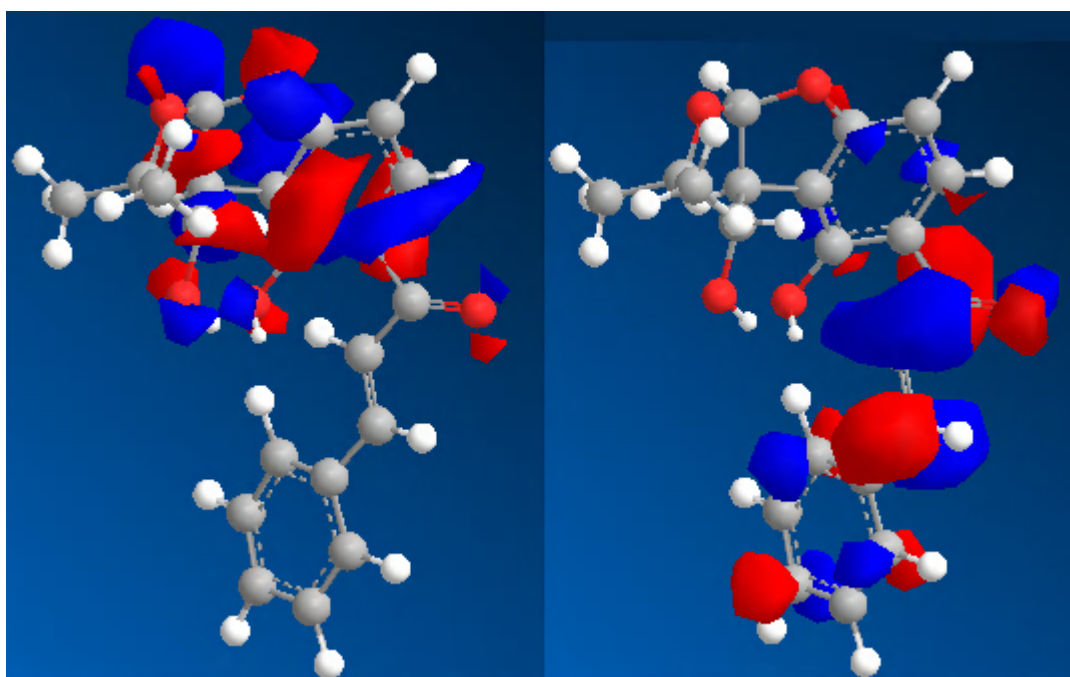
Figure 4.18 Compound m26 docked to NF-κB with Chem Score. On the left on the right with Hünkel orbital sufaces are shown. Red as positive, blue as negative orbital surface.

All compounds observed above are aligned inside the binding site of NF-κB as shown in Figure 4.19 below. Figure 4.20 shows that m26 and m21 are covering the binding site from the back and cyclocephalosideI, m144 are covering from front and m14 and m163 are covering the binding site from top. Molecular volume affects the covering areas.
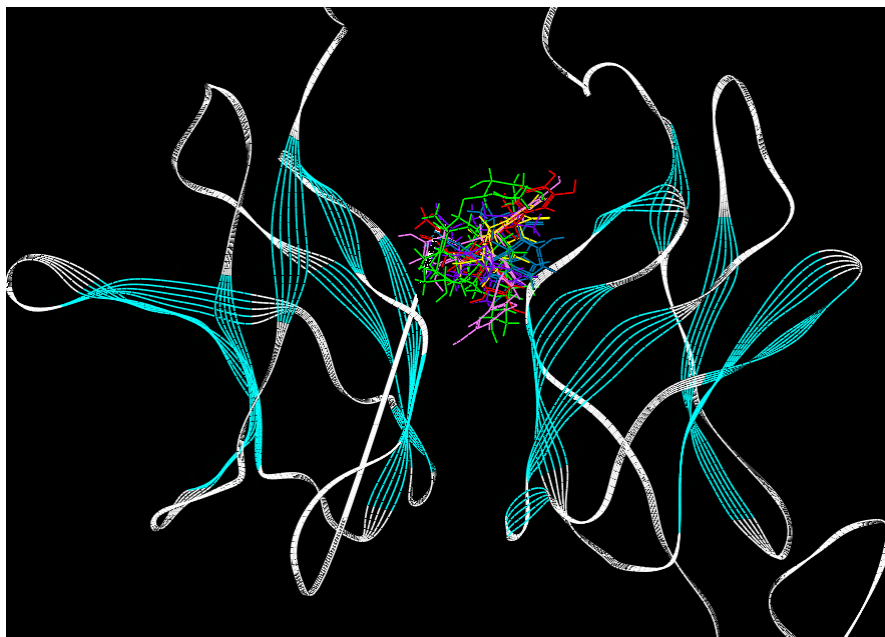
Figure 4.19 NF-κB   with best six compound aligned inside the binding site. Green refers to cyclocephalosideI, red refers to m14, blues as m21, yellow as m26, pink as m144, purple as m163.
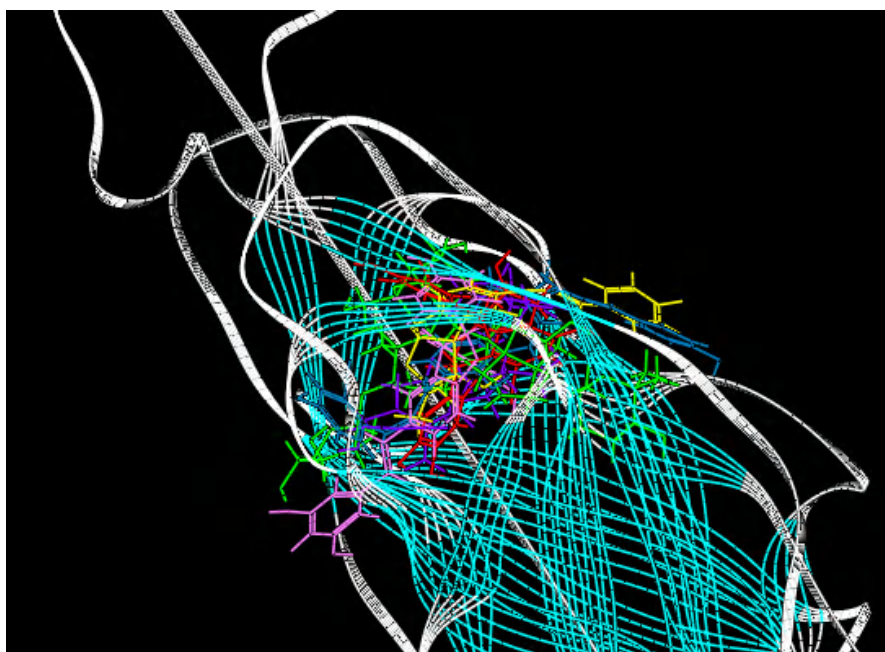
Figure 4.20 NF-κB with best six compound aligned inside the binding site.

Best binding modes of the top two scored compounds by Gold Score align in close proximity. And the best binding modes of top two scored compounds via Chem Score align similarly too. Aligned top two binding modes of each scoring function is shown in Figure 4.21. But the best binding modes of the top two scored compounds by AutoDock do not align.
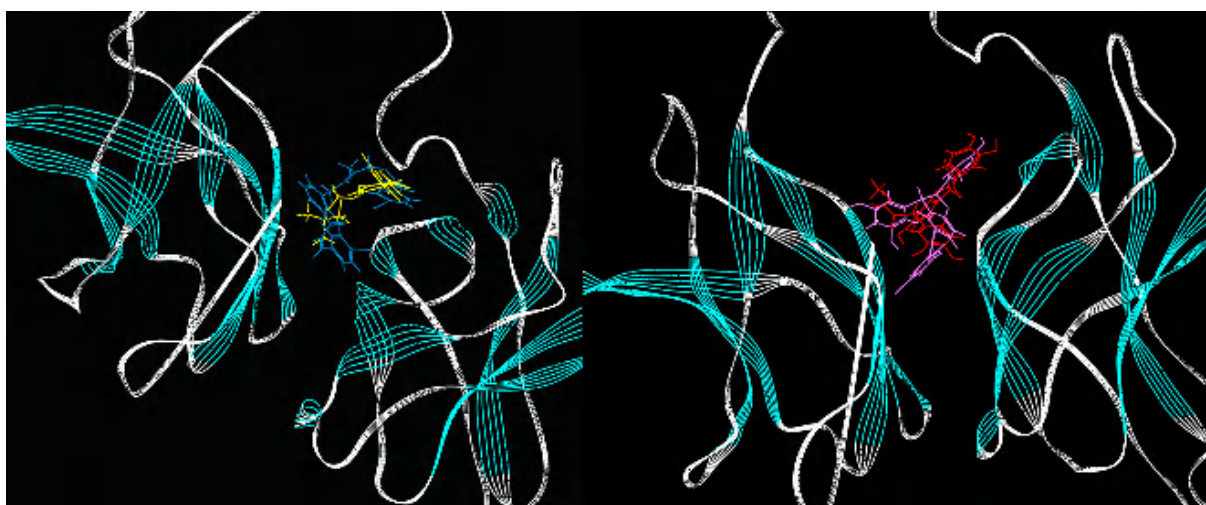
Figure 4.21 On the right Gold best results m14 and m144 are aligned; red and pink in the same order. On the left Chem Score function best results are aligned m21 and m26 as blues and yellow in the same order.

Comparison of compound properties: (i)Hydrogen bonds formed, (ii)Dispersion coefficient(logD), (iii)Total accessible solvent area, (iv)Hydrogen bonds formed on frontier molecular orbitals of six molecules are shown in Table 4.4. Compounds with lipophilicity, which is defined by the logD of around 2, are suitable for a drug and hydrogen bonds make the binding energetically more stable. Mostly hydrogen bonds built on binding modes are generated from the frontier orbital of the molecules.

Scoring function used	Compound Code	# of H-bonds	(LogD)	Total_ASA($Å^2$)	Hydrogen bonds on HOMO of compound	Hydrogen bonds on LUMO of compound
Au	cyclocephalosidel	6	-1.12	738	1(200)	—
Au	m163	1	2.46	536	1 (201)	—
Gold	m144	4	6.5	701	—	—
Gold	m14	5	2.81	657	—	1 (254)
Chem	m21	7	6.65	757	1(201)	1(202)
Chem	m26	5	3.55	539	1(200)	1(202)

Table 4.4 Compounds are compared with number of hydrogen bonds, logD and hydrogen bond origins in the table.

One compound from Table 4.4 is chosen for the comparison of the three scoring functions. Although three scoring functions rank one molecule differently the binding mode found from three dockings may be aligned well. Here m14 which is a Flavonoid derivative is chosen as a good example, because m14 is the most possible lead with its logD of 2.8 and it forms five hydrogen bonds with the target protein (NF-κβ). Compound m14 is ranked as the second in Gold Score with the highest fitness score (56.85). Compound m14 has Chem Score binding free energy of -15.4 kcal/mol whereas, it has AutoDock binding free energy of -7.05 kcal/mol. The binding modes found by the three scoring functions are aligned and shown in Figure 4.22.
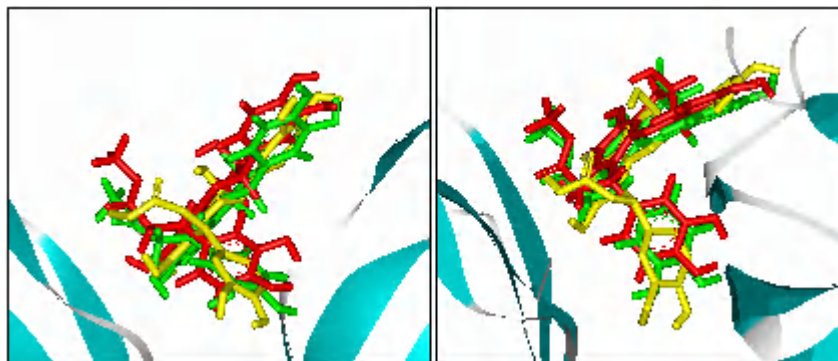
Figure 4.22 Binding modes of compound m14 are alligned in figure. Yellow mode is docked by AutoDock, green mode is docked by Gold Score, red mode is docked by Chem Score. front and top views are shown in both sides.

Three different scoring function docked results are aligned well although the ranking for the three functions are different. Compound best binding mode by AutoDock has four hydrogen bonds, best binding mode by Chem Score has seven hydrogen bonds and best binding mode with Gold Score has six hydrogen bonds. Compound m14 and NF-κB docking results along with hydrogen bonds shown in Figure 4.23, 4.24 and 4.25. Two amino acids (Serine [203rd], Glutamic acid [211th]) of NF-κB are common in the best binding modes generated by all three methods.
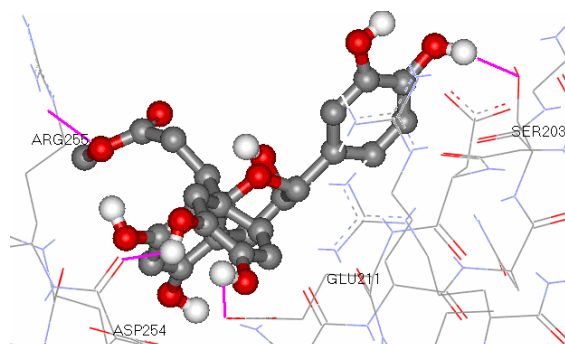
Figure 4.23 Best binding mode of m14 by AutoDock. Hydrogen bonds are Serine[203rd residue], Glutamic acid[211th], Ariginine[255th], and Aspartic acid[254th] on NF-κB .
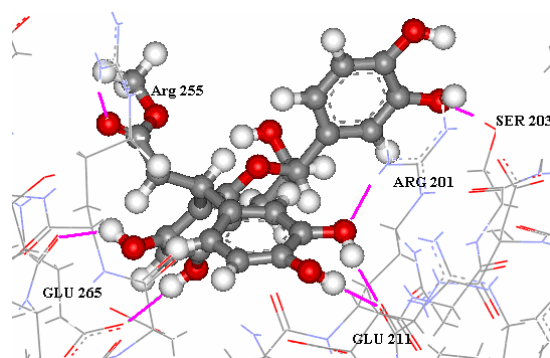
Figure 4.24 Best binding mode of m14 by Chem Score. Hydrogen bonds are Serine [203rd], Arginine[201st], Glutamic acid[211th], Glutamic acid[265th], and Arginine[255th] on NF-κB .
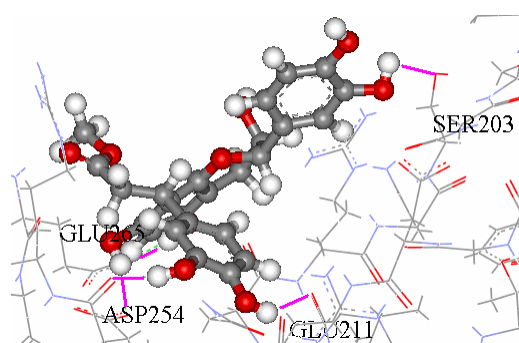
Figure 4.25 Best binding mode of m14 by Gold Score. Hydrogen bonds are Serine[203rd], Glutamic acid[211th], Glutamic acid[265th], and Aspartic acid[254th] on NF-κB .

Gold Score favors stilbenoids as lead compounds. More than 70% of the top ten molecules selected belong to the stilbenoid group, while the stilbenoid ratio for the total data set is 19%. Top ten molecules have high ligand internal score (S(int); including ligand van der Waals energy and ligand internal torsional term scores) and protein-ligand van der Waals energy score (Svdw-ext). Average S(int) is -19,9 and average S(vdw-ext) is 33.27. Values that are higher thean the average values are accepted as high. Among 46 Stilbenoid compounds, 89% of the molecules have high S(int) and 76% of the molecules have high S(vdw-ext). These two components of Gold Score are used to select the best molecules among stilbenoids. Two components of the scoring function, are highly effective on the best ranked compounds, which are based on both ligand itself and protein-ligand pair properties. In conclusion, Gold Score selection is based on the protein-ligand pair. Hence, selectivity of Gold Score is on the specific system present.

On the other hand AutoDock favors cycloartane molecules. Among the top ten molecules selected, the cycloartane triterpenoids ratio is over 70%, while cycloartane triterpenoids for the total data set is 15%. Top ten molecules have low final internal energies for the ligands. Low torsional energy and high unbound energy is beneficial for low binding free enegy but not necessary as observed from the data. Although cycloartane molecules have high torisonal energy and low unbound energy molecules are still the top ranking for AutoDock. Average final internal total free energy of ligands is -1.36 kcal/mol. Seven of the top ten molecules have low final internal total free energy so does cycloartane triterpene molecules. Cycloartane triterpenoids rank as the best group of molecules with this property. Over 36 cycloartane triterpenoids 86% of the molecules have low final internal total energy. One components of the energy equation, is highly effective on the best ranked compounds, based on the ligand itself. Since the final internal free energy of ligands has an important role in the overall evaluation, we can conclude that AutoDock has the selectivity for a specific group of ligand structure based on the ligand property itself, not on the property of the system (ligand-protein pair).

In conclusion we observed that scoring and ranking over a data set may differ for each scoring function although for a specific ligand and protein couple all search algorithms suggest similar translocation and orientations of binding that can be easily aligned [5].

Chapter 5

OTHER SYSTEMS, CALCULATIONS AND RESULTS

In this thesis two systems are used for docking studies; the first system is NF-κB and the second system is Fgfr2 protein which is a target for cancer therapy [1], [2]. Docking studies with the same data set and another protein are compared to observe the specificity of the ligands and the selectivity of the scoring functions.

5.1   Introduction; Fgfr2

Fgfr2 (Fibroblastic growth factor-2) is one of the target proteins for cancer therapy. Misregulation of Fgfr2 is observed in breast, lung, ovary and stomach cancer patients. Fgfr2 is a receptor of FGF-2(Fibroblastic growth factor). When the Fgfr2 and FGF-2 proteins are bounded the resulting complex is internalized and translocated into the nucleus. Translocation of FGF-2 complexes mediates pro-survival and anti-apoptotic activities. Fgfr2 is a cell membrane protein with three main domains.
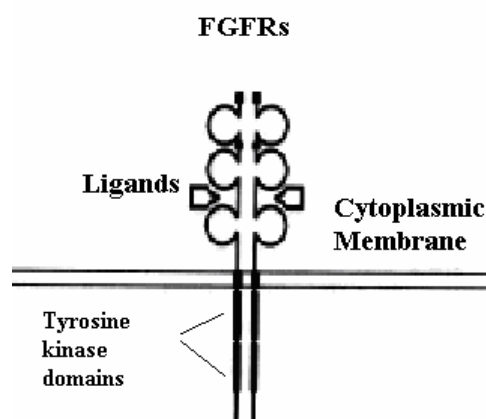
Figure 5.1 Fibroblastic growth factor receptor family proteins are demonstrated as above, consist of 3 main domains as shown.

Fgfr2 consists of extracellular region, transmembrane region and cytoplasmic tyrosine kinase, i.e. TK, region. The extracellular part of the protein interacts with FGF. As the binding occurs, two FGF and two Fgfr2 form a tetramer. After dimerization; the autophosphorylation and signaling is initiated. Signals trigger regulation of cell growth, cell differentiation, cell migration, wound healing and angiogenesis.

Fgfr2 mutations cause severe syndromes or cancer. Mutation in Fgfr2 makes the FGF-Fgfr2 interaction stronger and as a result permanent signaling of differentiation and growth occurs. Drug development targeting the binding and the mutations are of interest for cancer researchers. Drugs against cytoplasmic region containing TK have been developed but this causes severe side effects because other TK containing proteins are also affected. Since, targeting the extracellular domain is more specific, consequently we targeted extracellular domain as a strategy in this thesis.

The three dimensional structure of Fgfr2 is obtained from protein data bank with the code 1ii4. 1ii4 is a complex crystallized Fgfr2 with Heparin-binding growth factor 2. The protein is Homo *sapiens* derived sequence, expressed on Escherichia *coli* expression system. The crystal structure of the protein is a S252W mutant. This mutation is reverted changed back to the wild type form. The mutation is on 252nd residue, and it is changed into the original form as serine. The binding pocket involves 242nd to 262nd and 282nd to 285th residues. The calculations are made over this new model. The target binding site is presented in Figure 5.2.
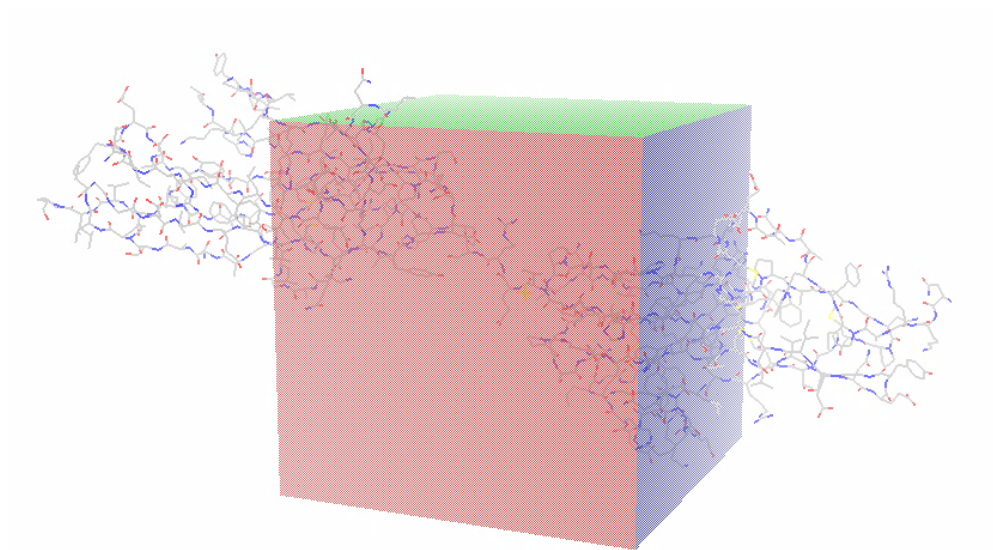
Figure 5.2 Fgfr2 Binding site on D3 domain.

5.2   Fgfr2 Results and their Comparison with results of NF-κB

5.2.1   Fgfr2 Results

The data set of 236 natural compounds are docked to Fgfr2 similar to the procedure for NF-κB by three different docking algorithms; Gold Score, Chem Score and AutoDock. The top ten ranked molecules of each algorithm are listed below. All twenty seven groups of structures in the data set are docked with protein Fgfr2. Contrary to the docking results for NF-κB, Fgfr2 with Gold Score hit compounds are in a variety of groups among the data set. However, seven out of the ten Chem Score hits are in the stilbenoids group, and again six of the ten AutoDock hits are in the cycloartane tirterpenoids. Thus, the two scoring functions, Chem Score and AutoDock, have the tendency to select specific molecular structure groups.

The AutoDock scoring function ranked cycloartane triterpenoids selectively against Fgfr2. AutoDock ranks the same structural ligand group selectively with the same data set and different proteins. AutoDock has the tendency to select the ligand itself but not according to the system.

AutoDock	Gold Score	Chem Score
Elongatoside *(cycloartane triterpenoid)*	vaticanolC *(resveratrol derivative)*	Pterostilbene-dehydrodimer *(resveratrol derivative)*
TrojanosideC *(cycloartane triterpenoid)*	hopeaphenol *(harmol alkaloid)*	res32 *(resveratrol derivative)*
TrojanosideJ *(cycloartane triterpenoid)*	m144 *(stilbenoid)*	AcetylastragalosideI *(cycloartane triterpenoid)*
m14 *(flavonoid)*	m123 *(lignan)*	res27 *(resveratrol derivative)*
MacrophyllosaponinB *(cycloartane triterpenoid)*	Resrevatrol-trans-dehydrodime *(resveratrol derivative)*	m23 *(flavonoid)*
CyclocephalosideI *(cycloartane triterpenoid)*	m76 *(flavonoid)*	res20 *(resveratrol derivative)*
m163 *(withanolide)*	m100 *(flavonoid)*	Resrevatrol-trans-dehydrodimer
m149 *(withanolide)*	m67 *(benzofuran)*	res16 *(resveratrol derivative)*
TrojanosideB *(cycloartane triterpenoid)*	m125 *(lignan)*	res31 *(resveratrol derivative)*
m41 *(stilbenoid)*	m80 *(flavonoid)*	res25 *(resveratrol derivative)*

Table 5.1 Top ten docking results against Fgfr2; Chem Score top compounds are ranked with their binding free energy results, Gold Score top compounds are ranked for their fitness score, AutoDock top compounds are ranked with their binding free energies calculated.

While Gold Score has no significant tendency to select any structural groups, Chem Score has tendency to select resveratrol structures as hits for target protein Fgfr2. Chem Score has eight resveratrol derivatives among ten hits for Fgfr2. As can be seen from Table 5.2 average result hits Chem Score has the same trend to select stilbenoids also. AutoDock shows two different trends for best result hits and averaged result hits as observed in other system shown in Chapter

4. AutoDock averaged results hits are withanolides and norwithanolides (which are also derived from withanolide) structures by a majority.

Chem Score Average	Gold Score Average	AutoDock Average
m29 *(norwithanolide)*	m123 *(lignan)*	m14 *(flavonoid)*
m23 (flavonoid)	Resrevatrol-trans-dehydrodimer *(resveratrol derivative)*	Elongatoside *(cycloartane triterpenoid)*
Pterostilbene-dehydrodimer (resveratrol derivative)	m67 (benzofuran)	AstrasieversianinXV *(cycloartane triterpenoid)*
res13 (resveratrol derivative)	vaticanolC *(resveratrol derivative)*	TrojanosideJ *(cycloartane triterpenoid)*
m72 (diarylheptanoid)	m100 *(flavonoid)*	m163 *(withanolide)*
res27 (resveratrol derivative)	res16 *(resveratrol derivative)*	m46 (withanolide)
res30 (resveratrol derivative)	m115 (flavonoid)	m45 (withanolide)
res31 (resveratrol derivative)	m76 (flavonoid)	m47 (withanolide)
m34 (norwithanolide)	res34 *(resveratrol derivative)*	m160 (withanolide)
res9 (resveratrol derivative)	m80 (flavonoid)	m29 (norwithanolide)

Table 5.2 Average values are calculated with for all runs done for each compound docked with Fgfr2. C. S (Chem Score), G. S. (Gold Score), A. (AutoDock).

According to the results observed in Chapter 4, the best results obtained from three scoring functions are differ from each other. And again there is no correlation between best results of the three scoring functions. In Figures 5.3, 5.4, and 5.5 we search for an overall correlation on the full data set. Charts are non-scaled and no significant correlation between scoring function results is observed. Gold Score vs Chem Score graph, shown in Figure 5.3, has better linear regression (R^2=0.251) than other best correlation graphs with the best results. Averaged results are shown in Figures 5.7, 5.8, and 5.9 with variable linear regressions. Average results of Gold Score and Chem Score graphs show a more significant linear regression (R^2=0.513) while other correlation graphs with the averaged scoring are not significant. Plots below show the correlations obtained

by different scoring algorithms applied to all data set consisting of 236 compounds.
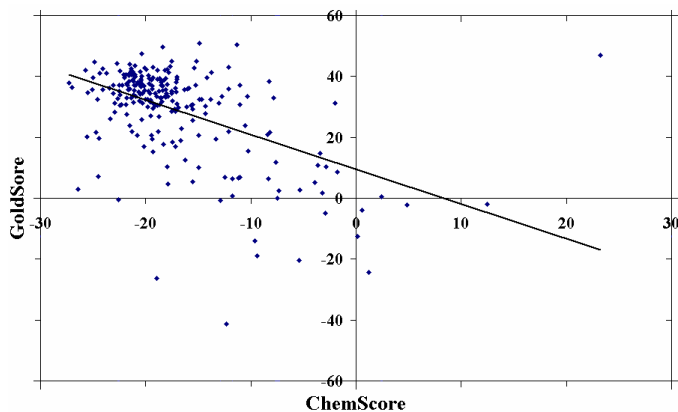
Figure 5.3 Results in AutoDock binding free energy and Gold Score fitness are compared in graph. Linear regression is not significant; R^2: 0.251.
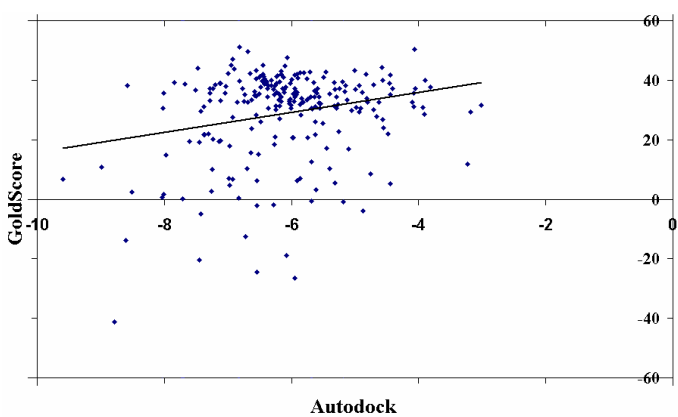
Figure 5.4 Results in AutoDock binding free energy and Gold Score fitness are compared in graph. Linear regression is not significant; R^2: 0.0553.
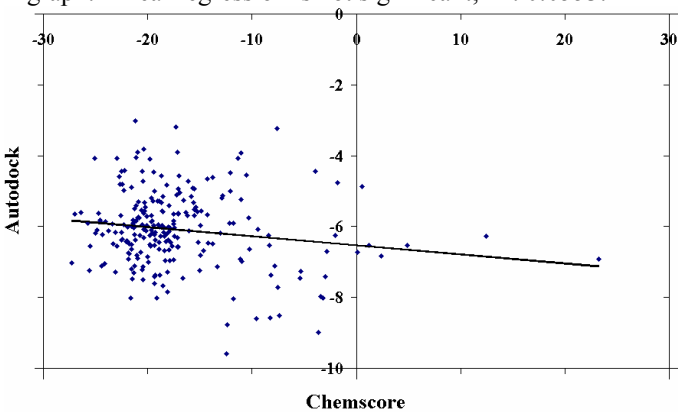
Figure 5.5 Results in AutoDock binding free energy and Chem Score fitness are compared in graph. Linear regression is not significant; R^2: 0.0262.
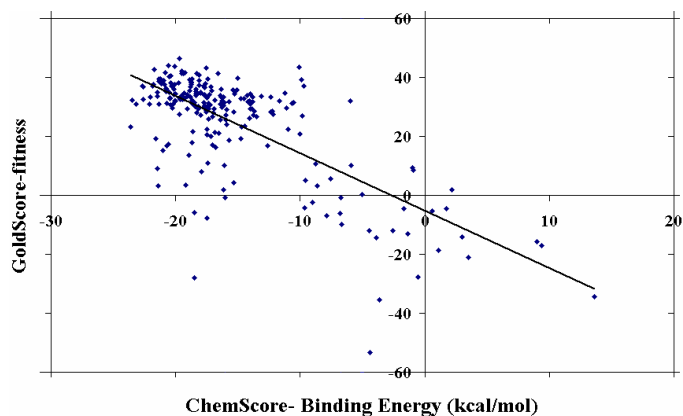
Figure 5.6 Average results in Chem Score binding free energy and Gold Score fitness are compared in graph. Linear regression is significant; R^2: 0.5134.
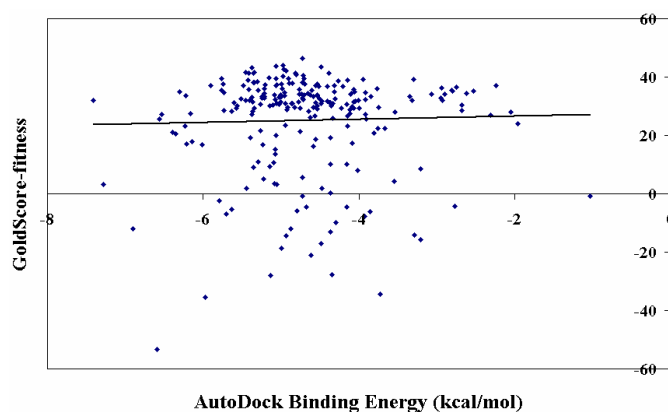
Figure 5.7 Average results in AutoDock binding free energy and Gold Score fitness are compared in graph. Linear regression is significant; R^2: 0.0009.
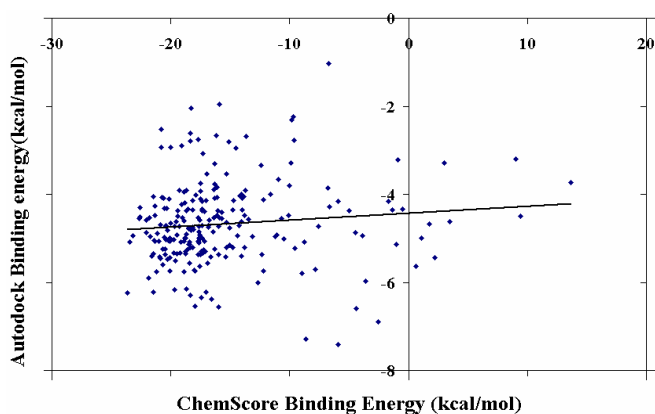
Figure 5.8 Average results in Chem Score binding free energy and AutoDock fitness are compared in graph. Linear regression is significant; R^2: 0.0108

Plots above shows the correlation of the algorithms applied to all data set with average values. As in Chapter 4, the best docking results for Fgfr2 obtained from three scoring functions are different from each other as can be seen from correlation plots. Here, we take a closer look at the best two results of each scoring function in order to have a better understanding. Investigating the best results individually will give us insight to determine which tool and scoring function is the most suitable for our data set and new system with Fgfr2. Detailed results of selected compounds are demonstrated in the rest of this chapter.

Results of frontier orbital analysis of Fgfr2 are presented in Figure 5.10. Frontier Orbital calculations are done on the binding pocket shown in Figure 5.9. Lowest unoccupied molecular orbital (LUMO) and highest occupied molecular orbital (HOMO) of 23 residues are calculated with Hückel charges by ChemBioOffice software 2008 [68]. Possible hydrogen bonding between the ligand and protein could be predicted from the frontier orbital of the binding pocket.
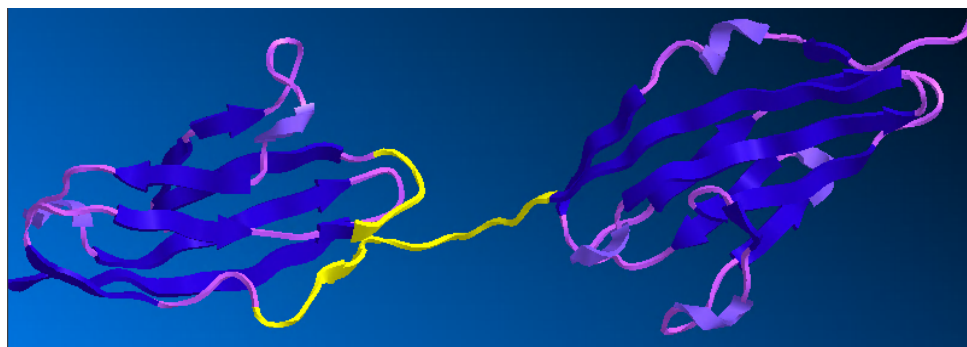
Figure 5.9 Fgfr2 binding pocket including 23 residues which are anlyzed for frontier orbital.
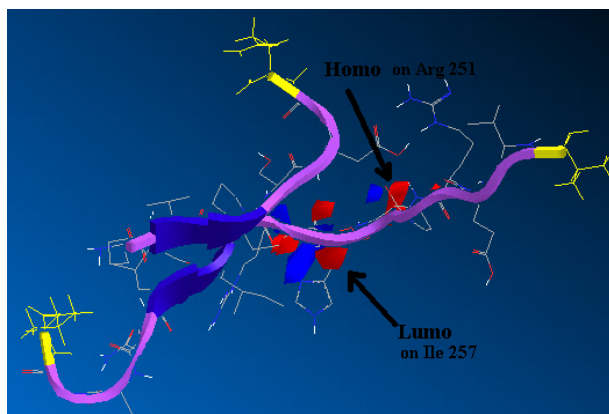
Figure 5.10 Fgfr2 binding pocket is anlysed with Hückel charges. Frontier orbital are shown both for LUMO and HOMO. HOMO is on Serine [252nd] and Arginine[251st] with -10.22 eV, LUMO is on Isoleusine [257th] with -1.07 eV.

5.3 Detailed Results of Selected Compounds

Top two ranking compounds are selected for each scoring function. But, unlike our strategy for the NF-κB, the top compounds are selected even if they are not ranked in all three scoring lists. The most important criteria to be selected picked as a lead is not to be a lead for the other protein. Compound TrojanosideC (2nd ranked in AutoDock), m149 (9th ranked in AutoDock), Compound res32 (2 nd ranked in Chem Score), Pterostilbene-dehydrodimer (1st ranked in Chem Score), compound m123 (4th ranked in Gold Score), compound m100 (7th ranked in Gold Score) are selected and details are analyzed in the following pages. Table 3 gives all results related to the six molecules selected.

Molecule Name	AutoDock binding free enegy (kcal\mol)	(AutoDock) Average binding energy	Gold Score fitness	(Gold Score) Average fitness	Chem Score binding energy kcal\mol	(Chem Score) Average binding energy
gm149	**-8.01**	-6.17	35.71	27.36	-21.63	-19.69
gTrojanosideC	**-8.99**	-5.63	10.87	-5.37	-3.64	0.60
gres32	-6.90	-5.60	37.78	31.18	**-26.47**	-20.7
gPterostilbene-dehydrodimer	-6.43	-4.94	38.75	30.83	**-28.43**	-23.2
gm123	-6.07	-4.72	**47.46**	46.40	-22.51	-19.69
gm100	-6.45	-5.37	**44.96**	43.15	-20.28	-17.52

Table 5.3 Two hits of each docking tool are shown.

5.3.1 Compound m149 (9th ranked in AutoDock)

After docking best binding mode with compound m149 has three hydrogen bonds build between ligand with Leusine, serine and Arginine residues.
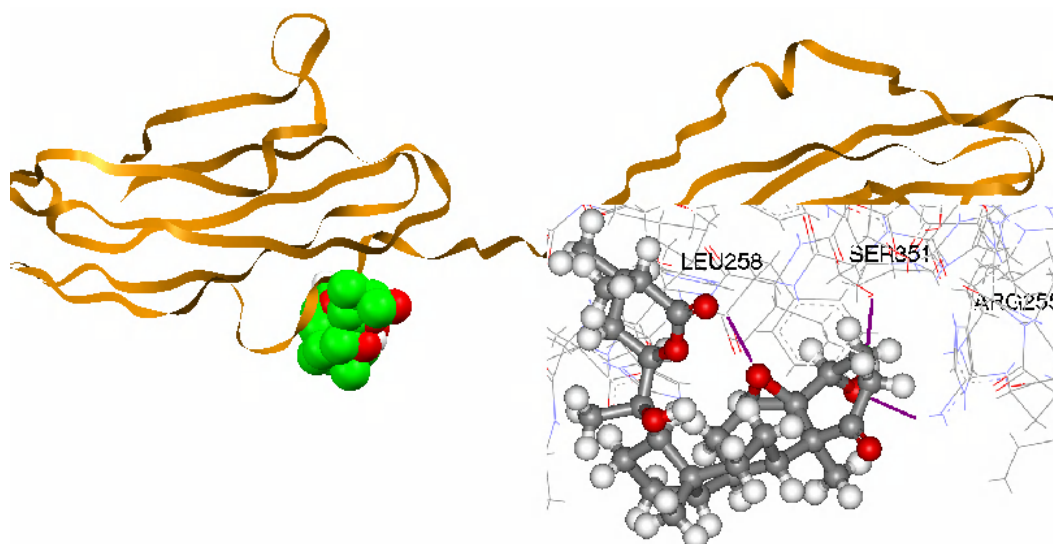
Figure 5.11 Docked Fgfr2 and Pterostilbene best binding mode. A close look is on the right corner; H-bonds within distance of 2.5A; three residues on the Fgfr2 as shown in purple contact points. Contact points are on residues; Leusine[258th], Arginine [255th] and Serine[351st].

Compound m149 is a Withanolide compound with 2 donors and five acceptors. AutoDock scoring function docked the structure with Fgfr2 and hydrogen bonds built by two acceptors and a donor of the molecule. Molecule has a lipophilic structure with positive logarithmic dispersion parameter (logD) value of 3.44 and Pho/phi_ASA ratio of 6.33. Compound with logD values over 2 are highly lipophilic. Lipophilicity is needed for permeabilitiy of a drug but high lipophilicity may cause non specific bindings. But compound m149 is not the highest lipophilic structure among six leads. Moreover, compound m149 has a binding mode of strong interactions with the highest number of hydrogen bonds built among six leads.

5.3.2 Compound TrojanosideC (2nd ranked in AutoDock)

TrojanosideC have one hydrogen bond with Isoleusine (257th residue) of Fgfr2.
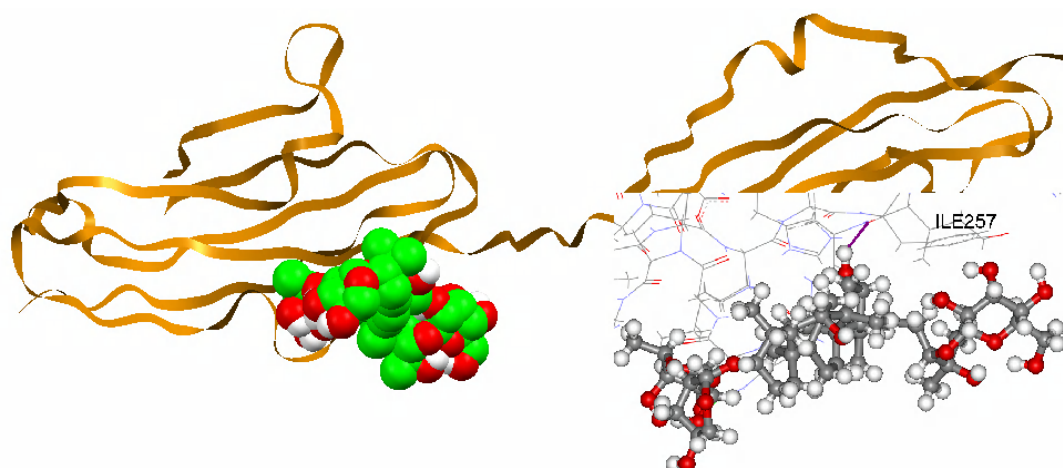
Figure 5.12 Docked Fgfr2 and Pterostilbene best binding mode. A close look is on the right corner; H-bonds within distance of 2.5A; just one residues on the Fgfr2 as shown in purple contact points. Contact points are on residues; Isoleusine[257th].

TrojanosideC is a Cycloartane derivative with twelve acceptors and eighteen donors. A donor builds a hydrogen bond with the protein. The molecule is hydrophilic with negative logD and apolar/polar solvent accessible area ratio (Pho\phi_ASA) of 2.81. Lipophilic structure is better for permeability of a drug than hydrophilic structure. TrojanosideC has a binding mode of least connectivity among six lead compounds with only one hydrogen bond. TrojanosideC may not be a good choice for a drug candidate but AutoDock scoring function rankes as the second best bound ligand among 236 compounds.

5.3.3 Compound m100 (7th ranked in Gold Score)

Compound m100 docked to Fgfr2 with Gold Score. Best binding mode has two hydrogen bonds.
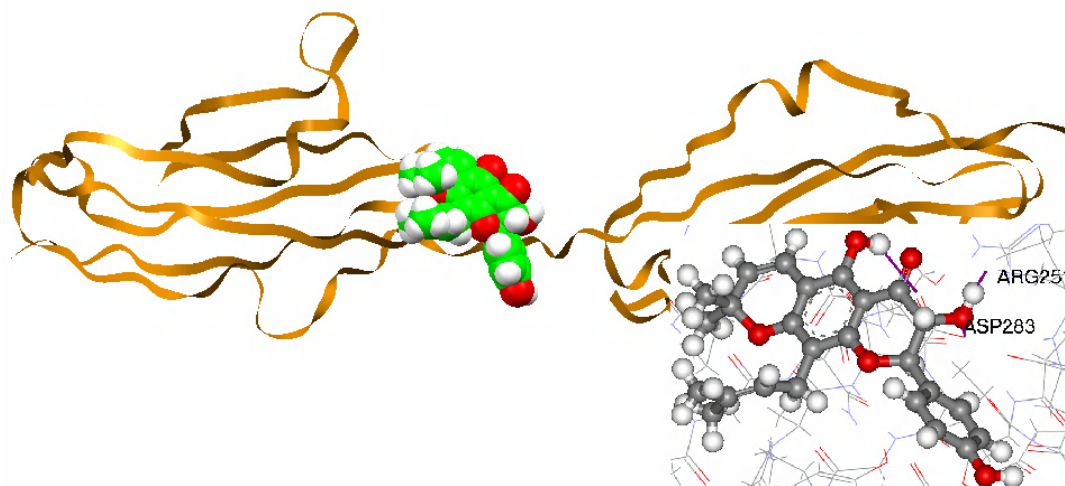
Figure 5.13 Docked Fgfr2 and m100 best binding mode. A close look is on the right corner; H-bonds within distance of 2.5A; two residues on the Fgfr2 as shown in purple contact points. Contact points are on residues; Arginine [251st], Aspartic acid [283rd].

Compound m100 is a Flavonoid with three acceptors and six donors. Hydrogen bonds build with two residues; Arginine and Aspartic acid with three contacts. Molecule is a lipophilic compound with positive logD (4.49) and Pho/phi_ASA ratio of 3.53. High lipophilic structures may have non-specific binding but compound has three hydrogen bonds built in this binding mode. Compound m100 may be a good candidate drug molecule for Fgfr2.

5.3.4 m123(4th ranked in Gold Score)

Three H bonds with two residues. Compound m123 is lignan structure. Lignans in our data set are lipophilic structures as observed in Chapter 3. Structure has logD value of 3.12 and Pho\phi_ASA ratio of 4.18.
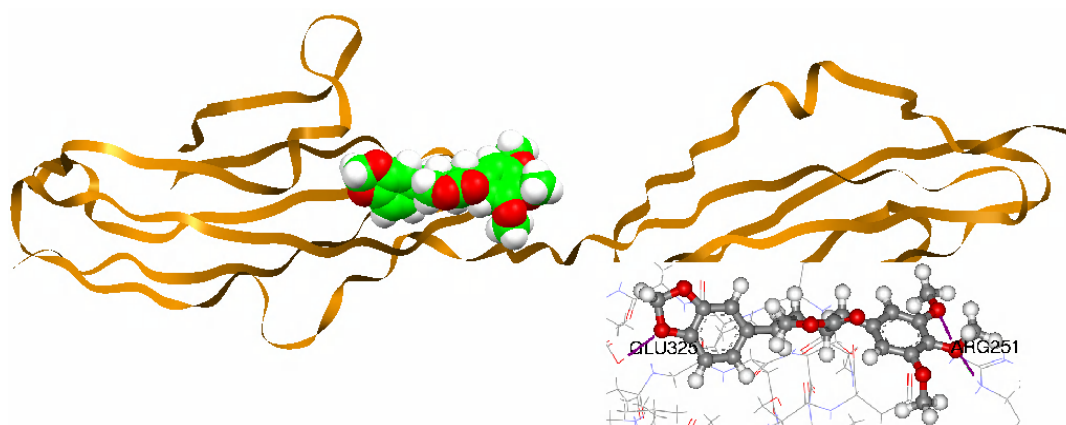
Figure 5.14 Fgfr2 and m123 best binding mode. On the right corner; hydrogen bonds with residues (Arginine [251st], Glutamic acid [325th]) on the Fgfr2 shown in purple contact points.

Molecule has six donors and no acceptor. Hydrogen bonds are built two residues on Fgfr2 with compound m123. Keeping in mind that high lipophilicity may cause non-specific, compound m123 is a good candidate drug molecule for Fgfr2.

5.3.5 Compound res32 (2nd ranked in Chem Score)

Compound res32 is docked with Fgfr2 with Chem Score scoring function.
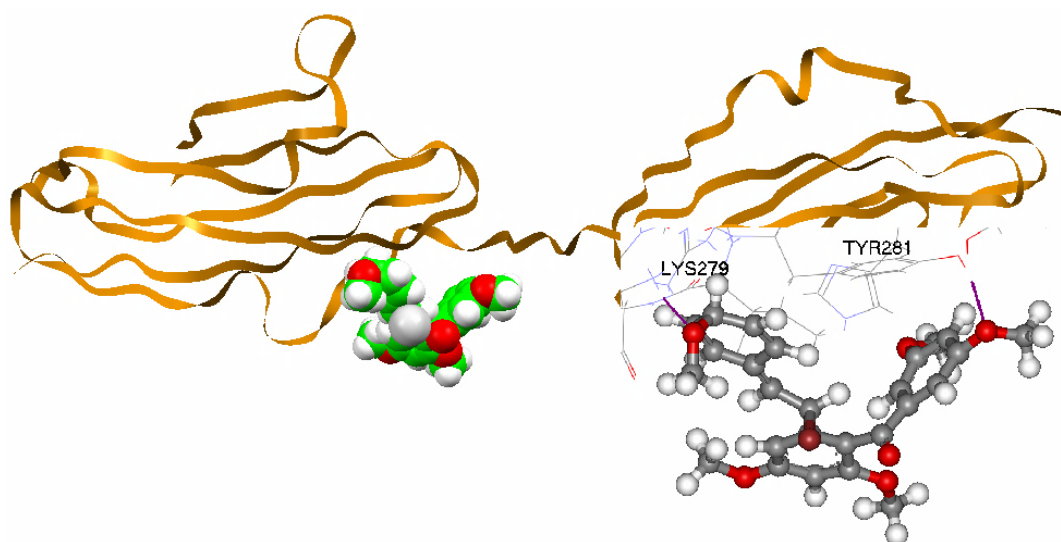
Figure 5.15 Docked Fgfr2 and compound res32 best binding mode. A close look is on the right corner; hydrogen bonds within distance of 2.5A; two residues on the Fgfr2 as shown in purple contact points. Contact points are on residues; Lysine [279th], Tyrosine [281st].

Two hydrogen bonds are build with two 279th and 281st residue on protein shown in Figure 5.15. Compound res32 is a resveratrol derivative, a lipophilic structure having logD value of 3.92 and Pho\phi_ASA ratio of 8.22. Six of the two donors on molecule build hydrogen bonds. Resveratrol derivative res32 has the highest Pho\phi_ASA ratio among six lead compounds while not having the highest logD. Structure is highly lipophilic and has a large solvent accessible surface area which allows larger interaction surface are.

5.3.6 Pterostilbene-dehydrodimer (1st ranked in Chem Score)

Pterostilbene-dehydrodimer is docked with Chem Score scoring function by GOLD program. Best binding mode is built with three hydrogen bonds shown in Figure 5.16.
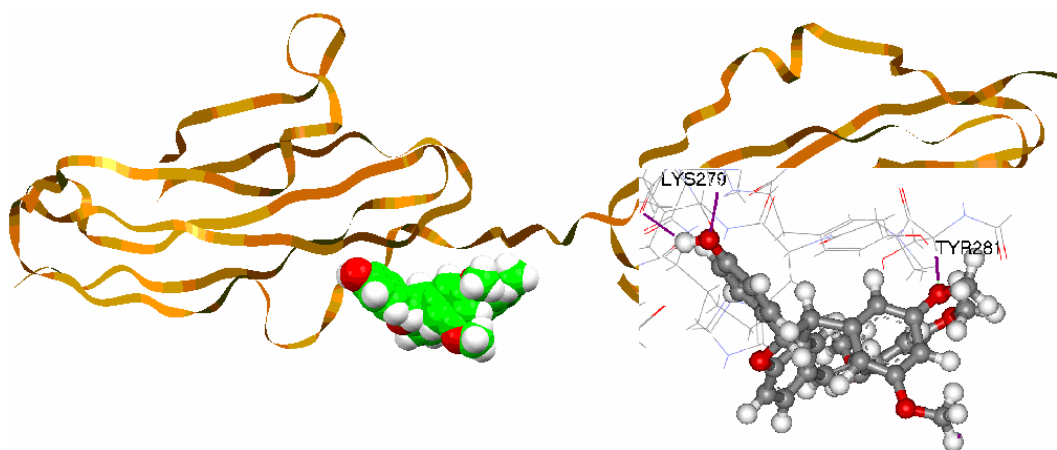
Figure 5.16 Docked Fgfr2 and Pterostilbene-dehydrodimer best binding mode. A close look is on the right corner; H-bonds within distance of 2.5A; two residues on the Fgfr2 as shown in purple contact points. Contact points are on residues; Tyrosine [281st] and Lysine [279th] with two contacts.

Compound Pterostilbene-dehydrodimer is a resveratrol derivative with one acceptor and six donors. Two acceptors and a donor is build in the best binding mode with 279th and 281st residues on protein. Molecule has a lipophilic structure with highly positive logarithmic dispersion parameter (highest logD of 6.66 among six lead compounds) and Pho/phi_ASA ratio of 6.47. High lipophilicity may cause non-specific binding and this may not a good drug property.

Comparison of six lead structures and binding modes on Fgfr2 is shown in Figure 5.17. Binding mode orientations of same scoring function has similar alignments.
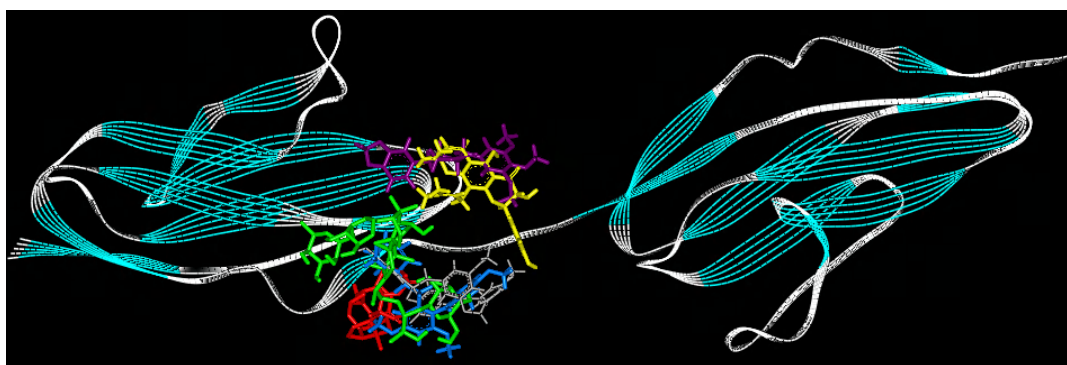
Figure 5.17 Binding modes of six compounds are aligned. Compound m149 is shown in red, TojanosideC is shown in green, compound m100 is shown in yellow, compound m123 is shown in purple, compound res32 is shown in blue and pterostilbene-dehydrodimer shown in gray.

Six compounds are compared in Table 5.4 and aligned in Figure 5.17. As shown in Table 5.4 molecular structures are docked in positions of binding modes according to specified scoring functions. Number of hydrogen bonds built with protein and lipophilic properties shown as dispersion parameter and hydrophobic over hydrophilic accessible surface area ratio are compared in the table. Figure 5.17 shows that each scoring function suggest different orientation for a good binding to Fgfr2. Compound m123 and m100 are the hits of Gold Score and aligned at the same location. But hits of Chem Score: compound res32 and pterostilbene-dehydrodimer align at another location. As in Chapter 4 two hits compounds of AutoDock: compound m149 and TrojanosideC does not align.
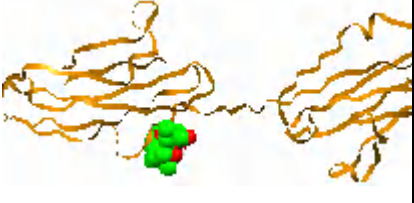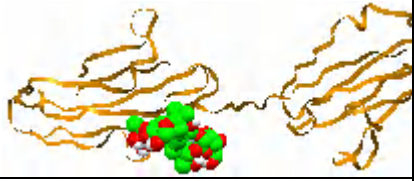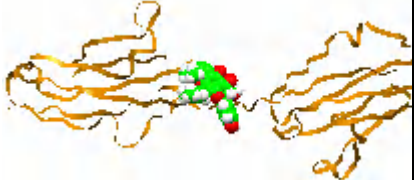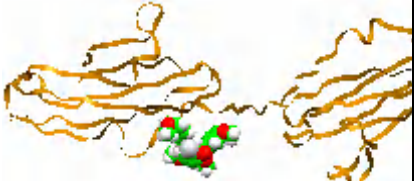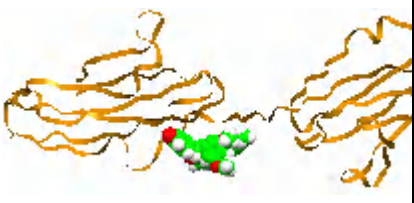
Molecule Scoring	# of H-bonds	# Donor / Acceptor	LogD	Pho/ Phi ratio	Structures
m149 (Au)	3	2 / 5	3.44	6.33	
Trojanoside-C (Au)	1	12 / 18	-1.73	2.81	
m100 (Gold)	2	3 / 6	4.49	3.53	
m123 (Gold)	3	0 / 6	3.12	4.18	
res32	2	0 / 6	3.92	8.22	
Pterostilbene-Dehydrodimer (Chem)	2	1 / 6	6.66	6.47	

Table 5.4 Compounds are compared with number of hydrogen bonds, logarithmic dispersion parameter and hydrophobic / hydrophilic solvent accessible surface area ratios in the table.

AutoDock \ NF-κB	AutoDock \ Fgfr2	Gold Score\ NF-κB	Gold Score \ Fgfr2	Chem Score\ NF-κB	Chem Score\ Fgfr2
TrojanosideJ (cycloartane)	*Elongatoside (cycloartane)*	*m144 (stilbenoid)*	*vaticanolC (resveratrol der.)*	m21 *(flavonoid)*	Pterostilbene-dehydrodimer *(resveratrol der.)*
AskendosideD *(cycloartane)*	TrojanosideC *(cycloartane)*	m14 *(falavonoid)*	hopeaphenol *(harmol alkaloid)*	m54 *(benzenoid)*	res32 *(resveratrol der.)*
m163 (withanolide)	*TrojanosideJ (cycloartane triterpenoid)*	*Resrevatrol-transdehydrodime (stilbeneoid)*	*m144 (stilbenoid)*	m26 *(flavonoid)*	AcetylastragalosideI *(cycloartane triterpenoid)*
CyclocephalosideI (cycloartaned)	m14 *(flavonoid)*	*vaticanolC (stilbenoid)*	m123 *(lignan)*	m02 *(alkaloid)*	res27 *(resveratrol der.)*
Elongatoside (cycloartane triterpenoid)	Macrophyllo-saponinB *(cycloartane)*	m24 *(flavonoid)*	*Resrevatrol-trans-dehydrodime*	m165 *(withanolide)*	m23 *(flavonoid)*
Astrasieversianin nXV *(cycloartane triterpenoid)*	*CyclocephalosideI (cycloartane triterpenoid)*	res16 *(stilbenoid)*	m76 *(flavonoid)*	CycloasgeninC *(cycloartane)*	res20 *(resveratrol der.)*
m29 *(norwithanolide)*	*m163 (withanolide)*	res25 *(stilbenoid)*	m100 *(flavonoid)*	MMD	Resrevatrol-trans-dehydrodimer
TrojanosideB (cycloartane triterpenoid)	m149 *(withanolide)*	Hopea-phenol *(stiilbenoid)*	m67 *(benzofuran)*	m48*(withanolide)*	res16 *(resveratrol der.)*
TorjanosideH *(cycloartane)*	*TrojanosideB (cycloartane)*	res24 *(stilbenoid)*	m125 *(lignan)*	m124 *(lignan)*	res31 *(resveratrol der.)*
m160 (withanolide)	m41 *(stilbenoid)*	m41 (stilbenoid)	m80 *(flavonoid)*	m151 *(alkaloid)*	res25 *(resveratrol der.)*

Table 5.5 Top ten results of Nf-κβ and Fgfr2 are given in descending order. Compound names indicated as italic, in gray and excluded fron lead search.

Results of two proteins are merged in Table 5.5 with all three scoring function hit listing. The ligand molecules suggested by a scoring function for both two proteins may be a structural selectivity trend of the function itself. Even assuming that there is no scoring function selectivity, in biological point of view same affinity to many proteins indicates that the compound is not specific. A drug with non-specific bindings, may bind to other important molecules in the cell and be toxic. After determining the ones that are suspected as non-specific in Table 5.5, are excluded in lead search.

According to results in Table 5.3 and 5.5 a possible lead compound is selected and is analyzed with alignment. Compound to be selected has good results in each scoring function. Compound is a resveratrol derivative coded as res32 and is observed structurally in Figure 5.18.
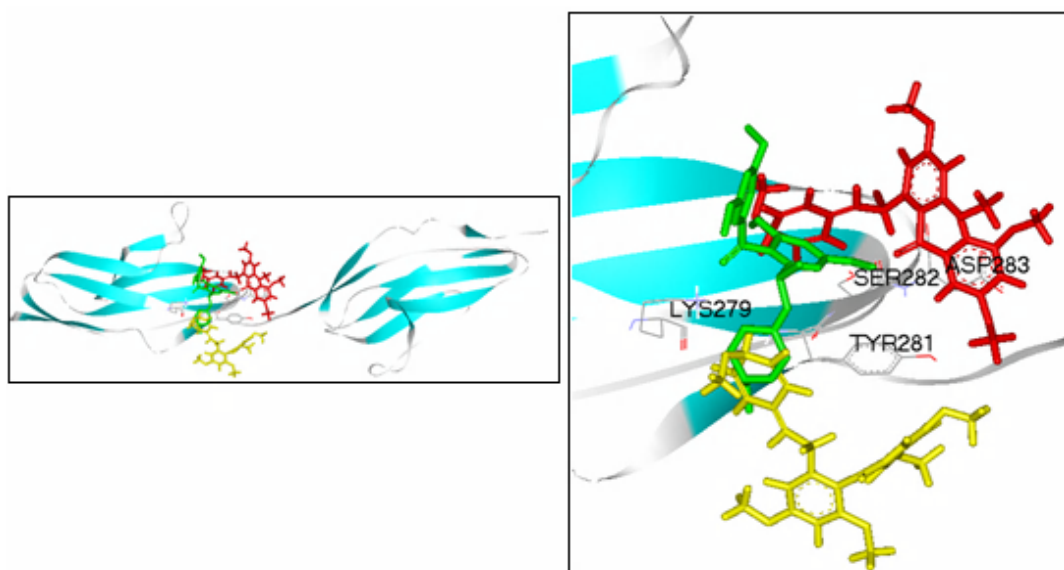
Figure 5.18 On the left compound res32 is docked with Fgfr2 by three different scoring functions are shown. Best binding mode with Chem Score is demonstrated in yellow, Gold Score in red and AutoDock in green. On the right with zoom residues on protein is also shown. Yellow binding mode has interaction with 279th and 281st residues, red binding mode has interactions with 282nd and 283rd residues and green structure has two hydrogen bonds with 279th ad 282nd residues.

Compound res32 best binding modes for each method is shown in different colors in Figure 5.18. Although lead compound for NF-κB shows good alignment for all scoring functions, lead compound res32 for Fgfr2 has different orientations on the binding pocket for all three scoring functions used. But the interaction with residues on protein is in the same range (281st, 282nd, 283rd residues) for all three scoring functions. Docking of compound res 32 with Fgfr2 has binding free energy of -26.47 kcal/mol by Chem Score, 35.71 of fitness by Gold Score and -6.9 kcal/mol by AutoDock scoring functions. Ranking is different but interactions of the best binding modes are similar for three functions.

Chem Score scoring function selects resveratrol derivatives (basically stilbenoids) as lead compounds. 80% of the top ten ranked molecules are stilbenoid structures. While the stilbenoid ratio for the total data set is 19%. The top ten molecules have high S(lipo) (protein-ligand lipophilic contribution) and Chem Score fitness. Average S(lipo) is 131. Values over average are accepted as high valeus. 65% of 46 the stilbenoids have high S(lipo), 78% is with high Chem Score fitness and 80% have binding free energy change (DG) lower than -17 kcal/mol (average DG value). And most of the stilbenoids are with low S(hbond) and H(rot) values. S(lipo) component of Chem Score is the most effective component for selecting best molecules among stilbenoids. The major component S(lipo) is based on protein-ligand pair. Hence, we can conclude that selectivity of Chem Score is a protein-ligand pair based selectivity.

On the other hand AutoDock favors Cycloartane triterpenoid molecules for Fgfr2 as in NF-κB docking results. 60% of the top ten ranked molecules are cycloartane triterpenoid structures. While the cycloartane triterpenoid ratio for the total data set is 15%. Cycloartane triterpenoids are ranked as the best group of molecules with low unbound energy, high final internal total free energy. Both of the energy terms mentioned are dependent on ligand.

As far as final internal free energy of ligands have important role over all evaluation, we can conclude that AutoDock has tendency to select a specific group of ligand structure according to the ligand property itself.

Chapter 6

CONCLUSIONS

In this study each scoring function (AutoDock, Gold Score, Chem Score) shows different rankings patterns. None of the correlation curves obtained using two different scoring functions have a significant linear regression coefficient. The best ranked NF-κB docking results with Gold Score scoring function are stilbenoids in majority, while the best ranked results of AutoDock with the same protein are cycloartane tritepenoid molecules. It has been shown that the best ranked Fgfr2 docking results with Chem Score scoring function are stilbenoids in majority, while Gold Score has no selectivity for Fgfr2 whereas AutoDock is selective against the same structural group (cycloartane triterpenoids) with Fgfr2 docking as in NF-κB docking. Scoring functions choose hits among the major structural groups as shown in Table 6.1.

Scoring Function	Target	Major group	Top ten ratio
AutoDock	NF-κB	**C. Triterpenoid	7/10
	Fgfr2	**C. Triterpenoid	6/10
Gold Score	NF-κB	Stilbenoid	8/10
	Fgfr2	-	-
Chem Score	NF-κB	-	-
	Fgfr2	Stilbenoid	8/10

Table 6.1 Selective results of each scoring function

Since the selectivity of each scoring function is based on the main equations and on the structural group properties of molecules, detailed results explain the tendency of each function.

Firstly, the main equations are analyzed in detail. Gold Score results of stilbenoids have high S(int) and and high S(vdw-ext). Gold Score main equation is dominated by both protein-ligand van der Waals score and internal ligand van der Waals energy score; selectivity based on system. Chem Score has also selectivity based on all system docked with dominant score of protein-ligand lipophilic interactions. But according to the similar results in both proteins AutoDock's selectivity is dominated by unbound and final internal free energy of the ligand.

Secondly, structural group properties are analyzed in detail. The mostly selected two major groups for specific scoring functions, which mentioned above, are studied for general structural properties more in detail. Cycloartane triterpenoids, mostly selected as AutoDock hit molecules in our data set, having much more hydrogen bonding donor and acceptor sites, having lower logD values meaning implying that more hydrophilic structure and higher total Accessible Surface Area indicating more complex structure than average in data set. While Stilbenoids, the mostly selected Gold Score hit molecules for NF-κB in our data set, have higher logD values and higher Pho/phi accessible surface area ratios meaning more lipophilic structure and higher total accessible surface area meaning more complex structures than average values in data set. Moreover, resveratrol derivatives, a special group from stilbenoids are mostly selected as Chem Score hit molecules for Fgfr2, with same properties as all stilbenoid but with higher total accessible surface area values and higher Pho/phi_ASA ratios than other stilbenoids.

After analyzing the main equation of the scoring function and structural properties we can conclude the reason for selecting some specific structural groups for each scoring function. Autodock has the lipophilic interactions inside the main equation but the reason for Cycloartane triterpenoids to be selected is that they have low unbound energy term and low final internal energy terms in the equation. Chem Score, has a main equation with lipophilic interactions are included and Stilbenoids have high values of S(lipo) in the equation. On the other hand, Gold Score underestimates hydrophobic interactions, searches for hydrogen bonding motifs but Stilbenoids with highly

lipophilic properties are selected by as hits. This result may be a result of underestimation of hydrophobic interactions causing solvent-exposed hyrdophobic groups and strong hydrogen bonds with protein [40]. These types of wrong solutions are reported in G.Jones et al., 97 [40]. Also hit compounds of Gold score are shown in detail in Table 6.2.

Scoring function used	NF-kB hits Compound Code	# of H-bonds	(LogD)	Total_ASA(Å²)
Au	cyclocephalosideI	6	—1.12	738
Au	m163	1	2.46	536
Gold	m144	4	6.5	701
Gold	m14	5	2.81	657
Chem	m21	7	6.65	757
Chem	m26	5	3.55	539

Scoring function used	Fgfr2 hits Compound Code	# of H-bonds	(LogD)	Total_ASA(Å²)
Au	m149	3	3.44	513.5
Au	Trojanoside-C	1	-1.73	895.1
Gold	m100	2	4.49	619.2
Gold	m123	3	3.12	666.7
Chem	res32	2	3.92	802.4
Chem	Pterostilbene-dehydrodimer	2	6.66	912.1

Table 6.1 Red circle shows the high hyrdophobicity values of Gold Score and Chem Score hit molecule propertied. Yellow circles shown the hydrogen bonds build for Gold Score and Chem Score hit molecules docked with target proteins. First table shows docking resultant hits with NF-κB and second table shows docking resultant hits for Fgfr2.

Data set diversity may be one of the reasons of the tendency to specific structural groups. In the literature it is shown that docking with good affinity results are observed with diverse data sets, but with a specified group of structures success rate is low in predicting affinities of protein-ligand pairs. Our data set is small and not diverse enough. Analysis over larger data sets may give a better understanding.

On the other hand, as reported in recent studies, the ranking of a docked protein-ligand pair is different for each scoring function, while the orientation and the interaction results of docking of different scoring functions are similar [5].

In this study, the same ligand and protein pairs docked with three scoring functions and the resultant three binding modes are analyzed. The results in Chapter 4 are consistent with three aligned binding modes of compound m14 with NF-κB. The results in Chapter 5 are consistent with three binding modes of compound res32 with same interaction location on Fgfr2. But the three different binding modes of res32 with Fgfr2 do not align.

We have observed that same molecules are ranked similarly for different targets in Table 5.5 . Possible reasons are; non-specific binding of the compound, target may have similar binding pockets of the target or the scoring function choosing the major group. For better understanding, the possiblities needed to be investigated as mentioned above.

After evaluating the scoring functions, assessment of results for selecting good leads for our target proteins is needed. Evaluation of scoring functions shows us which molecules to be eliminated as candidates. Non specific molecules are selected via comparison of each target ranks. Leads found for Nfκβ are CyclocpalosideI, m163, m144i m14, m21 and m26. Leads found for Fgfr2 are m149, TrojanosideC, m100, m123, res32 and Pterostilbene-dehydrodimer.

As a future work in drug discovery, lead optimizazion; modification of the found leads for better binding and drug properties is needed. LD $_{50}$ (lethal death) studies on cancer cell lines *in vitro* needed in order to understand the drug effectivity in living cell. Analysis over larger and more diverse data set may give a better understanding for evaluating different scoring functions.

Appendix 1

SUPPLEMENTARY MATERIAL for CHAPTER 5

Table A.1 Results of the 237 compound data set Docked with Nf-κβ

Molecule Name	Chem score (kcal/mol)	Avearge binding free energy Chem score	Gold score	Average fitness score Gold score	Autodock (kcal/mol)	Avearge binding free energy Autodock
gAcetylastragalosideI	-23.79	-18.40	22.67	16.02	-4.74	-0.15
gampelopsinD	-22.04	-19.42	41.93	40.21	-5.42	-4.77
gAskendosideD	-15.61	-7.87	-18.14	-30.24	-9.27	-5.26
gAstragalosideI	-14.91	-9.64	22.82	4.89	-7.04	-3.51
gAstragalosideIV	-18.72	-14.22	31.77	20.87	-7.38	-5.63
gAstragalosideVI	-10.77	-7.25	12.62	-4.65	-7.75	-2.42
gAstrasieversianinII	-12.36	-7.51	18.42	7.29	-6.69	-3.58
gAstrasieversianinX	-13.64	-9.80	11.93	5.33	-5.89	-3.94
gAstrasieversianinXV	-10.77	-10.19	9.49	1.92	-8	-5.00
gATA	-20.1	-18.61	40.2	37.53	-4.35	-2.59
gBaibutoside	-12.07	-9.00	3.47	-8.74	-6.17	-1.97
gCycloasgeninC	-25.49	-19.90	18.15	13.63	-5.47	-3.71
gCyclocanthogenin	-22.54	-21.19	9.68	5.65	-5.66	-3.99
gCyclocanthosideG	-8.84	-4.59	20	3.71	-5.61	3.73
gCyclocephalogenin	-21.48	-20.74	14.47	10.23	-5.23	-4.55
gCyclocephalosideI	-21.85	-13.34	8.39	0.66	-8.87	-5.03
gCyclosieversigenin	-24.03	-20.71	23.61	17.80	-6.62	-5.57
gDHMC5	-14.83	-11.09	44.97	39.47	-4.9	-2.92
gdiptoindonesinA	-17.07	-15.96	46.07	44.08	-6.94	-5.28
gElongatoside	-23.26	-15.10	19.67	8.77	-8.68	-6.44
gGA	-19.73	-17.61	30.86	30.01	-3.67	-3.20
gharmin	-18.06	-16.71	34.99	34.51	-5.1	-4.94
gharmol	-19.7	-18.92	34.76	32.36	-5.15	-5.14
ghopeaphenol	-16.98	-11.87	53.44	46.19	-6.61	-4.41
gIC-1	-17.9	-13.71	16.9	2.09	-7.13	-4.62
gIC-2	-16.99	-9.00	38.07	24.52	-6.59	-0.82
gIC-3	-14.95	-12.46	17.84	9.19	-7.37	-4.97
gIC-4	-10.25	-8.97	13.94	-2.97	-5.51	-0.04

gIsopaucifloralF	-23.89	-20.18	42.77	41.22	-5.33	-4.75
gm01	-24.01	-20.82	29.22	28.95	-6.44	-5.58
gm02	-26.74	-18.97	12.05	8.02	-5.85	-4.62
gm03	-21.56	-19.58	44.17	39.32	-6.12	-4.91
gm04	-19.94	-15.54	43.41	36.19	-5.4	-4.03
gm05	-21.94	-20.92	48.02	44.63	-6.21	-5.54
gm06	-20.87	-19.53	42.1	40.45	-6.18	-5.27
gm07	-19.97	-18.40	34.36	30.07	-6.58	-5.61
gm08	-18.04	-16.64	17.68	14.46	-5.67	-4.47
gm10	-18.06	-16.69	44.31	38.71	0.64	-6.29
gm100	-18.86	-19.05	44.32	40.68	-6.6	-5.50
gm102	-19.45	-19.52	41.46	38.28	-6.2	-5.50
gm103	-19.03	-17.23	36.01	33.74	-5.57	-4.90
gm105	-20.63	-18.48	40.85	38.71	-5.84	-5.30
gm106	-20.06	-18.38	30.92	27.98	-6	-5.36
gm107	-21.51	-18.83	38.26	33.03	-5.49	-4.55
gm108	-20.55	-18.86	40.77	38.31	-6.23	-5.29
gm109	-18.93	-18.18	45.08	43.39	-6.8	-6.31
gm111	-20.77	-19.05	38.66	36.60	-5.72	-5.23
gm112	-19.35	-17.96	49.37	40.80	-7.1	-5.82
gm113	-20.98	-19.96	45.45	41.17	-6.11	-4.86
gm114	-22.02	-18.32	37.82	36.76	-5.9	-5.03
gm115	-18.33	-16.50	48.35	45.60	-6.1	-5.27
gm116	-18.61	-17.02	41.52	37.90	-5.6	-5.20
gm12	-24.99	-20.35	42.3	40.71	-6.54	-5.99
gm120	-24.06	-20.70	42.49	37.03	-4.1	-2.46
gm121	-18.86	-18.37	33.38	32.26	-4.47	-4.13
gm122	-21.31	-20.19	44.03	36.95	-5.4	-4.93
gm123	-18.4	-17.58	44.87	40.94	-6.1	-5.03
gm124	-25.15	-18.62	45.33	41.77	-6.3	-5.59
gm125	-19.03	-18.58	43.55	39.18	-5.7	-4.60
gm126	-20.53	-18.95	43.82	37.24	-4.4	-3.75
gm127	-17.22	-17.34	29.37	28.44	-4.52	-4.50
gm128	-17.3	-17.03	33.58	32.44	-5.62	-5.31
gm129	-22.53	-20.27	41.38	41.07	-5.97	-5.67
gm12viniferin	-19.92	-18.49	25.82	19.02	-6.25	-4.40
gm13	-20.63	-19.39	43.25	42.41	-4.12	-3.51
gm130	-21.67	-21.32	41.18	39.39	-5.77	-5.59
gm131	-16.62	-16.29	19.37	18.91	-4.79	-4.66
gm132	-20.55	-19.96	26.09	24.78	-5.11	-5.06
gm133	-17.77	-16.90	34.22	31.33	-5.79	-5.06
gm134	-14.78	-14.68	35.14	34.89	-5.85	-5.31
gm135	-18.02	-15.63	34.65	31.49	-5.96	-5.02
gm137	-19.2	-18.89	25.71	21.89	-7.06	-6.24
gm138	-21.64	-15.77	31.25	28.60	-6.99	-5.70

gm139	-15.4	-10.88	25.09	20.29	-7.05	-4.56
gm14	-20.85	-15.02	56.85	49.75	-7.65	-5.72
gm140	-21.94	-21.06	36.81	34.01	-5.42	-4.54
gm141	-23.21	-21.83	37.65	35.44	-5.52	-4.67
gm142	-22.55	-16.73	34.17	33.03	-5.67	-4.37
gm144	-20.69	-18.96	64.05	47.63	-6.98	-6.37
gm145	-21.61	-21.13	40.9	38.21	-5.33	-4.69
gm147	-21.12	-17.28	-12.08	-14.95	-5.61	-5.12
gm148	-22.29	-22.15	36.49	33.35	-7.04	-5.61
gm149	-23.38	-19.29	33.59	30.59	-7.7	-7.19
gm15	-23.37	-20.29	44.61	42.74	-6.86	-5.68
gm150	-19.55	-19.39	37.13	36.64	-5.6	-5.44
gm151	-25.13	-21.67	46.97	43.52	-6.67	-5.97
gm152	-22.71	-20.16	41.8	35.80	-6.04	-5.06
gm153	-16.78	-16.21	40.13	38.40	-6.47	-5.54
gm155	-18.11	-16.98	49.76	47.64	-7.17	-5.61
gm156	-18.42	-17.74	44.23	37.88	-5.89	-5.30
gm157	-19.5	-17.13	44.16	42.12	-6.03	-5.73
gm16	-17.81	-15.51	42.2	39.36	-5.41	-4.69
gm160	-17.29	-16.00	44.84	40.52	-7.78	-6.57
gm161	-18.87	-18.83	39.88	38.02	-6.87	-6.12
gm162	-23.15	-18.24	19.91	15.55	-6.81	-5.94
gm163	-20.27	-20.37	42.79	38.19	-8.94	-8.33
gm164	-18.57	-17.69	39.15	29.46	-6.01	-4.97
gm165	-25.63	-20.26	21.57	19.02	-6.64	-5.94
gm166	-20.52	-19.64	19.62	16.54	-6.08	-5.50
gm17	-19.67	-19.06	45.01	40.39	-6.33	-4.80
gm18	-20.68	-18.50	44.71	39.32	-6.51	-5.31
gm19	-19.93	-19.03	39.73	37.32	-6.09	-5.21
gm21prodrg	-29.84	-23.01	42.63	35.24	-5.92	-4.34
gm22	-19.96	-17.71	42.83	39.58	-5.84	-5.52
gm23	-19.81	-18.48	34.36	30.19	-6.43	-4.88
gm24	-19.25	-18.29	55	44.17	-6.61	-5.33
gm25	-17.23	-15.11	39.3	36.74	-6.46	-5.88
gm26	-27.18	-25.71	37.4	33.58	-6.5	-6.04
gm27	-18.46	-17.19	43.42	40.16	-6.83	-6.23
gm28	-17.03	-16.46	45.41	44.39	-6.1	-6.03
gm29	-20.02	-17.87	32.71	32.39	-7.99	-7.92
gm30	-24.96	-23.48	36.9	34.82	-7.36	-7.17
gm31G	-23.61	-20.45	26.41	21.76	-7.3	-7.01
gm32	-18.9	-17.16	34.09	31.26	-7.67	-5.87
gm33	-17.79	-16.36	39.89	37.92	-5.98	-5.44
gm34	-18.28	-16.88	44.75	39.40	-7.13	-5.58
gm35	-20.86	-20.71	37.53	36.52	-7.03	-6.71
gm36	-18.51	-15.84	39.12	35.98	-6.37	-5.75

gm37	-20.2	-20.20	30.35	27.96	-5.85	-4.64
gm38	-20.04	-18.54	41.53	37.50	-6.17	-5.05
gm39	-23.54	-22.45	43.72	39.79	-6.09	-5.01
gm40	-23.11	-20.82	39.96	37.53	-6.24	-5.26
gm41	-21.01	-13.90	52.73	47.44	-7.34	-5.42
gm42	-21.65	-18.57	47.22	42.58	-7.33	-5.49
gm43	-19.27	-17.77	40.04	30.17	-5.37	-4.07
gm44	-23.36	-19.12	20.12	16.50	-5.23	-4.18
gm45	-19.15	-17.93	25.69	20.33	-6.89	-6.53
gm46	-23.67	-20.38	37.88	33.70	-6.41	-6.32
gm47	-21.89	-18.69	24.19	22.20	-7.24	-6.94
gm48	-25.37	-18.63	28.48	22.04	-7.5	-5.26
gm49	-24.41	-21.52	32.19	28.79	-6.41	-5.83
gm50	-19.61	-18.46	39.59	39.32	-5.23	-5.09
gm51	-23.11	-21.47	38.53	38.28	-6.75	-5.77
gm52	-22.82	-21.41	42.3	38.27	-5.64	-5.27
gm53	-20.72	-20.33	39.8	38.05	-5.21	-4.67
gm54	-27.47	-20.82	47.54	42.67	-6.46	-5.76
gm55	-22.35	-20.54	43.14	42.00	-5.61	-4.93
gm56	-20.52	-18.69	39.38	36.49	-5.42	-4.44
gm57	-17.98	-17.75	30.84	29.31	-4.13	-3.77
gm58	-17.88	-17.87	34.91	32.02	-4.82	-4.40
gm59	-22.71	-17.43	44.78	41.27	-5.78	-4.35
gm60	-20.78	-19.80	33.86	33.14	-3.77	-3.33
gm61	-17.02	-16.90	29.68	29.35	-3.71	-3.32
gm62	-18.89	-18.72	31.04	29.04	-4.28	-3.94
gm63	-19.66	-19.61	33.51	31.76	-4.78	-3.84
gm64	-21.54	-20.26	44.55	41.08	-6.03	-5.17
gm65	-21.77	-19.09	37.04	34.21	-6.12	-5.42
gm66	-20.41	-20.30	39.45	39.07	-5.68	-5.38
gm67	-22.11	-21.23	50.16	47.45	-5.88	-5.15
gm68	-17.19	-16.08	29.59	24.91	-7.09	-5.83
gm70	-17.62	-16.91	33.12	32.37	-4.83	-4.80
gm71	-16.43	-14.69	51.72	46.20	-6.69	-5.13
gm72	-24.77	-22.49	49.05	44.65	-5.71	-4.94
gm73	-18.77	-18.72	25.82	23.80	-6.35	-5.77
gm74	-18.77	-17.88	23.33	21.69	-5.59	-5.03
gm75	-23	-18.60	36.18	31.42	-3.86	-2.44
gm76	-22.63	-21.07	50.25	47.37	-6.35	-5.39
gm77	-22.05	-20.09	40.34	38.97	-5.63	-4.70
gm78	-20.84	-18.66	45.54	41.61	-6.11	-5.22
gm79	-22.49	-18.71	49.44	40.86	-6.94	-5.55
gm80	-20.84	-19.35	48.74	44.15	-6.66	-5.76
gm81	-19.36	-18.66	42.71	40.40	-7.36	-7.29
gm82	-17.99	-17.68	43.86	39.33	-5.68	-4.78

gm83	-20.97	-19.77	40.26	39.18	-5.9	-4.79
gm84	-23.81	-20.94	38.15	36.39	-6.23	-5.67
gm85	-20.79	-19.98	40.96	38.00	-5.5	-5.00
gm86	-20.5	-18.79	40.98	39.39	-6.33	-5.58
gm88	-23.56	-21.71	41.4	39.37	-6.09	-5.33
gm89	-18.01	-17.31	34.8	32.76	-6.09	-4.87
gm90	-19.79	-18.12	48.69	45.72	-6.5	-5.86
gm92	-20.02	-20.48	37.95	36.56	-5.98	-5.57
gm93	-20.73	-18.37	39.39	36.89	-5.9	-5.11
gm94	-20.03	-18.57	41.76	41.45	-5.61	-5.48
gm95	-17.77	-16.84	40.82	39.48	-5.48	-5.00
gm96	-23.35	-19.83	39.88	36.48	-5.88	-4.84
gm98	-19.18	-18.51	40.32	37.65	-6.03	-5.27
gm99	-16.29	-15.38	49.25	41.78	-5.82	-4.65
gMacrophyllogenin	-23.18	-19.41	12.72	-0.41	-6.35	-4.92
gMacrophyllosaponinA	-12.87	-10.65	23.63	9.85	-7.55	-3.07
gMacrophyllosaponinB	-15.52	-11.78	13.47	-0.37	-5.96	-2.97
gMacrophyllosaponinC	-16.6	-13.05	3.02	-7.56	-6.36	-3.09
gMacrophyllosaponinD	-12.97	-7.79	4.5	-3.03	-5.65	0.82
gMacrophyllosaponinE	-16.48	-10.33	18.64	0.69	-4.9	-2.00
gMMD	-25.4	-19.86	13.35	8.21	-5.1	-4.82
gpaucifloralF	-22.39	-20.35	51.91	48.11	-6.3	-5.93
gPMC	-21.31	-17.75	46.08	41.41	-5.46	-3.71
gPMC-A	-20.33	-18.09	48.4	44.13	-6.25	-4.03
gPMC-B	-19.81	-16.56	47.39	43.79	-4.6	-3.05
gPMC-C	-23.53	-19.98	45.46	41.55	-4.57	-3.46
gPMC-D	-18.7	-18.02	45.64	41.05	-4.3	-3.41
gPMC-E	-21.96	-19.56	47.58	43.49	-4.88	-3.32
gPMC-F	-19.71	-17.59	45.14	40.48	-4.15	-3.10
gPMC-G	-23.05	-19.14	45.27	44.43	-4.63	-3.45
gPMC-H	-19.72	-18.36	43.01	40.82	-4.56	-3.30
gPMC-I	-24.41	-21.18	42.04	39.30	-4.4	-3.44
gPMC-J	-16.86	-16.69	29.85	28.75	-3.58	-3.56
gPterostilbene	-21.42	-20.46	41.69	38.89	-4.66	-4.19
gPterostilbene-dehydrodimer	-24.62	-22.17	45.71	41.49	-5.58	-4.82
gquadrangularinA	-21.42	-19.86	47.29	43.46	-6.31	-5.46
gres10	-16.77	-16.70	32	31.72	-3.66	-3.55
gres11	-24.31	-20.17	51.26	40.67	-5.84	-4.66
gres13	-24.56	-22.10	43.9	43.15	-5.79	-5.64
gres14	-20.19	-15.58	42.03	37.57	-5.01	-4.23
gres15	-17.32	-16.58	46.33	45.47	-5.26	-4.58
gres16	-22.35	-20.26	54.33	49.12	-6.86	-5.32
gres17	-17.88	-15.23	43.57	37.31	-6.84	-4.52
gres18	-23.48	-20.73	41.36	38.27	-6.11	-5.32

gres20	-21.26	-19.77	43.19	37.75	-6.07	-4.72
gres23	-19.65	-17.93	42.64	40.98	-6.44	-5.58
gres24	-21.8	-18.99	52.98	44.86	-6.66	-5.52
gres25	-23.55	-19.15	54.11	39.84	-5.78	-4.59
gres27	-23.14	-18.54	42.43	39.46	-5.32	-4.97
gres28	-19.62	-18.10	27.14	22.54	-5.67	-3.94
gres30	-20.31	-16.53	46.36	41.60	-6	-5.45
gres31	-24.32	-20.07	44.54	36.30	-6.55	-5.14
gres32	-20.09	-19.25	52.48	45.33	-7.05	-5.44
gres33	-19.72	-18.09	47.96	46.62	-5.99	-4.88
gres34	-22.93	-20.51	38.97	37.91	-6.51	-5.40
gres36	-20.73	-14.34	34.5	30.91	-5.37	-4.49
gres37	-20.54	-17.79	46.76	34.71	-5.79	-5.17
gres38	-19.16	-19.45	43.91	42.61	-5.48	-4.85
gres9	-20.47	-18.46	41.48	36.77	-5.36	-4.89
gResrevatrol-trans-dehydrodime	-24.59	-5.07	56.25	48.73	-5.83	-5.07
gResveratrol	-20.9	-19.55	39.53	37.23	-4.72	-4.11
gsjh23	-22.55	-21.26	45.97	40.95	-5.25	-4.19
gTorjanosideH	-20.31	-10.60	15.65	5.62	-7.81	-0.95
gTrojanosideA	-21.29	-14.25	18.15	9.95	-7.07	-3.63
gTrojanosideB	-14.73	-7.02	36.69	17.15	-7.84	-3.30
gTrojanosideC	-12.49	-6.26	16.78	3.08	-6.75	-1.28
gTrojanosideD	-11.66	0.64	12.24	-0.63	-6.84	0.11
gTrojanosideE	-3.3	3.43	16.23	-8.26	-3.72	52.63
gTrojanosideF	-3.71	4.30	1.12	-19.26	-3.76	10.07
gTrojanosideI	-15.13	-8.98	21.1	10.28	-6.86	-3.20
gTrojanosideJ	-10.02	-9.24	-37.22	-47.94	-9.29	-4.36
gTrojanosideK	-13.33	-7.01	2.73	-5.94	-5.21	1.12
gvaticanolC	-14.69	-10.86	55.36	48.23	-1.67	-4.73

Appendix 2

SUPPLEMENTARY MATERIAL for CHAPTER 6

All comparison data for best results of docking Against Nf-κβ and Fgfr2 are below table.

Table A.2 All docking results for each protein

Molecule Name	Nf-κβ Chem Score (kcal/mol)	FGFR2 Chem Score	Nf-κβ Gold Score	FGFR2 Gold Score	Nf-κβ Autodock (kcal/mol)	FGFR2 Autodock
gAcetylastragalosideI	-23.79	-24.54	22.67	7.05	-4.74	-5.87
gampelopsinD	-22.04	-12.78	41.93	33.4	-5.42	-5.13
gAskendosideD	-15.61	-5.41	-18.14	-20.45	-9.27	-7.45
gAstragalosideI	-14.91	-26.43	22.82	3.07	-7.04	-5.61
gAstragalosideIV	-18.72	-10.25	31.77	15.59	-7.38	-6.64
gAstragalosideVI	-10.77	-9.6	12.62	-13.92	-7.75	-8.6
gAstrasieversianinII	-12.36	-3.92	18.42	5.25	-6.69	-4.44
gAstrasieversianinX	-13.64	-11.16	11.93	6.73	-5.89	-6.92
gAstrasieversianinXV	-10.77	-7.51	9.49	0.12	-8	-7.71
gATA	-20.1	-11.05	40.2	30.76	-4.35	-3.93
gBaibutoside	-12.07	-9.42	3.47	-18.82	-6.17	-6.08
gCycloasgeninC	-25.49	-1.81	18.15	8.52	-5.47	-4.76
gCyclocanthogenin	-22.54	-22.59	9.68	-0.56	-5.66	-5.69
gCyclocanthosideG	-8.84	0.57	20	-3.88	-5.61	-4.88
gCyclocephalogenin	-21.48	-15.57	14.47	5.44	-5.23	-5.32
gCyclocephalosideI	-21.85	-11.78	8.39	0.73	-8.87	-8.03
gCyclosieversigenin	-24.03	-11.78	23.61	6.3	-6.62	-5.91
gDHMC5	-14.83	-12.09	44.97	43.14	-4.9	-5.01
gdiptoindonesinA	-17.07	-18.52	46.07	37.19	-6.94	-5.59
gElongatoside	-23.26	-12.45	19.67	6.81	-8.68	-9.6

gGA	-19.73	-10.55	30.86	23.91	-3.67	-4.55
gharmin	-18.06	-17.53	34.99	34.34	-5.1	-5.64
gharmol	-19.7	-10.35	34.76	33.38	-5.15	-5.76
ghopeaphenol	-16.98	-11.31	53.44	50.35	-6.61	-4.07
gIC-1	-17.9	-3.19	16.9	1.7	-7.13	-8.01
gIC-2	-16.99	-8.44	38.07	20.85	-6.59	-6.26
gIC-3	-14.95	-2.86	17.84	10.26	-7.37	-6.7
gIC-4	-10.25	-14.99	13.94	10.04	-5.51	-7.25
gIsopaucifloralF	-23.89	-19.07	42.77	34.42	-5.33	-5.29
gm01	-24.01	-23.34	29.22	25.99	-6.44	-6.48
gm02	-26.74	-12.9	12.05	-0.83	-5.85	-5.19
gm03	-21.56	-21.99	44.17	39.64	-6.12	-6.4
gm04	-19.94	-20.43	43.41	38.49	-5.4	-4.71
gm05	-21.94	-19.95	48.02	42.3	-6.21	-7
gm06	-20.87	-18.37	42.1	36.88	-6.18	-6.28
gm07	-19.97	-18.1	34.36	35.55	-6.58	-7.28
gm08	-18.04	-19.46	17.68	20.57	-5.67	-5.97
gm10	-18.06	-18.2	44.31	42.09	0.64	-4.83
gm100	-18.86	-15.17	44.32	44.96	-6.6	-6.45
gm102	-19.45	-20.11	41.46	35.37	-6.2	-5.97
gm103	-19.03	-18.94	36.01	33.38	-5.57	-6.63
gm105	-20.63	-21.57	40.85	38.41	-5.84	-7.2
gm106	-20.06	-18.91	30.92	30.55	-6	-5.29
gm107	-21.51	-17.17	38.26	36.04	-5.49	-6.58
gm108	-20.55	-20.69	40.77	35.03	-6.23	-6.73
gm109	-18.93	-18.02	45.08	34.32	-6.8	-6.38
gm111	-20.77	-19.66	38.66	37.65	-5.72	-6.09
gm112	-19.35	-20.04	49.37	32.48	-7.1	-6.85
gm113	-20.98	-21.27	45.45	40.91	-6.11	-5.49
gm114	-22.02	-19.61	37.82	31.82	-5.9	-5.86
gm115	-18.33	-20.07	48.35	42.67	-6.1	-5.71
gm116	-18.61	-21.82	41.52	32.88	-5.6	-6.75
gm12	-24.99	-21.54	42.3	35.9	-6.54	-4.89

gm120	-24.06	-17.26	42.49	29.22	-4.1	-3.18
gm121	-18.86	-16.91	33.38	29.86	-4.47	-5.05
gm122	-21.31	-21.56	44.03	40.33	-5.4	-5.99
gm123	-18.4	-20.4	44.87	47.46	-6.1	-6.07
gm124	-25.15	-18.89	45.33	42	-6.3	-6.45
gm125	-19.03	-22.74	43.55	44.26	-5.7	-4.58
gm126	-20.53	-20.99	43.82	40.56	-4.4	-5.76
gm127	-17.22	-18.71	29.37	26.89	-4.52	-4.56
gm128	-17.3	-18.5	33.58	32.96	-5.62	-6.16
gm129	-22.53	-17.53	41.38	35.19	-5.97	-5.92
gm12viniferin	-19.92	-18.33	25.82	17.81	-6.25	-6.97
gm13	-20.63	-21	43.25	40.02	-4.12	-3.9
gm130	-21.67	-17.04	41.18	39.21	-5.77	-5.65
gm131	-16.62	-17.2	19.37	16.99	-4.79	-5.1
gm132	-20.55	-16.58	26.09	22.78	-5.11	-5.26
gm133	-17.77	-18.64	34.22	30.62	-5.79	-5.55
gm134	-14.78	-17.55	35.14	28.41	-5.85	-5.85
gm135	-18.02	-15.5	34.65	25.57	-5.96	-5.51
gm137	-19.2	-13.3	25.71	19.5	-7.06	-7.14
gm138	-21.64	-7.83	31.25	33.05	-6.99	-7.12
gm139	-15.4	-19.33	25.09	15.13	-7.05	-6.52
gm14	-20.85	-8.25	56.85	38.25	-7.65	-8.58
gm140	-21.94	-17.04	36.81	31.03	-5.42	-5.63
gm141	-23.21	-14.86	37.65	32.41	-5.52	-5.57
gm142	-22.55	-20.34	34.17	34.4	-5.67	-5.57
gm144	-20.69	-18.42	64.05	49.64	-6.98	-6.69
gm145	-21.61	-19.52	40.9	31.81	-5.33	-5.27
gm147	-21.12	-18.99	-12.08	-26.42	-5.61	-5.95
gm148	-22.29	-22.92	36.49	28.36	-7.04	-6.56
gm149	-23.38	-21.63	33.59	35.71	-7.7	-8.01
gm15	-23.37	-20.37	44.61	37.57	-6.86	-3.82
gm150	-19.55	-21.15	37.13	33.96	-5.6	-5.84
gm151	-25.13	-22.1	46.97	40.5	-6.67	-6.42

gm152	-22.71	-13.78	41.8	41.21	-6.04	-6.49
gm153	-16.78	-17.91	40.13	33.35	-6.47	-6.03
gm155	-18.11	-22.26	49.76	37.09	-7.17	-5.95
gm156	-18.42	-14.65	44.23	32.51	-5.89	-6.69
gm157	-19.5	-14.3	44.16	30.27	-6.03	-5.67
gm16	-17.81	-18.7	42.2	39.31	-5.41	-5.33
gm160	-17.29	-20.47	44.84	36.76	-7.78	-7.51
gm161	-18.87	-27.31	39.88	37.84	-6.87	-7.03
gm162	-23.15	-16.65	19.91	19.05	-6.81	-7.45
gm163	-20.27	-19.1	42.79	30.46	-8.94	-8.02
gm164	-18.57	-20.57	39.15	31.97	-6.01	-5.54
gm165	-25.63	-21.54	21.57	23.82	-6.64	-6.63
gm166	-20.52	-17.46	19.62	18.5	-6.08	-6.27
gm17	-19.67	-19.68	45.01	39.42	-6.33	-5.2
gm18	-20.68	-18.13	44.71	33.08	-6.51	-7.28
gm19	-19.93	-19.55	39.73	32.16	-6.09	-5.85
gm21prodrg	-29.84	-20.15	42.63	37	-5.92	-6.11
gm22	-19.96	-13.03	42.83	31.27	-5.84	-6.19
gm23	-19.81	-25.49	34.36	34.72	-6.43	-6.55
gm24	-19.25	-14.19	55	39.62	-6.61	-6.46
gm25	-17.23	-17.13	39.3	38.21	-6.46	-6.32
gm26	-27.18	-20.43	37.4	38.99	-6.5	-7.13
gm27	-18.46	-19.28	43.42	43.04	-6.83	-6.16
gm28	-17.03	-24.38	45.41	41.16	-6.1	-6.23
gm29	-20.02	-21.83	32.71	27.54	-7.99	-6.92
gm30	-24.96	-23.1	36.9	37.2	-7.36	-7.28
gm31	-23.61	-24.43	26.41	19.77	-7.3	-7.12
gm32	-18.9	-23.1	34.09	34.41	-7.67	-6.16
gm33	-17.79	-24.12	39.89	35.74	-5.98	-7.04
gm34	-18.28	-21.26	44.75	38.96	-7.13	-7.12
gm35	-20.86	-21.04	37.53	39.78	-7.03	-6.82
gm36	-18.51	-17.98	39.12	42.25	-6.37	-6.65

gm37	-20.2	-14.94	30.35	20.62	-5.85	-5.98
gm38	-20.04	-19.77	41.53	35.27	-6.17	-6.51
gm39	-23.54	-19.37	43.72	35.78	-6.09	-6.23
gm40	-23.11	-21.24	39.96	41.48	-6.24	-6.22
gm41	-21.01	-18.09	52.73	39.32	-7.34	-7.84
gm42	-21.65	-21.41	47.22	41.82	-7.33	-5.95
gm43	-19.27	-20.41	40.04	25.89	-5.37	-5.63
gm44	-23.36	-16.23	20.12	12.52	-5.23	-5.69
gm45	-19.15	-18.84	25.69	21.81	-6.89	-7.38
gm46	-23.67	-19.32	37.88	29.61	-6.41	-7.42
gm47	-21.89	-21.32	24.19	22	-7.24	-7.31
gm48	-25.37	-19.38	28.48	19.45	-7.5	-7.6
gm49	-24.41	-19.25	32.19	31.08	-6.41	-7.38
gm50	-19.61	-19.62	39.59	34.39	-5.23	-5.7
gm51	-23.11	-18.62	38.53	32.9	-6.75	-5.96
gm52	-22.82	-16.34	42.3	35.57	-5.64	-5.6
gm53	-20.72	-22.42	39.8	33.91	-5.21	-4.81
gm54	-27.47	-21.9	47.54	40.57	-6.46	-6.43
gm55	-22.35	-20.92	43.14	33.53	-5.61	-5.74
gm56	-20.52	-15.86	39.38	36.26	-5.42	-5.16
gm57	-17.98	-13.8	30.84	32.54	-4.13	-4.62
gm58	-17.88	-15.54	34.91	29.18	-4.82	-4.93
gm59	-22.71	-11.06	44.78	37.25	-5.78	-5.24
gm60	-20.78	-16.13	33.86	29.99	-3.77	-4.71
gm61	-17.02	-12.11	29.68	21.88	-3.71	-4.48
gm62	-18.89	-14.32	31.04	27.81	-4.28	-4.66
gm63	-19.66	-22.68	33.51	32.61	-4.78	-4.8
gm64	-21.54	-21.77	44.55	37.61	-6.03	-6.48
gm65	-21.77	-20.5	37.04	37.06	-6.12	-7.23
gm66	-20.41	-20.74	39.45	38.22	-5.68	-6.26
gm67	-22.11	-24.86	50.16	44.84	-5.88	-6.1
gm68	-17.19	-25.57	29.59	20.2	-7.09	-7.24
gm70	-17.62	-15.67	33.12	30.54	-4.83	-4.91

gm71	-16.43	-25.05	51.72	35.72	-6.69	-4.08
gm72	-24.77	-17.82	49.05	41.27	-5.71	-6.44
gm73	-18.77	-20.79	25.82	20.7	-6.35	-5.75
gm74	-18.77	-20.2	23.33	17.05	-5.59	-5.45
gm75	-23	-21.19	36.18	31.48	-3.86	-3.01
gm76	-22.63	-17.54	50.25	44.98	-6.35	-6.95
gm77	-22.05	-18.06	40.34	33.61	-5.63	-6.02
gm78	-20.84	-22.48	45.54	41.91	-6.11	-6.16
gm79	-22.49	-19.3	49.44	39.57	-6.94	-6.39
gm80	-20.84	-21.51	48.74	44.06	-6.66	-7.47
gm81	-19.36	-19.75	42.71	40.35	-7.36	-6.55
gm82	-17.99	-12.13	43.86	36	-5.68	-5.91
gm83	-20.97	-21.35	40.26	32.9	-5.9	-6.03
gm84	-23.81	-21.91	38.15	30.67	-6.23	-6.01
gm85	-20.79	-20.37	40.96	40.08	-5.5	-6.32
gm86	-20.5	-18.55	40.98	38.7	-6.33	-7.67
gm88	-23.56	-17.07	41.4	34.87	-6.09	-6.58
gm89	-18.01	-18.19	34.8	33.41	-6.09	-6.2
gm90	-19.79	-17.93	48.69	37.29	-6.5	-6.14
gm92	-20.02	-19.79	37.95	30.03	-5.98	-6.24
gm93	-20.73	-20.52	39.39	34.25	-5.9	-5.95
gm94	-20.03	-18.49	41.76	31.68	-5.61	-6.01
gm95	-17.77	-21.16	40.82	38.27	-5.48	-6.37
gm96	-23.35	-21.74	39.88	37.08	-5.88	-6.77
gm98	-19.18	-16.02	40.32	35.97	-6.03	-6.11
gm99	-16.29	-20.92	49.25	37.04	-5.82	-5.31
gMacrophyllogenin	-23.18	-8.37	12.72	6.29	-6.35	-6.54
gMacrophyllosaponinA	-12.87	-11.01	23.63	6.87	-7.55	-6.98
gMacrophyllosaponinB	-15.52	-7.36	13.47	2.49	-5.96	-8.51
gMacrophyllosaponinC	-16.6	-2.94	3.02	-5.01	-6.36	-7.42
gMacrophyllosaponinD	-12.97	-17.91	4.5	4.59	-5.65	-6.97
gMacrophyllosaponinE	-16.48	-5.34	18.64	2.7	-4.9	-7.26
gMMD	-25.4	-17.98	13.35	10.39	-5.1	-5.4

gpaucifloralF	-22.39		51.91	39.02	-6.3	-6.23
gPMC	-21.31	-17.08	46.08	38.06	-5.46	-4.93
gPMC-A	-20.33	-15.27	48.4	37.15	-6.25	-5.59
gPMC-B	-19.81	-19.81	47.39	32.52	-4.6	-4.1
gPMC-C	-23.53	-21.09	45.46	37.16	-4.57	-4.05
gPMC-D	-18.7	-18.52	45.64	38.16	-4.3	-5.15
gPMC-E	-21.96	-19.45	47.58	38.89	-4.88	-4.46
gPMC-F	-19.71	-22.22	45.14	37.15	-4.15	-4.41
gPMC-G	-23.05	-20.55	45.27	41.63	-4.63	-4.45
gPMC-H	-19.72	-17	43.01	39.85	-4.56	-4.57
gPMC-I	-24.41	-22.54	42.04	35.09	-4.4	-4.45
gPMC-J	-16.86	-17.17	29.85	28.49	-3.58	-3.91
gPterostilbene	-21.42	-27.01	41.69	36.36	-4.66	-5.64
gPterostilbene-dehydrodimer	-24.62	-18.22	45.71	38.75	-5.58	-6.43
gquadrangularinA	-21.42	-18.35	47.29	35.41	-6.31	-6.2
gres10	-16.77	-22.93	32	31.11	-3.66	-4.08
gres11	-24.31	-24.06	51.26	42.49	-5.84	-5.87
gres13	-24.56	-18.87	43.9	38.98	-5.79	-6.02
gres14	-20.19	-19.35	42.03	36.92	-5.01	-5
gres15	-17.32	-24.59	46.33	32.57	-5.26	-5.82
gres16	-22.35	-17.69	54.33	43.19	-6.86	-6.56
gres17	-17.88	-22.27	43.57	30.74	-6.84	-4.98
gres18	-23.48	-21.38	41.36	40.67	-6.11	-5.9
gres20	-21.26	-21.38	43.19	42.62	-6.07	-5.81
gres23	-19.65	-23.75	42.64	41.04	-6.44	-6.13
gres24	-21.8	-21.54	52.98	41.06	-6.66	-6.51
gres25	-23.55	-22.5	54.11	37.06	-5.78	-6.38
gres27	-23.14	-25.72	42.43	42.07	-5.32	-5.91
gres28	-19.62	-24.7	27.14	21.71	-5.67	-5.63
gres30	-20.31	-23.41	46.36	36.08	-6	-5.93
gres31	-24.32	-18.77	44.54	35.05	-6.55	-6.27
gres32	-20.09	-20.99	52.48	43.64	-7.05	-6.9

gres33	-19.72	-15.4	47.96	42.8	-5.99	-5.46
gres34	-22.93	-19.56	38.97	27.02	-6.51	-6.19
gres36	-20.73	-15.56	34.5	30.68	-5.37	-5.69
gres37	-20.54	-22.43	46.76	30.55	-5.79	-6.36
gres38	-19.16	-25.01	43.91	39.49	-5.48	-6.18
gres9	-20.47	-1.99	41.48	31.23	-5.36	-6.25
gResrevatrol-trans-dehydrodime	-24.59	23.26	56.25	47.03	-5.83	-6.92
gResveratrol	-20.9	-20.3	39.53	31.23	-4.72	-5.33
gsjh23	-22.55	-21.08	45.97	35.49	-5.25	-5.4
gTorjanosideH	-20.31	-8.21	15.65	21.63	-7.81	-7.36
gTrojanosideA	-21.29	4.85	18.15	-2.15	-7.07	-6.54
gTrojanosideB	-14.73	-3.39	36.69	14.84	-7.84	-7.97
gTrojanosideC	-12.49	-3.64	16.78	10.87	-6.75	-8.99
gTrojanosideD	-11.66	2.42	12.24	0.41	-6.84	-6.83
gTrojanosideE	-3.3	12.48	16.23	-1.99	-3.72	-6.28
gTrojanosideF	-3.71	1.23	1.12	-24.39	-3.76	-6.54
gTrojanosideI	-15.13	-7.63	21.1	11.7	-6.86	-3.23
gTrojanosideJ	-10.02	-12.36	-37.22	-41.31	-9.29	-8.78
gTrojanosideK	-13.33	0.12	2.73	-12.66	-5.21	-6.72
gvaticanolC	-14.69	-14.91	55.36	51.02	-1.67	-6.82

Appendix 3

NF-κB INHIBITORS

Results of known inhibitors against NF-κB (p50) with binding site of DBR(DNA binding region). Three docking methods with best binding mode scores and averages of binding modes of each molecule are shown on the table below.

	Binding site DBR (p50 protein ; centered as 65th residue, 20Ǻ sphere) 1nfk				
	Auto Dock	**Gold**		**Gold\ChemScore**	
	Binding Free energy (kcal\mol)	Gold Score fitting	Gold Score fit. average	Binding Free energy(kcal\mol)	Chem Score Average
ATA	**-6,38**	41,02	35,14	**-22,69**	-17,19
GA	-5,22	28,83	27,05	-18,09	-17,67
DHMC5	-5,47	**53,53**	44	-22,01	-15,83

Table A.3 Dock results of NF-κB (p50) with known inhibitors

Appendix 4

SUPPLEMENTARY MATERIAL FOR CHAPTER 3

PROPERTIES of 237 compound data set

- If polar **BOLD** in ASA-P value, If pozitifly charged **BOLD** in / ASA+

- MarvinSkecth is used in every calculation of structures in .mol2 format.

Table A.4 Properties of molecules in drug data set

Molecule Name	IUPAC Name	@ pH7.4				
		# of H bond Donor	# of H bond Accep.	LoD	Polar Surface Area (ASA-P)	Solvent access. Area (ASA)/ ASA+
AcetylastraAlosideI	(2S,3S,4S,5R)-3-(acetyloxy)-5-hydroxy-2-{[(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-14-hydroxy-15-[(2R,5S)-5-(2-hydroxypropan-2-yl)-2-methyloxolan-2-yl]-7,7,12,16-tetramethyl-9-{[(2S,3R,4R,5R)-3,4,5-trihydroxyoxan-2-yl]oxy}pentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-6-yl]oxy}oxan-4-yl acetate	6	13	0.41	156.7	846.3 / **677.8**
AmpelopsinD	(1E,2S,3S)-3-(3,5-dihydroxyphenyl)-2-(4-hydroxyphenyl)-1-[(4-hydroxyphenyl)methylidene]-2,3-dihydro-1H-indene-4,6-diol	6	6	6.40	222.1	570.5 / **386.3**
AskendosideD	(2S,3S,4S,5S)-2-{[(2S,3R,4S,5R)-4,5-dihydroxy-2-{[(1S,3S,5S,7S,8S,10S,11R,13S,14R,15R)-13-hydroxy-14-[(2R,5R)-5-(2-hydroxypropan-2-yl)-2-methyloxolan-2-yl]-6,6,11,15-tetramethyl-8-{[(2S,3R,4R,5R)-3,4,5-trihydroxyoxan-2-yl]oxy}pentacyclo[8.7.0.0^{1,3}.0^{3,7}.0^{11,15}]heptadecan-5-yl]oxy}oxan-3-yl]oxy}oxane-3,4,5-triol	10	17	-2.29	196.2	734.4 / **580.4**
AstraAlosideI	(2S,3S,4S,5R)-3-(acetyloxy)-5-hydroxy-2-{[(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-14-hydroxy-15-[(2R,5S)-5-(2-hydroxypropan-2-yl)-2-methyloxolan-2-yl]-7,7,12,16-tetramethyl-9-{[(2S,3R,4R,5R)-3,4,5-trihydroxyoxan-2-yl]oxy}pentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-6-yl]oxy}oxan-4-yl acetate	6	13	0.41	156.7	846.3 / **677.8**
AstraAlosideIV	(2R,3R,4R,5S,6R)-2-{[(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-14-hydroxy-15-[(2R,5S)-5-(2-hydroxypropan-2-yl)-2-methyloxolan-2-yl]-7,7,12,16-tetramethyl-6-{[(2S,3S,4S,5R)-3,4,5-trihydroxyoxan-2-yl]oxy}pentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-9-yl]oxy}-6-(hydroxymethyl)oxane-3,4,5-triol	9	14	-1.12	208.1	796.2 / **635.9**

AstraAlosideVI	(2R,3R,4R,5S,6R)-2-{[(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-6-{[(2S,3S,4S,5R)-4,5-dihydroxy-3-{[(2S,3S,4S,5S,6S)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxy}oxan-2-yl]oxy}-14-hydroxy-15-[(2R,5S)-5-(2-hydroxypropan-2-yl)-2-methyloxolan-2-yl]-7,7,12,16-tetramethylpentacyclo [9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-9-yl]oxy}-6-(hydroxymethyl)oxane-3,4,5-triol	12	19	-3.14	237.5	820.8 / **644.4**
Astrasieversianin II	(2S,3S,4S,5R)-3-(acetyloxy)-5-hydroxy-2-{[(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-14-hydroxy-15-[(2R,5S)-5-(2-hydroxypropan-2-yl)-2-methyloxolan-2-yl]-7,7,12,16-tetramethyl-9-{[(2S,3R,4R,5R)-3,4,5-trihydroxyoxan-2-yl]oxy}pentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-6-yl]oxy}oxan-4-yl acetate	6	13	0.41	156.7	846.3 / **677.8**
Astrasieversianin X	(2S,3S,4S,5R)-2-{[(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-14-hydroxy-15-[(2R,5S)-5-(2-hydroxypropan-2-yl)-2-methyloxolan-2-yl]-7,7,12,16-tetramethyl-9-{[(2S,3R,4R,5R)-3,4,5-trihydroxyoxan-2-yl]oxy}pentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-6-yl]oxy}oxane-3,4,5-triol	8	13	-0.45	160.7	748.7 / **600.4**
Astrasieversianin XV	(2S,3S,4S,5R,6R)-2-{[(2S,3S,4S,5R)-4,5-dihydroxy-2-{[(1S,3R,6S,8S,9S,11S,12R,14S,15R,16R)-14-hydroxy-15-[(2R,5S)-5-(2-hydroxypropan-2-yl)-2-methyloxolan-2-yl]-7,7,12,16-tetramethyl-9-{[(2S,3R,4R,5R)-3,4,5-trihydroxyoxan-2-yl]oxy}pentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-6-yl]oxy}oxan-3-yl]oxy}-6-methyloxane-3,4,5-triol	10	17	-1.48	171	854.8 / **692.8**
ATA	5-[(3-carboxy-4-hydroxyphenyl)[(1R)-3-carboxy-4-oxocyclohexa-2,5-dien-1-yl]methyl]-2-hydroxybenzoic acid	2	9	-6.23	**295.2**	522.9 / 192.4
Baibutoside	(2R,3R,4R,5S,6R)-2-{[(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-6-{[(2S,3S,4S,5R)-3-{[(2S,3R,4S)-3,4-dihydroxy-4-(hydroxymethyl)oxolan-2-yl]oxy}-4,5-dihydroxyoxan-2-yl]oxy}-14-hydroxy-15-[(2R,5R)-5-(2-hydroxypropan-2-yl)-2-methyloxolan-2-yl]-7,7,12,16-tetramethylpentacyclo [9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-9-yl]oxy}-6-(hydroxymethyl)oxane-3,4,5-triol	11	18	-2.9	208.6	829.3 / **659.1**
CycloasgeninC	(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-15-[(2R,5S)-5,6-dihydroxy-6-methylheptan-2-yl]-7,7,12,16-tetramethylpentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecane-6,9,14-triol	5	5	2.86	84.2	566.1 / **452.4**
Cyclocanthogenin	(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-15-[(2R,5S)-5,6-dihydroxy-6-methylheptan-2-yl]-7,7,12,16-tetramethylpentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecane-6,9,14-triol	5	5	2.86	77.2	538.2 / **432.5**
CyclocanthosideG	(2R,3R,4R,5S,6R)-2-{[(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-6-{[(2S,3S,4S,5R)-4,5-dihydroxy-3-{[(2S,3S,4S,5S,6S)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxy}oxan-2-yl]oxy}-15-[(2R,5S)-5,6-dihydroxy-6-methylheptan-2-yl]-14-hydroxy-7,7,12,16-tetramethylpentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-9-yl]oxy}-6-(hydroxymethyl)oxane-3,4,5-triol	13	19	-2.81	257.5	895.0 / **704.0**
Cyclocephalogenin	(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-15-[(2R,5S)-5-hydroxy-2,6,6-trimethyloxan-2-yl]-7,7,12,16-tetramethylpentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecane-6,9,14-triol	4	5	2.44	56.5	500.3 / **408.7**

Molecule Name	IUPAC Name	# of H bond Donor	# of H bond Accep.	LoD	Polar Surface Area (ASA-P)	Solvent accesible Area (ASA) / ASA+
Cyclocephaloside I	(2R,3R,4R,5S,6R)-2-{[(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-14-hydroxy-15-[(2R,5S)-5-hydroxy-2,6,6-trimethyloxan-2-yl]-7,7,12,16-tetramethyl-6-{[(2S,3S,4S,5R)-3,4,5-trihydroxyoxan-2-yl]oxy}pentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-9-yl]oxy}-6-(hydroxymethyl)oxane-3,4,5-triol	9	14	-1.12	183.3	738.1 / **591.3**
Cyclosieversigenin	(1S,3R,6S,8R,9S,11S,12S,14S,15R,16R)-6,9-dihydroxy-15-[(2R,5S)-5-(2-hydroxypropan-2-yl)-2-methyloxolan-2-yl]-7,7,12,16-tetramethylpentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-14-yl acetate	3	5	2.87	79.4	629.9 / **505.0**
DHMC5	methyl (2S)-2-({[4-(dimethylamino)-6-({[(2S)-3-(1H-imidazol-4-yl)-1-methoxy-1-oxopropan-2-yl]amino}methyl)-3,5-dimethylpyridin-2-yl]methyl}amino)-3-(1H-imidazol-4-yl)propanoate	4	8	-0.20	126	828.3 / **594.8**
DiptoindonesinA	(10R)-4,6,12,14-tetrahydroxy-10-(4-hydroxyphenyl)tricyclo[9.4.0.0^{3,8}]pentadeca-1(11),3,5,7,12,14-hexaene-2,9-dione	4	7	2.66	216.6	490.0 / **274.2**
ElonAtoside	(2S,3S,4S,5S)-2-{[(2S,3S,4S,5R)-2-{[(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-9,14-dihydroxy-15-[(2R,5R)-5-(2-hydroxypropan-2-yl)-2-methyloxolan-2-yl]-7,7,12,16-tetramethylpentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-6-yl]oxy}-4,5-dihydroxyoxan-3-yl]oxy}oxane-3,4,5-triol	8	13	-0.45	153.1	738.1 / **593.8**
GA	3,4,5-trihydroxybenzoic acid	3	5	-2.56	201.1	308.3 / 132.8
harmin	7-methoxy-1-methyl-9H-pyrido[3,4-b]indole	1	2	1.83	45.1	443.0 / **313.2**
harmol	1-methyl-9H-pyrido[3,4-b]indol-7-ol	2	2	1.79	71.1	380.8 / **244.9**
hopeaphenol	(1R,8R,9S,16R)-8,16-bis(4-hydroxyphenyl)-9-[(1R,8R,9S,16R)-4,6,12-trihydroxy-8,16-bis(4-hydroxyphenyl)-15-oxatetracyclo[8.6.1.0^{2,7}.0^{14,17}]heptadeca-2(7),3,5,10,12,14(17)-hexaen-9-yl]-15-oxatetracyclo[8.6.1.0^{2,7}.0^{14,17}]heptadeca-2(7),3,5,10,12,14(17)-hexaene-4,6,12-triol	10	12	10.92	322.6	889.7 / **605.3**
IC-1	(2S,3R,4S,5S)-2-{[(2S,3R,4S,5R)-2-{[(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-9,14-dihydroxy-15-[(2R,4S,5S)-4-hydroxy-5-(2-hydroxypropan-2-yl)-2-methyloxolan-2-yl]-7,7,12,16-tetramethylpentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-6-yl]oxy}-4,5-dihydroxyoxan-3-yl]oxy}oxane-3,4,5-triol	9	14	-1.57	181	730.1 / **586.9**

Molecule Name	IUPAC Name	# of H bond Donor	# of H bond Accep.	LoD	Polar Surface Area (ASA-P)	Solvent accesible Area (ASA) / ASA+
IC-2	(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-15-[(2R,4S,5S)-4-(acetyloxy)-5-(2-hydroxypropan-2-yl)-2-methyloxolan-2-yl]-6-{[(2S,3R,4S,5R)-4,5-dihydroxy-3-{[(2S,3R,4S,5S)-3,4,5-trihydroxyoxan-2-yl]oxy}oxan-2-yl]oxy}-9-hydroxy-7,7,12,16-tetramethylpentacyclo [9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-14-yl 2-hydroxyacetate	8	15	-1.57	214.8	792.2 / **613.5**
IC-3	(2S,3R,4S,5S)-2-{[(2S,3R,4S,5R)-2-{[(1R,2R,3R,6S,8R,11S,13S,14S,16S,17S,19S,21R,22S)-14,22-dihydroxy-21-(2-hydroxypropan-2-yl)-1,3,12,12,17-pentamethyl-20,24-dioxaheptacyclo [19.2.1.0^{2,19}.0^{3,17}.0^{6,8}.0^{6,16}.0^{8,13}]tetracosan-11-yl]oxy}-4,5-dihydroxyoxan-3-yl]oxy}oxane-3,4,5-triol	8	14	-0.33	165.8	702.5 / **561.5**
IC-4	(2S,3R,4S,5S,6R)-2-({2-[(2S,3S,5R)-5-[(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-6-{[(2S,3R,4S,5R)-4,5-dihydroxy-3-{[(2S,3R,4S,5S)-3,4,5-trihydroxyoxan-2-yl]oxy}oxan-2-yl]oxy}-9,14-dihydroxy-7,7,12,16-tetramethylpentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}] octadecan-15-yl]-3-hydroxy-5-methyloxolan-2-yl]propan-2-yl}oxy)-6-(hydroxymethyl)oxane-3,4,5-triol	12	19	-3.69	241	815.1 / **646.7**
IsopaucifloralF	(2R,3R)-2-(3,5-dihydroxyphenyl)-3-(4-hydroxyphenyl)-2,3-dihydro-1H-indene-4,6-diol	5	5	4.53	197.3	510.5 / **313.4**
M01	N-[(1S,2R,10S,11R,14S,15S)-14-[(1S)-1-(dimethylamino)ethyl]-6-hydroxy-2,15-dimethyltetracyclo[8.7.0.0^{2,7}.0^{11,15}]heptadec-4-en-5-yl]benzamide	3	2	1.37	58.49	584.98 / **437.73**
M02	*[(1S)-1-[(1S,2S,5R,7S,10R,11S,14S,15S,16R)-16-hydroxy-5-{[(1S)-1-hydroxy-3-methylbut-2-en-1-yl](methyl)amino}-2,15-dimethyltetracyclo[8.7.0.0^{2,7}.0^{11,15}]}heptadecan-14-yl]ethyl]dimethylazanium	4	2	1.09	43.5	604.0/ **491.2**
M03	4-[3-(4-hydroxyphenyl)propyl]-2-(3-methylbut-2-en-1-yl)benzene-1,3-diol	4	2	5.6	84.7	505 / **359.8**
M04	4-[(1S,2S)-1-ethoxy-3-hydroxy-1-(4-hydroxy-3-methoxyphenyl)propan-2-yl]-2-methoxyphenol	4	2	2.15	145.1	636.3 / **497.3**
M05	5-[5-hydroxy-7-(3-methylbut-2-en-1-yl)-1-benzofuran-2-yl]benzene-1,3-diol	3	3	4.42	135.5	531.8 / **387.5**
M06	5-(6-hydroxy-1-benzofuran-2-yl)-4-(3-methylbut-2-en-1-yl)benzene-1,3-diol	3	3	4.42	125.7	536.1 / **375.1**

Molecule Name	IUPAC Name	# of H bond Donor	# of H bond Accep.	LoD	Polar Surface Area (ASA-P)	Solvent accesible area (ASA) / ASA+
M07	4-[(1S,2S,4R,5S,6S,9S,10S,13S,15S,18S)-15-{[(2R,3S,4S,5R,6R)-3,5-dihydroxy-4-methoxy-6-methyloxan-2-yl]oxy}-9-hydroxy-5,18-dimethyl-3-oxapentacyclo[8.8.0.0^{2,4}.0^{5,9}.0^{13,18}]octadecan-6-yl]-2,5-dihydrofuran-2-one	3	8	1.62	135.8	620.6 / **470**
M08	4-[(1S,2S,4R,5S,6S,9S,10S,13S,15S,18S)-15-{[(2S,3S,4S,5R,6R)-3,5-dihydroxy-4-methoxy-6-methyloxan-2-yl]oxy}-9-hydroxy-5,18-dimethyl-3-oxapentacyclo[8.8.0.0^{2,4}.0^{5,9}.0^{13,18}]octadecan-6-yl]-2,5-dihydrofuran-2-one	3	8	1.62	119.5	575.4 / **444.9**
M10	(2R,5R,7S)-5-(4-hydroxyphenyl)-13-methoxy-6-oxatricyclo[8.4.0.0^{2,7}]tetradeca-1(14),3,10,12- tetraen-12-ol	2	4	3.71	102.8	527.1 / **381.2**
M100	(2R,3R)-3,5-dihydroxy-2-(4-hydroxyphenyl)-8,8-dimethyl-10-(3-methylbut-2-en-1-yl)-2H,3H,4H,8H-pyrano[3,2-g]chromen-4-one	3	6	4.49	136.7	619.2 / **423.5**
M102	11-methoxy-12-(7-methoxy-2H-1,3-benzodioxol-5-yl)-5,13-dioxatricyclo[7.4.0. 0^{2,6}]tyrideca-1,3,6,8,11-pentaen-10-one	0	6	1.88	132.1	581 / **462.8**
M103	3,5,7-trihydroxy-8-methoxy-2-phenylchromen-4-one	2	6	-0.7	149.8	462.4 / **284.9**
M105	(2Z)-3-(2H-1,3-benzodioxol-5-yl)-1-(5,7-dimethoxy-2,2-dimethylchromen-6-yl)-3-hydroxyprop-2-en-1-one	0	7	1.06	140.8	695.7 / **498.8**
M106	(2S)-2-(2H-1,3-benzodioxol-5-yl)-5,7-dimethoxy-8-(3-methylbut-2-en-1-yl)-3,4-dihydro-2H-1-benzopyran-4-one	0	6	3.57	103.2	544.6 / 508
M107	(2Z)-1-(5,7-dimethoxy-2,2-dimethylchromen-6-yl)-3-hydroxy-3-phenylprop-2-en-1-one	0	5	1.42	65.4	637.2 / **459.1**
M108	3,5,7-trihydroxy-2-[4-hydroxy-3-(3-methylbut-2-en-1-yl)phenyl]chromen-4-one	3	6	0.95	177.6	581.4 / **370.5**
M109	(2S)-5,7-dihydroxy-2-(4-hydroxyphenyl)-8-(3-methylbut-2-en-1-yl)-2,3-dihydro-1-benzopyran-4-one	2	5	3.7	136.9	509.4 / **319.8**
M111	(2Z)-2-[(3,4-dihydroxyphenyl)methylidene]-6-hydroxy-1-benzofuran-3-one	2	5	1.50	164	443.1 / **231.2**
M112	2-(2,4-dihydroxyphenyl)-3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-5,7-dihydroxychromen-4-one	3	6	4.15	161.2	702 / **475.1**
M113	(2E)-1-[2,4-dihydroxy-3-(3-methylbut-2-en-1-yl)phenyl]-3-(2,4-dihydroxyphenyl)prop-2-en-1-one	4	5	4.54	147.4	573.1 / **386.6**
M114	2-(3,4-dimethoxyphenyl)-5-hydroxy-3,7-dimethoxychromen-4-one	1	7	1.77	125.3	650.9 / **525.3**
M115	3-(3,4-dimethoxyphenyl)-7-methoxychromen-4-one	0	5	2.9	86.9	600.4 / **482.6**
M116	5-hydroxy-2-(4-hydroxy-3-methoxyphenyl)-3,7-dimethoxychromen-4-one	2	7	1.73	154.6	606.9 / **471.8**

Molecule Name	IUPAC Name	# of H bond Donor	# of H bond Accep.	LoD	Polar Surface Area (ASA-P)	Solvent accesible area (ASA) / ASA+
M12	(4R,12S)-12-(2,4-dihydroxyphenyl)-4-(2-hydroxypropan-2-yl)-5,13-dioxatricyclo[7.4.0.0^{2,6}]trideca-1,6,8-trien-10-one	3	6	1.79	124.8	521.3 / **372.1**
M120	(2S)-2,3-dihydroxypropyl hexadecanoat	2	3	4.23	100.2	731.3 / **541.1**
M121	(4S)-4-hydroxy-4-[(1E,3R)-3-hydroxybut-1-en-1-yl]-3,5,5-trimethylcyclohex-2-en-1-one	2	3	1.39	70.0	369.8 / **291.4**
M122	(10R,11S,15R)-10-(3,4,5-trimethoxyphenyl)-4,6,13-trioxatetracyclo[7.7.0.0^{3,7}.0^{11,15}]hexadeca-1(9),2,7-trien-12-one	0	6	2.63	137.1	633.9 / **494.4**
M123	(3R,4R)-4-(2H-1,3-benzodioxol-5-ylmethyl)-3-[(3,4,5-trimethoxyphenyl)methyl]oxolan-2-one	0	6	3.12	128.8	666.7 / **519.1**
M124	(3R,3aS,6R,6aR)-3,6-bis(4-hydroxy-3- methoxyphenyl)-tetrahydro-3H-furo[3,4-c]furan-1-one	2	6	2.03	148	588.5 / **434.0**
M125	(3R,4R)-4-[(7-methoxy-2H-1,3-benzodioxol-5-yl)methyl]-3-[(3,4,5-trimethoxyphenyl)methyl] oxolan-2-one	0	7	2.87	131.4	655.0 / **517.9**
M126	(3R,3aS,6S,6aS)-3,6-bis(3,4,5-trimethoxyphenyl)-tetrahydro-3H-furo[3,4-c]furan-1-one	0	8	1.59	122.1	819.9 / **675.4**
M127	(6R,7aS)-6-hydroxy-4,4,7a-trimethyl-6,7-dihydro-5H-1-benzofuran-2-one	1	2	1.29	70.9	301.5 / **204.5**
M128	methyl9-hydroxy-4,4-dimethyl-3-oxatricyclo[8.4.0.0^{2,7}]tetradeca- 1(14),2(7),5,8,10,12-hexaene-8-carboxylate	1	3	4.31	73.6	460.0 / **320.7**
M129	(1S,2R,3S)-4-(4-methoxyphenyl)-2,3-dihydro-1H-phenalene-1,2,3-triol	3	4	2.10	93.4	507.5 / **393.0**
M12viniferin	5-[(2S,3S)-6-hydroxy-2-(4-hydroxyphenyl)-4-[(E)-2-(4-hydroxyphenyl)ethenyl]-2,3-dihydro-1-benzofuran-3-yl]benzene-1,3-diol	5	6	6.5	168.7	581.1 / **383.2**
M13	4-[(2S)-7-hydroxy-8-(3-methylbut-2-en-1-yl)-3,4-dihydro-2H-1-benzopyran-2-yl]benzene-1,3-diol	3	4	4.68	109.7	500.9 / **358.5**
M130	2-hydroxy-9-(4-hydroxyphenyl)phenalen-1-one		3	2.55	92.8	425.4 / **234.7**
M131	(1R,4S,4aS,8aR)-4-isopropyl-1,6-dimethyl-3,4,4a,7,8,8a-hexahydro-2H-naphthalen-1-ol	1	1	3.07	17.7	359.2 / **287.8**
M132	(1S,2S,5R,8R,9R)-4,4,8-trimethyltricyclo[6.3.1.0^{1,5}]dodecane-2,9-diol	2	2	2.36	35.6	310.6 / **246.9**
M133	(1S,2R,6R,7S,9R,10R,12S)-10-hydroxy-9-methyl-7-[(3-methylbut-1-en-2-yl)oxy]-5,13-dimethylidene-4-oxo-3-oxatricyclo[7.4.0.0^{2,6}]tridecan-12-yl acetate	1	4	1.92	83.9	506.2 / **384.6**
M134	(1S,2S,6S,8S,9S,10R,13S)-9,13-dihydroxy-9,13-dimethyl-5-methylidene-4,12-dioxo-3-oxatricyclo[8.3.0.0^{2,6}]tridecan-8-yl2-methylpropanoate	2	5	1.77	101.8	464.4 / **322.5**

Molecule Name	IUPAC Name	# of H bond Donor	# of H bond Accep.	LoD	Polar Surface Area (ASA-P)	Solvent accesible area (ASA) / ASA+
M135	(1R,2R,4R,5R,7R,9E,11R)-4-hydroxy-4,9-dimethyl-14-methylidene-8,13-dioxo-6,12-dioxatricyclo[9.3.0.0^{5,7}]tetradec-9-en-2-yl2-methylpropanoate	1	5	1.98	105.7	458.5 / **311.5**
M137	(1R,2S,3R,6R,8S,12S,13S,14R,15R,16S,17S)-3,12,15,16-tetrahydroxy-9,13,17-trimethyl-5,18-dioxapentacyclo[12.5.0.0^{1,6}.0^{2,17}.0^{8,13}]nonadec-9-ene-4,11-dione	4	7	-1.59	119.9	365.2 / **253.2**
M138	(1R,2S,3R,6R,8S,13S,14R,15R,16S,17R)-17-(acetyloxy)-10,15,16-trihydroxy-9,13-dimethyl-4,11-dioxo-5,18-dioxapentacyclo[12.5.0.0^{1,6}.0^{2,17}.0^{8,13}]nonadec-9-en-3-yl3-methylbut-2-enoate	2	8	0.27	125.9	514.6 / **343.1**
M139	methyl(1R,2S,3R,6S,8R,13S,14R,15R,16S,17S)-15,16-dihydroxy-9,13-dimethyl-3-[(3-methylbutanoyl)oxy]-4,11-dioxo-10-{[(2R,3S,4R,5R,6S)-4,5,6-trihydroxy-2-(hydroxymethyl)oxan-3-yl]oxy}-5,18-dioxapentacyclo[12.5.0.0^{1,6}.0^{2,17}.0^{8,13}]nonadec-9-ene-17-carboxylate	6	13	-3.26	209.7	675.7 / **492.8**
M14	methyl (3S)-3-(3,4-dihydroxyphenyl)-3-[(2R,3R)-2-(3,4-dihydroxyphenyl)-3,5,7-trihydroxy-3,4-dihydro-2H-1-benzopyran-8-yl]propanoate	7	9	2.81	258.4	657.1 / **469.6**
M140	5-[(E)-2-(4-hydroxy-3-methoxyphenyl)ethenyl]benzene-1,3-diol	3	4	3.64	133.5	479.5 / **322.6**
M141	4-[(E)-2-(3,5-dihydroxyphenyl)ethenyl]benzene-1,3-diol	4	4	3.59	164	433.9 / **268.5**
M142	(2S,3R,4R,5S,6R)-5-{3-hydroxy-4-[(E)-2-(4-hydroxyphenyl)ethenyl]phenoxy}-6-(hydroxymethyl)oxane-2,3,4-triol	6	8	1.23	187.3	530.6 / **356.7**
M144	5-[(2S,3S)-5-[(E)-2-(3,5-dihydroxyphenyl)ethenyl]-2-(4-hydroxyphenyl)-2,3-dihydro-1-benzofuran-3-yl]benzene-1,3-diol	5	6	6.5	224.5	701.0 / **464.2**
M145	4-[(E)-2-(4-hydroxyphenyl)ethenyl]benzene-1,3-diol	3	3	3.89	125.2	428.6 / **270.7**
M147	(1R,2R,5S,8R,9S,10S,14R,17S,19S)-17-hydroxy-1,2,14,18,18-pentamethyl-8-(prop-1-en-2-yl)pentacyclo[11.8.0.0^{2,10}.0^{5,9}.0^{14,19}]henicosane-5-carboxylic acid	1	3	4.19	47.6	446.8 / **354.4**
M148	(1R,3'S,4R,4'S,6R,7S,8R,9R,10S,14S,18R,19S,21R)-7,8-dihydroxy-3',4',8,10,14-pentamethyl-5',15-dioxo-5,20-dioxaspiro[hexacyclo[11.9.0.0^{2,10}.0^{4,9}.0^{14,19}.0^{19,21}]docosane-6,2'-oxolan]-16-en-18-yl acetate	2	7	2.7	104.7	579.8 / **442.4**

Molecule Name	IUPAC Name	# of H bond Donor	# of H bond Accep.	LoD	Polar Surface Area (ASA-P)	Solvent accesible area (ASA) / ASA+
M149	(1S,2S,6R,7R,9R,11R,12S,15R,16R)-15-[(1R)-1-[(2R,4R,5R)-4,5-dimethyl-6-oxooxan-2-yl]-1-hydroxyethyl]-6-hydroxy-2,16-dimethyl-8-oxapentacyclo[9.7.0.0^{2,7}.0^{7,9}.0^{12,16}]octadec-4-en-3-one	2	5	3.44	70.1	513.5 / **398.8**
M15	(2S)-2-[4-hydroxy-3-(3-methylbut-2-en-1-yl)phenyl]-3,4-dihydro-2H-1-benzopyran-7-ol	2	3	4.99	87.6	495.3 / **360.2**
M150	5-methyl-5H,10H,11H-indolo[3,2-b]quinolin-11-one	1	2	2.59	42.4	420.4 / **254.6**
M151	1-{1-[5-(hydroxymethyl)furan-2-yl]-9H-pyrido[3,4-b]indol-3-yl}ethen-1-ol	2	3	1.26	119.2	517.7 / **347.3**
M152	(2E)-3-(4-hydroxy-3-methoxyphenyl)-N-[2-(4-hydroxyphenyl)ethyl]prop-2-enamide	3	4	2.98	130.9	570.3 / **392.6**
M153	(2R,3R)-2-(2,4-dihydroxyphenyl)-3,5,7-trihydroxy-3,4-dihydro-2H-1-benzopyran-4-one	4	7	1.17	202.7	433.6 / **238.3**
M155	(2S)-5,7-dihydroxy-2-(4-hydroxyphenyl)-6,8-bis(3-methylbut-2-en-1-yl)-3,4-dihydro-2H-1-benzopyran-4-one	2	5	5.12	107.6	651.9 / **450.2**
M156	2-(2,4-dihydroxyphenyl)-5,7-dihydroxy-4H-chromen-4-one	3	6	1.33	193.2	432.3 / **255.7**
M157	(2S)-2-(2,4-dihydroxyphenyl)-5,7-dihydroxy-3,4-dihydro-2H-1-benzopyran-4-one	3	6	1.96	159.8	389.1 / **231.7**
M16	(2S)-5,7-dimethoxy-4-oxo-2-phenyl-3,4-dihydro-2H-1-benzopyran-8-carbaldehyd	0	5	2.01	91.4	555.3 / **402.8**
M160	(1R,14R)-11,14-dihydroxy-17,18-dimethoxy-7,7-dimethyl-2,8,21-trioxapentacyclo[12.8.0.0^{3,12}.0^{4,9}.0^{15,20}]docosa-3,5,9,11,15(20),16,18-heptaen-13-one	2	8	2.15	128.6	640.0 / **501.9**
M161	(1S,14S)-11-hydroxy-17,18-dimethoxy-7,7-dimethyl-2,8,21-trioxapentacyclo[12.8.0.0^{3,12}.0^{4,9}.0^{15,20}]docosa-3,5,9,11,15(20),16,18-heptaen-13-one	1	7	2.95	105.9	628.4 / **491.8**
M162	(4aS,6aS,6bR,8aS,9R,10R,11R,12aR,12bR,14bS)-11-hydroxy-9-(hydroxymethyl)-10-{[(2E)-3-(4-hydroxyphenyl)prop-2-enoyl]oxy}-2,2,6a,6b,9,12a-hexamethyl-1,3,4,5,6,7,8,8a,10,11,12,12b,13,14b-tetradecahydropicene-4a-carboxylic acid	3	6	5.44	117.9	621.2 / **453.9**
M163	(1R,2R,6S,7S,9S,11S,12R,15S,16R)-15-[(1S)-1-[(2R)-4,5-dimethyl-6-oxo-3,6-dihydro-2H-pyran-2-yl]-1-hydroxyethyl]-6-hydroxy-16-(hydroxymethyl)-2-methyl-8-oxapentacyclo[9.7.0.0^{2,7}.0^{7,9}.0^{12,16}]octadec-4-en-3-one	3	6	2.46	92.5	535.8 / **407.2**
M164	(1R,2R,6R,7S,9S,11R,12R,14R,15R,16R)-15-[(1R,2S)-1-[(2R,3S,4S)-3,4-dimethyl-5-oxooxolan-2-yl]-1,2-dihydroxypropan-2-yl]-6,14-dihydroxy-2-methyl-8-oxapentacyclo[9.7.0.0^{2,7}.0^{7,9}.0^{12,16}]octadec-4-en-3-one	4	7	0.70	116.3	536.8 / **404.0**
M165	(1S,2R,4R,5R,10R,11R,14S,15R,18R)-15-[(1R)-1-[(2S)-4,5-dimethyl-6-oxo-3,6-dihydro-2H-pyran-2-yl]ethyl]-5,15-dihydroxy-10,14-dimethyl-3-oxapentacyclo[9.7.0.0^{2,4}.0^{5,10}.0^{14,18}]octadec-7-en-9-one	2	5	3.66	68.3	511.1 / **386.3**

Molecule Name	IUPAC Name	# of H bond Donor	# of H bond Accep.	LoD	Polar Surface Area (ASA-P)	Solvent accesible area (ASA) / ASA+
M166	(1R,2R,6R,7S,9S,11R,12S,14R,15R,16S)-15-[(1R)-1-[(2S,4S,5R)-4,5-dimethyl-6-oxooxan-2-yl]-1-hydroxyethyl]-6,14-dihydroxy-2,16-dimethyl-8-oxapentacyclo[9.7.0.0^{2,7}.0^{7,9}.0^{12,16}]octadec-4-en-3-one	3	6	2.13	90.5	529.0 / **410.2**
M17	(2S)-8-[(2S)-2-hydroxy-3-methylbut-3-en-1-yl]-5,7-dimethoxy-2-phenyl-2,3-dihydro-1-benzopyran-4-one	1	5	2.72	58.2	594.0 / **478.6**
M18	(2Z)-2-[1,2-bis(3,4-dihydroxyphenyl)-2-[(2Z)-6-hydroxy-3-oxo-1-benzofuran-2-ylidene]ethylidene]-6-hydroxy-1-benzofuran-3-one	4	10	1.79	264.1	671.5 / **336.8**
M19	(2S)-5-hydroxy-7-methoxy-8-[(1E)-3-oxobut-1-en-1-yl]-2-phenyl-2,3-dihydro-1-benzopyran-4-one	1	5	3.85	101.3	534.0 / **375.8**
M21prodrg	(2E)-4-{3-[(2Z)-3-(2,4-dihydroxyphenyl)prop-2-enoyl]-2,6-dihydroxyphenyl}-2-methylbut-2-en-1-yl (2E)-3-(4-hydroxyphenyl)prop-2-enoate	5	7	6.65	209.3	757.2 / 507.4
M22	(2R,3R)-3,5,8-trihydroxy-7-methoxy-2-phenyl-2,3-dihydro-1-benzopyran-4-one	3	6	1.93	134.3	449.2 / **315.8**
M23	(2Z)-1-(5,7-dimethoxy-2,2-dimethylchromen-6-yl)-3-methoxy-3-phenylprop-2-en-1-one	0	5	2.83	54.6	631.1 / **482.3**
M24	2-(2,4-dihydroxyphenyl)-3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-5,7-dihydroxychromen-4-one	3	6	4.15	173	698.3 / **470.3**
M25	(3R,4R,5R)-5-hydroxy-4-[(2S)-7-methoxy-4-oxo-2-phenyl-2,3-dihydro-1-benzopyran-8-yl]-2,2-dimethyloxolan-3-yl acetate	1	6	2.09	87.1	573.8 / **421.7**
M26	(2E)-1-[(2S,3S,6R)-3,12-dihydroxy-4,4-dimethyl-5,7-dioxatricyclo[6.4.0. 0^{2,6}]dodeca-1(12),8,10- trien-11-yl]-3-phenylprop-2-en-1-one	2	5	3.55	87.5	538.6 / **369.7**
M27	(10R,11R,15S,16R)-14-oxo-16-(3,4,5-trimethoxyphenyl)-4,6,13-trioxatetracyclo[7.7.0.0^{3,7}.0^{11,15}] hexadeca- 1,3(7),8-trien-10-yl acetate	0	7	1.89	156.5	690.3 / **546.9**
M28	(10S,11S,15R)-10-(3,4,5-trimethoxyphenyl)-4,6,13-trioxatetracyclo[7.7.0.0^{3,7}.0^{11,15}]hexadeca-1(9),2,7-trien-12-one	0	6	2.63	122.9	624.9 / **493.5**
M29	(1S,3S,4R,12S,13S,15S,16R,17R,18S,20R)-18-[(2R,3R,4R)-3,4-dimethyl-5-oxooxolan-2-yl]-17-hydroxy-4,17-dimethyl-19,21-dioxahexacyclo[11.6.1.1^{1,15}.0^{3,12}.0^{4,9}. 0^{16,20}]henicosa-6,9-dien-5-one	1	5	3.31	69.3	505.9 / **383.3**
M30	(1S,3S,4R,12R,13R,15S,16R,17R,18S,20S)-18-[(2R,3R,4R)-3,4-dimethyl-5-oxooxolan-2-yl]-13,17-dihydroxy-4,17-dimethyl-19,21-dioxahexacyclo [11.6.1.1^{1,15}.0^{3,12}.0^{4,9}. 0^{16,20}]henicosa-6,9-dien-5-one	2	6	2.03	85.9	514.4 / **388.7**
M31	(1S,2R,10S,11S,13S,14R,15S)-14-[(1R)-1-[(2R,4S,5R)-4,5-dimethyl-6-oxooxan-2-yl]-1-hydroxyethyl]-13-hydroxy-2-methyltetracyclo[8.7.0.0^{2,7}.0^{11,15}]heptadeca-4,7-diene-3,16-dione	2	5	2.88	60.5	495.5 / **381.4**

Molecule Name	IUPAC Name	# of H bond Donor	# of H bond Accep.	LoD	Polar Surface Area (ASA-P)	Solvent accesible area (ASA) / ASA+
M32	(1S,2R,10S,11S,13S,14R,15S)-14-[(1R,2R)-1-[(2R,3R,4R)-3,4-dimethyl-5-oxooxolan-2-yl]-1,2-dihydroxypropan-2-yl]-13-hydroxy-2-methyltetracyclo [8.7.0.0^{2,7}.0^{11,15}]heptadeca-4,7-diene-3,16-dione	3	6	1.88	**94.6**	**510.3 / 377.8**
M33	(1S,2S,10R,11S,13R,15R,16R,17S,19S)-15-[(2S,4S,5R)-4,5-dimethyl-6-oxooxan-2-yl]-13-ethoxy-10,15-dimethyl-9-oxo-14-oxapentacyclo [11.5.1.0^{2,11}.0^{5,10}.0^{16,19}]nonadeca-4,7-dien-17-yl formate	0	5	4.41	81.2	588.5 / **450.6**
M34	(1S,2S,10R,11S,13R,15R,16R,17S,19S)-15-[(2S,4S,5R)-4,5-dimethyl-6-oxooxan-2-yl]-13-methoxy-10,15-dimethyl-9-oxo-14-oxapentacyclo [11.5.1.0^{2,11}.0^{5,10}.0^{16,19}]nonadeca-4,7-dien-17-yl formate	0	5	4.07	88.5	584.3 / **447.5**
M35	(1S,3S,4R,7R,9S,10S,12S,15S,16R,17R,18R,20R)-18-[(2R,3R,4R)-3,4-dimethyl-5-oxooxolan-2-yl]-9,17-dihydroxy-4,17-dimethyl-19,21,22-trioxaheptacyclo [11.6.1.1^{1,15}.1^{7,10}.0^{3,12}.0^{4,9}.0^{16,20}]docosane-5,8-dione	2	8	1.38	108	497.7 / **358.7**
M36	(1R,5S,6R,14S)-6,11,14-trihydroxy-5,17,18-trimethoxy-7,7-dimethyl-2,8,21-trioxapentacyclo[12.8.0.0^{3,12}.0^{4,9}.0^{5,20}}]docosa-3(12),4(9),10,15(20),16,18-hexaen-13-one	3	10	0.60	146.2	669.8 / **535.6**
M37	(1S,2S,6R,7R,9R,10R,12S,13S)-12-(acetyloxy)-10,13-dihydroxy-9,13-dimethyl-5-methylidene-4-oxo-3-oxatricyclo[7.4.0. 0^{2,6}]tridecan-7-yl 2-methylpropanoate	2	5	0.90	115.9	543.4 / **394.7**
M38	5-[(E)-2-{3-[(2S)-2,3-dihydroxy-3-methylbutyl]-4-hydroxyphenyl}ethenyl]benzene-1,3-diol	5	5	3.12	134.7	504.0 / **344.1**
M39	4-[(E)-2-(3,5-dihydroxyphenyl)ethenyl]-6-(3-methylbut-2-en-1-yl)benzene-1,3-diol	4	4	5.13	156.9	585.7 / **389.9**
M40	5-[(E)-2-[4-hydroxy-3-(3-methylbut-2-en-1-yl)phenyl]ethenyl]benzene-1,3-diol	3	3	5.43	116.5	557.2 / **384.3**
M41	(2R,3S,4S,5S,6S)-5-{3-[(2R,3R)-5-[(E)-2-(3,5-dihydroxyphenyl)ethenyl]-2-(4-hydroxyphenyl)-2,3-dihydro-1-benzofuran-3-yl]-5-hydroxyphenoxy}-6-(hydroxymethyl)oxane-2,3,4-triol	8	11	3.84	263.6	761.7 / **521.0**
M42	4-[(E)-2-[(2S,3R)-3-(4-hydroxy-3,5-dimethoxyphenyl)-2-(hydroxymethyl)-2,3-dihydro-1,4-benzodioxin-6-yl]ethenyl]benzene-1,2-diol	4	8	3.88	171.9	715.5 / **524.2**
M43	methyl 3-[(1S,4R,5R,8S,9S,11S,12S,13R)-12-ethenyl-5-[(2R,5S)-5-hydroperoxy-6-methylhept-6-en-2-yl]-11-hydroxy-4,8-dimethyltetracyclo[7.5.0.0^{1,13}.0^{4,8}]tetradecan-13-yl]propanoate	2	4	5.45	107.4	608.6 / **470.0**
M44	3-[(1R,4R,5R,8S,11S,12S,13S)-12-ethenyl-11-hydroxy-4,8-dimethyl-5-[(2R)-6-methylhept-5-en-2-yl]tetracyclo [7.5.0.0^{1,13}.0^{4,8}]tetradecan-13-yl]propanoic acid	1	3	3.56	89.7	612.9 / **444.1**

Molecule Name	IUPAC Name	# of H bond Donor	# of H bond Accep.	LoD	Polar Surface Area (ASA-P)	Solvent accesible area (ASA) / ASA+
M45	(1R,2R,3'R,4S,4'R,6R,7S,8R,9S,10S,13R,14R,18S,19R,21R)-7,8,18-trihydroxy-3',4',8,10,14-pentamethyl-5,20-dioxaspiro[hexacyclo[11.9.0.0^{2,10}.0^{4,9}.0^{14,19}.0^{19,21}]docosane-6,2'-oxolane]-5',15-dione	3	7	2.05	104	525.3 / **395.6**
M46	(1R,2R,3'R,4S,4'R,6R,7S,8R,9R,10S,13S,14R,18S,19R,21R)-7,8,18-trihydroxy-3',4',8,10,14-pentamethyl-5,20-dioxaspiro[hexacyclo[11.9.0.0^{2,10}.0^{4,9}.0^{14,19}.0^{19,21}]docosane-6,2'-oxolan]-16-ene-5',15-dione	3	7	2.28	103.2	526.0 / **393.5**
M47	(1R,2R,6S,7R,9R,11R,12S,15R,16S)-15-[(1S)-1-[(2R,4S,5R)-4,5-dimethyl-6-oxooxan-2-yl]-1-hydroxyethyl]-6,15-dihydroxy-2,16-dimethyl-8-oxapentacyclo[9.7.0. 0$^{2,7.}$ 0$^{7,9.}$ 0^{12,16}]octadec-4-en-3-one	3	6	2.26	84	507.1 / **387.2**
M48	(1S,2R,6S,7R,9R,11S,12S,15R,16S)-6,15-dihydroxy-15-[(1S)-1-hydroxy-1-[(2R,4R,5R)-4-hydroxy-4,5-dimethyl-6-oxooxan-2-yl]ethyl]-2,16-dimethyl-8-oxapentacyclo [9.7.0. 0$^{2,7.}$ 0$^{7,9.}$ 0^{12,16}]octadec-4-en-3-one	4	7	0.74	104	523.8 / **402.8**
M49	(3R,4S,6S)-6-[(1R)-1-[(1R,2R,6S,9R,10R,14S,15S)-6,9-dihydroxy-2,15-dimethyl-3-oxotetracyclo[8.7.0.0^{2,7}.0^{11,15}]heptadeca-4,7-dien-14-yl]-1-hydroxyethyl]-3,4-dimethyloxan-2-one	3	5	3.04	92.2	526.4 / **404.5**
M50	11-methoxy-10H-indolo[3,2-b]quinoline	1	2	3.41	35	453.4 / **313.5**
M51	5H,10H-indolo[3,2-b]quinolin-11-one	2	2	3.64	57.6	398.7 / **232.1**
M52	(2R)-2-methyl-N-[(2R)-1-[(2E)-3-phenylprop-2-enoyl]pyrrolidin-2-yl]butanamide	1	2	3.12	32.5	510.0 / **373.3**
M53	2-methyl-N-[(2S)-1-[(2E)-3-phenylprop-2-enoyl]pyrrolidin-2-yl]propanamide	1	2	2.72	35.9	505.2 / **366.2**
M54	(1R,9S,13R,21S)-1-(2,4-dihydroxyphenyl)-17-(6-hydroxy-1-benzofuran-2-yl)-11-methyl-2,20-dioxapentacyclo [11.7.1.0^{3,8}.0^{9,21}.0^{14,19}]henicosa-3,5,7,11,14(19),15,17-heptaene-5,15-diol	5	7	7.01	189.1	674.2 / **487.3**
M55	2-[3-(4-hydroxyphenyl)propyl]-5-methoxyphenol	2	3	4.09	75.2	433.4 / **315.5**
M56	4-[3-(4-hydroxyphenyl)propyl]-3-methoxyphenol	2	3	4.09	71.9	378.6 / **282.8**
M57	(2E)-3-(3,4-dimethoxyphenyl)prop-2-enal	0	3	1.48	72.2	424.4 / **329.6**
M58	(2R)-2-hydroxy-1-(4-hydroxy-3-methoxyphenyl)propan-1-one	2	4	0.43	104.1	402.7 / **284.4**
M59	(2R)-3-hydroxy-1,2-bis(4-hydroxy-3-methoxyphenyl)propan-1-one	3	6	1.33	144.6	553.8 / **417.2**
M60	(2E)-3-(4-hydroxy-3-methoxyphenyl)prop-2-enoic acid	1	4	-1.69	139	391.4 / **230.5**

Molecule Name	IUPAC Name	# of H bond Donor	# of H bond Accep.	LoD	Polar Surface Area (ASA-P)	Solvent accesible area (ASA) / ASA+
M61	3,4,5-trihydroxybenzoic acid	3	5	-2.56	**199.8**	306.6 / **130.9**
M62	methyl 3,4,5-trihydroxybenzoate	3	4	1.25	163.7	352.8 / **243.3**
M63	(2E)-3-(4-hydroxy-3,5-dimethoxyphenyl)prop-2-enal	1	4	1.19	106.2	450.0 / **346.6**
M64	(2E)-3-[(2S,3R)-2-(4-hydroxy-3-methoxyphenyl)-3-(hydroxymethyl)-2,3-dihydro-1-benzofuran-5-yl]prop-2-enal	2	5	2.03	134.9	548.4 / **391.3**
M65	2-[5-hydroxy-3-methoxy-2-(3-methylbut-2-en-1-yl)phenyl]-1-benzofuran-6-ol	2	3	4.47	104.1	543.1 / **402.0**
M66	5-(6-hydroxy-1-benzofuran-2-yl)benzene-1,3-diol	3	3	2.88	148.4	430.4 / **258.3**
M67	5-[6-hydroxy-5-(3-methylbut-2-en-1-yl)-1-benzofuran-2-yl]benzene-1,3-diol	3	3	4.40	121.5	535.8 / **366.0**
M68	4-[(1S,2S,5R,7R,10S,11S,14R,15R)-5-{[(2S,3R,4R,5R,6S)-3,5-dihydroxy-4-methoxy-6-methyloxan-2-yl]oxy}-11-hydroxy-2,5,15-trimethyltetracyclo[8.7.0.0^{2,7}.0^{11,15}]heptadecan-14-yl]-2,5-dihydrofuran-2-one	3	7	2.76	120.9	621.9 / **480.5**
M70	7-hydroxy-6-methoxychromen-2-one	1	3	0.98	120.2	357.3 / **255.7**
M71	methyl(2S,3R,4R,5S,6R)-2,3-dihydroxy-10,12-dimethoxy-6-(4-methoxyphenyl)-5-phenyl-7-oxatricyclo[6.4.0.0^{2,6}]dodeca-1(12),8,10-triene-4-carboxylate	2	7	2.19	92.1	723.1 / **582.6**
M72	(4E,6E)-1,7-bis(4-hydroxyphenyl)hepta-4,6-dien-3-one	2	3	5.14	104.8	541.1 / **392.4**
M73	(1R,4aR,4bR,7S,10aR)-7-ethenyl-1,4a,7-trimethyl-3,4,4b,5,6,8,10,10a-octahydro-2H-phenanthrene-1-carboxylic acid	0	2	2.65	38.6	387.0 / **293.2**
M74	(1R,4aS,7S,10aR)-7-ethenyl-1,4a,7-trimethyl-9-oxo-2,3,4,5,6,8,10,10a-octahydrophenanthrene-1-carboxylic acid	0	3	1.89	55.5	393.4 / **285.3**
M75	(9Z,12Z)-nonadeca-9,12-dienoic acid	0	2	3.89	72.5	643.2 / **444.8**
M76	(2R)-7-hydroxy-2-[4-hydroxy-3-(3-methylbut-2-en-1-yl)phenyl]-2,3-dihydro-1-benzopyran-4-one	2	4	3.56	115.4	546.0 / **390.9**
M77	(2R,3R)-2-(4-hydroxyphenyl)-3-methyl-3,4-dihydro-2H-1-benzopyran-5,7-diol	3	4	3.55	128.5	437.1 / **301.5**
M78	(2S)-7-hydroxy-2-(4-hydroxyphenyl)-6-(3-methylbut-2-en-1-yl)-2,3-dihydro-1-benzopyran-4-one	1	4	3.47	104.3	576.1 / **352.4**
M79	(2E)-1-[2,4-dihydroxy-5-(3-methylbut-2-en-1-yl)phenyl]-3-(4-hydroxyphenyl)prop-2-en-1-one	3	4	4.90	134.4	560.7 / **373.3**

Molecule Name	IUPAC Name	# of H bond Donor	# of H bond Accep.	LoD	Polar Surface Area (ASA-P)	Solvent accesible area (ASA) / ASA+
M80	3,5,7-trihydroxy-2-[4-hydroxy-3-(3-methylbut-2-en-1-yl)phenyl]-8-(3-methylbut-2-en-1-yl)chromen-4-one	2	6	2.61	152.7	699.8 / **442.1**
M81	8-methoxy-12,12-dimethyl-4-phenyl-3,11-dioxatricyclo [8.4.0. $0^{2,7}$]tetradeca-1,4,7,9,13- pentaen-6-one	0	4	2.81	61.6	565.2/ **395.8**
M82	(2S)-5,7-dimethoxy-8-(3-methylbut-2-en-1-yl)-2-phenyl-2,3-dihydro-1-benzopyran-4-one	0	4	3.88	50.5	661.1/ **508.9**
M83	(2R,3S)-2-(3,4-dihydroxyphenyl)-3,4-dihydro-2H-1-benzopyran-3,5,7-triol	5	6	1.79	188.9	412.6 / **294.1**
M84	(2E)-1-(2,4-dihydroxyphenyl)-3-phenylprop-2-en-1-one	2	3	3.72	103.3	450.2 / **295.1**
M85	(2Z)-6-hydroxy-2-[(4-hydroxy-3-methoxyphenyl)methylidene]-1-benzofuran-3-one	1	5	1.54	137.3	494.7 / **289.0**
M86	5,7-dihydroxy-2-(4-hydroxy-3-methoxyphenyl)chromen-4-one	2	6	1.50	173.8	505.0 / **320.9**
M88	(2S,3R)-3,5,7-trihydroxy-2-(4-hydroxyphenyl)-2,3-dihydro-1-benzopyran-4-one	3	6	1.47	176.6	425.9 / **228.1**
M89	(2R,3S,4aS,5R,7S,8aS)-2-(3,4-dihydroxyphenyl)-3,5,7-trihydroxy-octahydro-1-benzopyran-4-one	5	7	-0.63	178	393.7/ **262.4**
M90	(2S)-2-(2,4-dihydroxyphenyl)-7-hydroxy-8-(3-methylbut-2-en-1-yl)-2,3-dihydro-1-benzopyran-4-one	2	5	3.16	116.5	512.4 / **330.8**
M92	5,7-dihydroxy-3-(4-hydroxyphenyl)chromen-4-one	2	5	3.07	154.3	422.9 / **224.3**
M93	(2E)-1-(2-hydroxy-4,6-dimethoxyphenyl)-3-phenylprop-2-en-1-one	1	4	3.57	81.3	550.7 / 409.7
M94	7-hydroxy-3-phenylchromen-4-one	1	3	3.11	85.4	412.9 / 269.5
M95	(2S)-5-hydroxy-7-methoxy-2-phenyl-2,3-dihydro-1-benzopyran-4-one	1	4	2.95	90.3	479.8 / 348.0
M96	(2E)-1-[2,4-dihydroxy-3-(3-methylbut-2-en-1-yl)phenyl]-3-(4-hydroxyphenyl)prop-2-en-1-one	3	4	4.87	126	585.1 / 383.4
M98	3,5,7-trihydroxy-2-[4-hydroxy-3-(3-methylbut-2-en-1-yl)phenyl]chromen-4-one	3	6	0.95	181.8	587.2 / 377.3
M99	(2R,3R)-3,5,7-trihydroxy-2-[4-hydroxy-3-(3-methylbut-2-en-1-yl)phenyl]-8-(3-methylbut-2-en-1-yl)-2,3-dihydro-1-benzopyran-4-one	3	6	4.45	128.7	609.6 / 424.8
Macrophyllogenin	(1S,3S,4S,6S,8R,10S,11S,12S,15R,16R)-15-[(2R,5S)-5,6-dihydroxy-6-methylheptan-2-yl]-7,7,12,16-tetramethyl pentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}octadecane-4,6,10-triol	5	5	2.88	67.1	526.5 / 428.0
Macrophyllosapo ninA	(3S,4R,5R,6R)-6-{[(3S,6R)-6-[(1S,3S,4S,6S,8R,10S,11S,12S,15R,16R)-4,10-dihydroxy-7,7,12,16-tetramethyl-6-{[(2R,3S,4S,5R,6R)-3,4,5-trihydroxy-6-methyloxan-2-yl]oxy}pentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}] octadecan-15-yl]-2-hydroxy-2-methylheptan-3-yl]oxy}-4,5-dihydroxyoxan-3-yl acetate	8	13	0.83	168.5	869.9 / 697.3

Molecule Name	IUPAC Name	# of H bond Donor	# of H bond Accep.	LoD	Polar Surface Area (ASA-P)	(ASA) / ASA+
MacrophyllosaponinB	(2R,3S,4S,5R,6R)-2-{[(1S,3S,4S,6S,8R,10S,11S,12S, 15R,16R)-4,10-dihydroxy-15-[(2R,5S)-6-hydroxy-6-methyl-5-{[(2R,3R,4S,5S)-3,4,5-trihydroxyoxan-2-yl]oxy} heptan-2-yl]-7,7,12,16-tetramethylpentacyclo[9.7.0.0^{1,3}. 0^{3,8}.0^{12,16}]octadecan-6-yl]oxy}-6-methyloxane-3,4,5-triol	9	13	0.41	173.2	769.1 / 616.5
MacrophyllosaponinC	(2R,3S,4S,5R,6R)-2-{[(1S,3S,4S,6S,8R,10S,11S,12S, 15R,16R)-4,10-dihydroxy-15-[(2R,5S)-5-hydroxy-6-methyl-6-{[(2S,3R,4R,5R)-3,4,5-trihydroxyoxan-2-yl]oxy} heptan-2-yl]-7,7,12,16–tetramethylpentacyclo[9.7.0.0^{1,3}. 0^{3,8}.0^{12,16}]octadecan-6-yl]oxy}-6-methyloxane-3,4,5-triol	9	13	0.41	193.2	797 / 641.2
MacrophyllosaponinD	(2R,3S,4S,5R,6R)-2-{[(1S,3S,4S,6S,8R,10S,11S,12S, 15R,16R)-15-[(2R,5S)-5-{[(2R,3R,4S,5S)-4,5-dihydroxy-3-{[(2S,3R,4R,5R)-3,4,5-trihydroxyoxan-2-yl]oxy}oxan-2-yl]oxy}-6-hydroxy-6-methylheptan-2-yl]-4,10-dihydroxy-7,7,12,16-tetramethylpentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}] octadecan-6-yl]oxy}-6-methyloxane-3,4,5-triol	11	17	-1.04	216	859.9 / 688.2
MacrophyllosaponinE	(2S,3R,4R,5S,6R)-2-{[(3S,6R)-6-[(1S,3S,4S,6S,8R,10S, 11S,12S,15R,16R)-4,10-dihydroxy-7,7,12,16-tetramethyl-6-{[(2R,3S,4S,5S,6S)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxy}pentacyclo [9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan -15-yl]-3-hydroxy-2-methylheptan-2-yl]oxy}-6-(hydroxymethyl)oxane-3,4,5-triol	11	15	-1.35	223.2	827.5 / 662.2
MMD	(10R)-2-methyl-10-[(9R)-2-methyl-4,5-dioxido-10-oxo-9,10-dihydroanthracen-9-yl]-9-oxo-9,10-dihydroanthracene-1,8-bis(olate)	4	6	8.36	158.5	619.7 / 424.4
PaucifloralF	(2S,3S)-3-(3,5-dihydroxyphenyl)-2-(4-hydroxyphenyl)-2,3-dihydro-1H-inden-1-one	3	4	4.0	151.2	510.2 / 318.8
PMC	(3S,4S,5S)-3,4-dihydroxy-5-(hydroxymethyl)-3-[(2E)-3-[(2R,3R)-3-nonyloxiran-2-yl]prop-2-enoyl]pyrrolidin-2-one	4	6	1.16	149.2	618.6 / 444.1
PMC-A	(3R,4S,5S)-3,4-dihydroxy-5-(hydroxymethyl)-3-[(2E,4E)-tetradeca-2,4-dienoyl]pyrrolidin-2-one	4	5	2.03	138.9	644.0 / 468.0
PMC-B	(3R,4S,5S)-3,4-dihydroxy-5-(hydroxymethyl)-3-tetradecanoylpyrrolidin-2-one	4	5	2.22	125.5	637.8 / 473.2
PMC-C	(3S)-3-tetradecanoylpyrrolidin-2-one	1	2	4.68	56.9	631.4 / 471.4
PMC-D	(3R)-3-hydroxy-3-tetradecanoylpyrrolidin-2-one	2	3	3.61	86.2	548.9 / 412.9
PMC-E	(3S,5R)-5-(hydroxymethyl)-3-tetradecanoylpyrrolidin-2-one	2	3	4.01	49.6	564.5 / 440.4
PMC-F	(3R,5R)-3-hydroxy-5-(hydroxymethyl)-3-tetradecanoylpyrrolidin-2-one	3	4	2.94	79.6	556.0 / 434.7
PMC-G	(3S)-3-tetradecanoyloxolan-2-one	0	2	6.03	65.8	653.7 / 472.6
PMC-H	(3R)-3-hydroxy-3-tetradecanoyloxolan-2-one	1	3	4.29	61.9	595.4 / 451.9

Molecule Name	IUPAC Name	# of H bond Donor	# of H bond Accep.	LoD	Polar Surface Area (ASA-P)	Solvent accesible area (ASA) / ASA+
PMC-I	(3R)-3-tetradecylpyrrolidin-2-one	1	1	5.07	31.3	599.3 / 460.4
PMC-J	(3S)-3-acetylpyrrolidin-2-one	1	2	-0.30	66.4	272.6 / 176.3
Pterostilbene	4-[(E)-2-(3,5-dimethoxyphenyl)ethenyl]phenol	1	3	3.97	78.1	549.4 / 402.3
Pterostilbene-dehydrodimer	4-[(2S,3S)-3-(3,5-dimethoxyphenyl)-5-[(E)-2-(3,5-dimethoxyphenyl)ethenyl]-2,3-dihydro-1-benzofuran-2-yl]phenol	1	6	6.66	122.1	912.1 / 707.0
gquadrangularinA	(1E,2R,3R)-2-(3,5-dihydroxyphenyl)-3-(4-hydroxyphenyl) -1-[(4-hydroxyphenyl)methylidene]-2,3-dihydro-1H-indene-4,6-diol	6	6	6.40	214.9	578.3 / 382.9
gres10	3,5-dimethoxybenzaldehyde	0	3	1.22	74.9	408.1 / 310.5
gres11	(S)-{2,4-dimethoxy-6-[(E)-2-(4-methoxyphenyl)ethenyl]phenyl}(3,5-dimethoxyphenyl)methanol	1	6	4.51	102.2	842.3 / 684.3
gres13	(1S,2S)-1-(3,5-dimethoxyphenyl)-5,7-dimethoxy-2-(4-methoxyphenyl)-2,3-dihydro-1H-indene	0	5	4.71	76.2	782.8 / 640.7
gres14	(1S,2S,3S)-3-(3,5-dimethoxyphenyl)-4,6-dimethoxy-2-(4-methoxyphenyl)-2,3-dihydro-1H-inden-1-yl 2,2,2-trifluoroacetate	0	9	5.10	203.1	893.8 / 655.6
gres15	(1S,2S,3S)-3-(3,5-dimethoxyphenyl)-4,6-dimethoxy-2-(4-methoxyphenyl)-2,3-dihydro-1H-inden-1-ol	1	6	3.56	96.8	800.6 / 662.9
gres16	(1S,2S,3S)-3-(3,5-dimethoxyphenyl)-4,6-dimethoxy-2-(4-methoxyphenyl)-1-{[(4-methoxyphenyl)methyl]sulfanyl}-2,3-dihydro-1H-indene	0	7	6.43	100.2	1042.8 / 835.2
gres17	(1R)-1-(3,5-dihydroxyphenyl)-2-(4-hydroxyphenyl)-3-[(4-hydroxyphenyl)methyl]-1H-indene-5,7-diol	6	6	6.37	178.7	531.0 / 372.8
gres18	2-bromo-1-[(E)-2-(3,5-dimethoxyphenyl)ethenyl]-3,5-dimethoxybenzene	0	4	4.54	72.9	680.0 / 547.9
gres20	(R)-{2-[(E)-2-(3,5-dimethoxyphenyl)ethenyl]-4,6-dimethoxyphenyl}(4-methoxyphenyl)methanol	1	6	4.51	87.2	738.6 / 621.7
gres23	(1E,2R,3R)-2-(3,5-dimethoxyphenyl)-4,6-dimethoxy-3-(4-methoxyphenyl)-1-[(4-methoxyphenyl)methylidene]-2,3-dihydro-1H-indene	0	6	6.62	86.2	909.7 / 754.7
gres24	(1R,2R,3E)-4-bromo-2-(3,5-dimethoxyphenyl)-5,7-dimethoxy-1-(4-methoxyphenyl)-3-[(4-methoxyphenyl)methylidene]-2,3-dihydro-1H-indene	0	6	7.42	99.2	990.7 / 798.1
gres25	(1R,2S,3E)-4-bromo-2-(2-bromo-3,5-dimethoxyphenyl)-5,7-dimethoxy-1-(4-methoxyphenyl)-3-[(4-methoxyphenyl)methylidene]-2,3-dihydro-1H-indene	0	6	8.21	96.8	953.8 / 761.7
gres27	(1R,8R,9R,16R)-1,3,11-tribromo-4,6,12,14-tetramethoxy-8,16-bis(4-methoxyphenyl)-9-methyltetracyclo [7.7.0.0^{2,7}.0^{10,15}]hexadeca-2(7),3,5,10(15),11,13-hexaene	0	6	8.25	88.9	935.7 / 745.0

Molecule Name	IUPAC Name	# of H bond Donor	# of H bond Accep.	LoD	Polar Surface Area (ASA-P)	Solvent accesible area (ASA) / ASA+
gres28	(1E,2S,3S)-3-(3,5-dimethoxyphenyl)-4,6-dimethoxy-2-(4-methoxyphenyl)-1-[(4-methoxyphenyl)methylidene]-2,3-dihydroindene	0	6	6.62	90.9	958.0 / 790.7
gres30	(1R,8R,9S,16S)-3,9,11-tribromo-4,6,12,14-tetramethoxy-8,16-bis(4-methoxyphenyl)tetracyclo[7.6.1.0^{2,7}.0^{10,15}]hexadeca-2(7),3,5,10(15),11,13-hexaene	0	6	7.74	85.5	928.3 / 730.7
gres31	{2,4-dimethoxy-6-[(E)-2-(4-methoxyphenyl)ethenyl]phenyl}(3,5-dimethoxyphenyl)methanone	0	6	4.72	88.1	775.4 / 643.5
gres32	{2-[(1R)-1-bromo-2-(4-methoxycyclohexa-2,5-dien-1-ylidene)ethyl]-4,6-dimethoxyphenyl}(3,5-dimethoxyphenyl)methanone	0	6	3.92	87	802.4 / 678.3
gres33	(9S,10R)-9-bromo-4,6,12,14-tetramethoxy-10-(4-methoxyphenyl)tricyclo[9.4.0.0^{3,8}]pentadeca-1(11),3,5,7,12,14-hexaen-2-one	0	6	4.35	87	743.1 / 606.3
gres34	(10R)-4,6,12,14-tetramethoxy-10-(4-methoxyphenyl)tricyclo[9.4.0.0^{3,8}]pentadeca-1(11),3,5,7,12,14-hexaen-2-one	0	6	3.98	91.9	791.6 / 655.4
gres36	(9R,10S)-5,7,13,15-tetramethoxy-10-(4-methoxyphenyl)-2-oxotricyclo[9.4.0.0^{3,8}]pentadeca-1(15),3(8),4,6,11,13-hexaen-9-yl acetate	0	7	3.27	114.6	806.2 / 665.0
gres37	(10S)-5,7,13,15-tetramethoxy-10-(4-methoxyphenyl)tricyclo[9.4.0.0^{3,8}]pentadeca-1(15),3(8),4,6,11,13-hexaene-2,9-dione	0	7	2.89	106.2	750.0 / 611.7
gres38	(9S,10S)-4,6,9,12,14-pentahydroxy-10-(4-hydroxyphenyl)tricyclo[9.4.0.0^{3,8}]pentadeca-1(11),3,5,7,12,14-hexaen-2-one	6	7	2.98	214.1	482.0 / 325.8
gres9	2-bromo-1,5-dimethoxy-3-[(E)-2-(4-methoxyphenyl)ethenyl]benzene	0	3	4.79	54.9	627.7 / 473.1
gResrevatrol-trans-dehydrodime	5-[(2S,3S)-5-[(E)-2-(3,5-dihydroxyphenyl)ethenyl]-2-(4-hydroxyphenyl)-2,3-dihydro-1-benzofuran-3-yl]benzene-1,3-diol	5	6	6.50	225.1	683.6 / 455.2
gResveratrol	5-[(E)-2-(4-hydroxyphenyl)ethenyl]benzene-1,3-diol	3	3	3.89	133.8	430.8 / 248.6
gsjh23	4-methyl-1-N-(3-phenylpropyl)benzene-1,2-diamine	3	2	3.72	55.6	509.0 / 354.5
gTorjanosideH	(2R,3R,4R,5S,6R)-2-{[(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-6-{[(2S,3S,4S,5R)-4,5-dihydroxy-3-{[(2S,3S,4S,5S)-3,4,5-trihydroxyoxan-2-yl]oxy}oxan-2-yl]oxy}-14-hydroxy-15-[(2R,5S)-5-(2-hydroxypropan-2-yl)-2-methyl oxolan-2-yl]-7,7,12,16-tetramethylpentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-9-yl]oxy}-6-(hydroxymethyl)oxane-3,4,5-triol	11	18	-2.56	241.9	872.3 / 688.0

Molecule Name	IUPAC Name	# of H bond Donor	# of H bond Accep.	LoD	Polar Surface Area (ASA-P)	Solvent accesible area (ASA) / ASA+
gTrojanosideA	(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-15-[(2R,5S)-5-(2-hydroxypropan-2-yl)-2-methyloxolan-2-yl]-7,7,12,16-tetramethyl-9-{[(2R,3R,4R,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxy}-6-{[(2S,3S,4S,5R)-3,4,5-trihydroxyoxan-2-yl]oxy}pentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}] octadecan-14-yl acetate	8	14	-0.69	197	865.4 / 692.4
gTrojanosideB	(2S,3R,4R,5S,6R)-2-({2-[(2S,5R)-5-[(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-14-hydroxy-7,7,12,16-tetra methyl-9-{[(2R,3R,4R,5S,6R)-3,4,5-trihydroxy-6-(hydroxy methyl)oxan-2-yl]oxy}-6-{[(2S,3S,4S,5R)-3,4,5-trihydroxyoxan-2-yl]oxy}pentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}] octadecan-15-yl]-5-methyloxolan-2-yl]propan-2-yl}oxy)-6-(hydroxymethyl)oxane-3,4,5-triol	12	19	-3.24	284.9	914.2 / 719.7
gTrojanosideC	(2S,3R,4R,5S,6R)-2-{[(3S,6R)-6-[(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-6-{[(2S,3S,4S,5R)-4,5-dihydroxy-3-{[(2S,3S,4S,5R,6R)-3,4,5-trihydroxy-6-methyloxan-2-yl]oxy}oxan-2-yl]oxy}-9,14-dihydroxy-7,7,12,16-tetra methylpentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-15-yl]-2-hydroxy-2-methylheptan-3-yl]oxy}-6-(hydroxymethyl)oxane-3,4,5-triol	12	18	-1.73	234.7	895.1 / 716.3
gTrojanosideD	(2R,3S,4S,5S,6S)-2-{[(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-14-hydroxy-15-[(2R,5S)-6-hydroxy-6-methyl-5-{[(2S,3R,4R,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl) oxan-2-yl]oxy}heptan-2-yl]-7,7,12,16-tetramethyl-9-{[(2R,3R,4R,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl) oxan-2-yl]oxy}pentacyclo [9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan -6-yl]oxy}-6-(hydroxymethyl)oxane-3,4,5-triol	14	20	-3.49	280.1	926.8 / 735.6
gTrojanosideE	(2S,3R,4R,5S,6R)-2-{[(3S,6R)-6-[(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-6-{[(2S,3S,4S,5R)-4,5-dihydroxy-3-{[(2S,3S,4S,5R,6R)-3,4,5-trihydroxy-6-methyloxan-2-yl]oxy}oxan-2-yl]oxy}-14-hydroxy-7,7,12,16-tetramethyl-9-{[(2R,3R,4R,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl) oxan-2-yl]oxy}pentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-15-yl]-2-hydroxy-2-methylheptan-3-yl]oxy}-6-(hydroxy methyl)oxane-3,4,5-triol	15	23	-3.85	264.8	988.2 / 788.9
gTrojanosideF	(2S,3R,4R,5S,6R)-2-{[(3S,6R)-6-[(1S,3R,6S,8S,9S,11S,12S,14S,15R,16R)-6-{[(2S,3S,4S,5R)-4,5-dihydroxy-3-{[(2S,3S,4S,5S)-3,4,5-trihydroxyoxan-2-yl]oxy}oxan-2-yl]oxy}-14-hydroxy-7,7,12,16-tetramethyl-9-{[(2R,3R,4R, 5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxy}pentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-15-yl]-2-hydroxy-2-methylheptan-3-yl]oxy}-6-(hydroxymethyl)oxane-3,4,5-triol	15	23	-4.26	330.4	1021.5 / 812.8

Molecule Name	IUPAC Name	# of H bond Donor	# of H bond Accep.	LoD	Polar Surface Area (ASA-P)	Solvent accesible area (ASA) / ASA+
gTrojanosideI	(1S,3R,6S,8S,9S,12R,14S,15R,16R)-6-{[(2S,3S,4S,5R)-3,4-bis(acetyloxy)-5-hydroxyoxan-2-yl]oxy}-15-[(2R,5S)-5-(2-hydroxypropan-2-yl)-2-methyloxolan-2-yl]-7,7,12,16-tetramethyl-9-{[(2R,3R,4R,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxy}pentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-14-yl acetate	6	14	0.17	175.8	878.0 / 709.8
gTrojanosideJ	(2S,3S,4S,5R)-5-(acetyloxy)-2-{[(1S,3R,6S,8S,9S,12R,15S,16R)-15-[(2R,5S)-5-(2-hydroxypropan-2-yl)-2-methyloxolan-2-yl]-7,7,12,16-tetramethyl-9-{[(2S,3R,4R,5R)-3,4,5-trihydroxyoxan-2-yl]oxy}pentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}]octadecan-6-yl]oxy}-3-{[(2S,3S,4S,5R,6R)-3,4,5-trihydroxy-6-methyloxan-2-yl]oxy}oxan-4-yl acetate	7	16	0.69	178.2	946.1 / 767.8
gTrojanosideK	(2R,3S,4R,5R,6R)-2-(hydroxymethyl)-6-{[(1S,3R,6S,8S,9S,11R,12S,14S,15R,16R)-15-[(2R,5S)-5-(2-hydroxypropan-2-yl)-2-methyloxolan-2-yl]-7,7,12,16-tetramethyl-14-{[(2R,3R,4R,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxy}-6-{[(2S,3S,4S,5R)-3,4,5-trihydroxyoxan-2-yl]oxy}pentacyclo[9.7.0.0^{1,3}.0^{3,8}.0^{12,16}] octadecan-9-yl]oxy}oxane-3,4,5-triol	12	19	-3.24	281.3	894.2 / 706.8
gvaticanolC	(1R,4R,5S,11S,12S,15R,16R,22R)-4,15-bis(3,5-dihydroxyphenyl)-5,11,16,22-tetrakis(4-hydroxyphenyl)-11-methyl-6,17-dioxahexacyclo[10.9.1.0^{2,10}.0^{3,7}.0^{13,21}.0^{14,18}]docosa-2(10),3(7),8,13(21),14(18),19-hexaene-9,19-diol	10	12	11.35	322.2	899.6 / 622.6

Appendix 5

SUPPLEMENTARY MATERIAL FOR CHAPTER 5

Scores from resultant binding modes of each molecule are used in coralation curves. Average of resultant binding modes are also searched for beter coralations as can be seen in table7.1. Starting with top 10% of binding mode results are averaged with periodic intervals. The highest regression constant is achived with averages of all data coralation curves.

Top % # of binding Mode Averages	Coralation Curves R^2		
	Chem Score & Gold Score	Chem Score& Autodock	Autodock & Gold Score
10%	0.127	0.051	0.0001
20%	0.19	0.15	0.01
30%	0.2297	0.11	0.041
40%	0.264	0.205	0.050
50%	0.291	0.244	0.070
60%	0.309	0.280	0.089
70%	0.3347	0.2692	0.0897
80%	0.3549	0.2928	0.1055
90%	0.3798	0.2967	0.108
All	0.404	0.319	0.117

Table A.5 Rearranged binding modes of each molecule are used in correlation curves averaged with periodic intervals. As can be seen regression constant increases with increasing percents of averaged data.

BIBLIOGRAPHY

[1]   J. Adnane, P. Gaudray, CA. Dionne, *BEK and FLG, two receptors to members of the FGF family, are amplified in subsets of human breast cancers.* Oncogene, 1991. 6: p. 659–663.

[2]   C. Nakanishi, M. Toi, *Nuclear factor-κβ inhibitors as sensitizers to anticancer drugs.* Nature reviews;cancer, 2005 april. V: 5.

[3]   G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, *Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function.* J. Computational Chemistry, 1998. 19: p. 1639-1662.

[4]   G. Jones, P. Willett and R. C. Glen, *Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation.* J. Mol. Biol., 1995. 245: p. 43-53.

[5]   A. Wolf, M. Zimmermann, M. Hofmann-Apitius, *GeneticALGORITHM-Alternative to Consensus ScoringsA New Approach Toward the Qualitative Combination of Docking Algorithms.* J. Chem. Inf. Model., 2007.

[6]   D. S. Latchman, *Transcription Factors: An overview,* Int. J. Biochem. Cell Biol., 1997, V: 29, 12: p. 1305-1312.

[7]   http://www.embl-grenoble.fr/groups/dna/p52.gif, last visited 02.09.08.

[8]   L.David, V. Jagt, M. D. Dorraine, F. Steve, *Method and compounds for cancer treatment utilizing NF-κB as a diret or ultimate target for small molecule inhibitore.* Patent Application Publication, 2006 nov. 16.

[9]   C. Nakanishi and T. Masakazu , *Nuclear Factor-κB Inhibitors As Sensitizers To Anticancer Drugs.* Nature Reviews; Cancer, 2005. V:5, p.297.

[10]   http://www.biochem.umd.edu/biochem/kahn/molmachines/enhancesomes/ NFkB.html, last visited 05.09.08.

[11]   A. Garg and B. B. Aggarwal, *Nuclear transcription factor-_B as a target for cancer drug development.* Leukemia, 2002. 16: p.1053-1068.

[12]   J.L. Luo, H. Kamata and M. Karin, *IKK/NF-κB signaling: balancing life and death – a new approach to cancer theraphy.* J. of Clinical Inv., 2005. V: 115, 10: p. 2625-2632.

[13]   J. Boik, *Natural Compounds in Cancer Therapy.* Quality Books , 2001. p. 51-62.

[14]   D.C. Duffey, Z. Chen, G. Dong, F.G. Ondrey, J.S. Wolf and K. Brown *et al.,* *Expression of a dominant-negative mutant inhibitor-kappaBalpha of nuclear factor-kappaB in human head and neck squamous cell carcinoma inhibits survival, proinflammatory cytokine expression, and tumor growth in vivo.* Cancer Res., 1999. 59: p. 3468–3474.

[15]   A. C. Bharti, B. B. Aggarwal, *Nuclear factor-kappa B and cancer: its role in prevention and therapy ,* Biochemical Pharmacology, 2002. 64: p. 883-888.

[16]   T. Huxford, D. Mishler, C. B. Phelps, D. B. Huang, L. L. Sengchanthalangsy, R. Reeves, C. A. Huges, E. A. Komives and G. Gosh, *Solvent Exposed Non-contacting Amino Acids Play a Critical Role in NF-κB/IκBα Complex Formation.* J. Mol. Biol., 2002. 324: p.587-597.

[17]   B. S DeDecker, *Allosteric drugs: thinking outside the active-site box.* Chemistry & Biology, 2000. V: 7, 5: p. 103-107.

[18]   V. Pande et. al., *A molecular modeling study of inhibitors of nuclear factor kappa-B (p50).* Journal of Computer-Aided Molecular Design, 2003. 17: p.825-836.

[19]   R. K. Sharma, B. S. Garg, H. Kurosaki, M. Goto, M. Otsuko, T. Yamamoto and J. I. Inoue, *Aurine Tricarboxylic Acid, a Potent Metal-Chelating Inhibitor of NFκB-DNA Binding.* Bioorganic & Medicinal Chemistry, 2000. 8: p. 1819-1823.

[20]   J.Latzer et al, *Induced Fit, Folding, and Recognition of the NF-κB -Nuclear Localization Signals by IkBα and IkBβ.* J.Mol. Biol., 2007. 367: p.162-274.

[21]   H. Shin, M. Kim, B. Kim, S. Jung, Y. Kim, H. Park, J. Hong, K. Min, Y. Kim, *Inhibitory action of novel aromatic diamine compound on lipopolysaccharide-induced nuclear translocation of NF-κB without affecting IκB degradation.* FEBS Letters, 2004. 571: p. 50-54.

[22]   Y. Dai, X. Pei, M. Rahmani, D. H. Conrad, P. Dent and S. Grant, *Interruption of the NF-κB pathway by Bay 11-7082 promotes UCN-01–mediated mitochondrial dysfunction and apoptosis in human multiple myeloma cells.* Blood, 2004. V: 103, 7: p.2761-2770.

[23]   Ikezoe, T. *et al.,* HIV-1protease inhibitor, ritonavir: a potent inhibitor of CYP3A4, enhanced the anticancer effects of docetaxel in androgen-independent prostate cancer *cells in vitro* and *in vivo*. Cancer Res., 2004. 64: p. 7426–7431.

[24]   Y. Li, et al., Apoptosis-inducing effect of chemotherapeutic agents is potentiated by soy isoflavone genistein, a natural inhibitor of NF-κB in BxPC-3 pancreatic cancer cell line. Pancreas, 2004. 28: p. 90–95.

[25]   V. Tergaonkar, V. Bottero, M. Ikawa, Q. Li, and I. M. Verma, *IκB kinase-independent IκB degradation pathway: functional NF-κB activity and implications for cancer therapy.* Mol. Cell. Biol., 2003. 23: p. 8070–8083.

[26]   V. Kolenko, T. Bloom, P. Rayman, R. Bukowski, E. Hsi and J. Finke, *Inhibition of NF-κB Activity in Human T Lymphocytes Induces Caspase-Dependent Apoptosis Without Detectable Activation of Caspase-1 and -3.* J. Immunol., 1999. 163: p. 590-598.

[27]   Y. Lin, S. Yao, R. A. Veachi T. R. Torgerson and J. Hawiger, *Inhibition of Nuclear Translocation of Transcription Factor NF-κB by Synthetic Peptide Containing a Cell Membrane-Permeable Motif and Nuclear Localization Sequence*, J. of Biol. Chem., 1995. V: 270, 24: p.14255-14258.

[28]   http://www.elmhurst.edu/~chm/vchembook/655cancer.html, last visited 15.09.2008.

[29] A. Salminen, M. Lehtonen, T. Suuronen, K. Kaarniranta, J. Huuskonen., *Terponoids: natural inhibitor of NF-κB signaling with anti-inflammatory and anticancer potential*, Cell Mol. Life Sci., 2008. V: 65, 19: p. 2979-99.

[30] R. Sen, and D. Baltimore, *Multiple nuclear factors interact with the ,immunoglobulin enhancer sequences*. Cell, 1986. 46: p. 705-*716*.

[31] *S. A. Snyder, A. L. Zografos and Y. Lin, Total Synthesis of resveratrol-based natural products: a chemoselective solution*. Angew. Chem. Int. Ed., 2007. 46: p. 8186-8191.

[32] E. Bedir, I. Calis, O. Zerbe and O. Sticher, *Cyclocephaloside I: A novel Cycloartane-type Glycoside from Astragalus microcephalus. J. Nat. Prod., 1998. 61: p. 503-505.*

[33] *A. D. Kinghorn et al., Natural Inhibitors of Carcinogenesis*. Planta Med, 2004. 70: 691-705.

[34] http://fendrri.blogspot.com/2007/05/art-of-frhttpwwwbloggercomimgglalignful.html, last visited 05.01.2009.

[35] T. Schulz-Gasch, M. Stahl, *Scoring functions for protein-ligand interactions: a critical prespective*. DDTEC, 2004. V: 1, 3: p. 231-239.

[36] http://www.unil.ch/webdav/site/dpt/shared/BachelorMaster/8semestre/LS_Drug_Design2.pdf, last visited at 28.09.08

[37] M. K. Anamala, K. K. Inampudi and L. Gutuprasad*, Docking of phosphate and trealose analog inhibitors into M. Tuberculosis mycolyltransferase Ag85C: Comparison of the two scoring fitness functions GoldScore and ChemScore, in the GOLD software*. Bioinformation, 2007. V:1, 9: p. 339-350.

[38] http://autodock.scripps.edu/ , last visited 05.01.2009.

[39] M.D. Eldride, C.W. Murray, T.R. Auton G.V. Paolini, R.P.Mee, *Emprical scoring functions. I: The development of fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes*, J.Computer-Aided Mol. Des., 1997. 11: p.425-445.

[40]   G. Jones, P. Willet, R. C. Clen, A. R. Leach and R. Taylor, *Development and validation of a genetic algorthim for flexible docking*. J. Mol. Biol., 1997. 267: p. 727-748.

[41]   D. E. Goldberg and J. H. Holand, *Genetic algorithms and machine learning*. Machine Learning, Kluwer Acedemic Publishers,1988. 3: 95-99.

[42]   J. Fang and D. S. Beattie, *Rotanone-insensitive NADH dehydrogenase is a potential source of superoxide in procyclic* T. Mol. Biochem. Parasitol, 2002. 123: p. 135-142.

[43]   A.K. Patick, H. Mo, M. Markowitz, et al., *Antiviral and resistance studies of AG1343, an orally bioavailable inhibitor of human* immunodeficiency virus protease. Antimicrob Agents Chemother, 1996. 40: p.292–297.

[44]   M. Markowitz, H. Mo, D.J. Kempf, D.W. Noebeck, T.N. Bhat, J.W. Erickson and D.D.Ho, *Selection and analysis of human immunodeficiency virus type 1 variants with increased resistance to ABT-538, a novel protease inhibitor*. J. Virol, 1995. 69: p. 701–706.

[45]   H. Schindler, Schutz G.J., M. Sonnleiter and P. Hinterdorfer, *Single molecule microscopy of biomembranes (review)*. Mol. Membr. Biol., 2000. V:17, 1: p.17-29.

[46]   http://www.angelo.edu/faculty/kboudrea/molecule_gallery/02_alkenes/00_alkenes.htmlast visited 27.10.08.

[47]   http://www.withanolide.com/ , last visited 05.01.09.

[48]   ChemBioDraw Ultra 11.0; Cambridge Soft licenced software, http://www.cambridgesoft.com/software/details/?ds=14&dsv=65 last visited 05.01.09.

[49]   WebLab Viewer Pro, commercial software by Molecular Simulations, Incorporated (MSI).

[50]   http://www.ccdc.cam.ac.uk/products/life_sciences/hermes/ , 05.01.09.

[51]   J. Gasteiger and M. Marsili, *A new model for calculating atomic charges in molecules*. Tetrahedron Letters, 1978. 34: 3181-3184.

[52]   T. Kutchan, *Taxol (paclitaxel) an effective antitumor compound a steroid alkaloid*. The Plant Cell, 1995. 7: p. 1059-1070.

[53]   http://www.acdlabs.com/products/phys_chem_lab/logd/ , last visited 05.01.09.

[54]   R.A. Scherrer, S.M. Howard, "Use of distribution coefficients in quantitative structure-activity relationships". *J Med Chem.*, 1977. V: 20, 1: p. 53–58.

[55]   J. C. Dearden, Partitioning and lipophilicity in Quantitative structure-Activity Relationships. Enviromental Health Perspectives, 1985. V: 61, p. 203-228.

[56]   T. I. Operea, A. M. Davis, S. J. Teague and P. D. Leeson, *Is there a difference between leads and drugs? A historical perspective*. J. Chem. Inf. Comput. Sci., 2001. 41: p. 1308-1315.

[57]   D. Engelmeier,
        http://www.phytochemie.botanik.univie.ac.at/staff/Amsterdam.pdf, last visited 05.01.09.

[58]   http://www.phytochemicals.info/phytochemicals/coumarin.php , 28.10.08

[59]   L. V. Alves, R. M. Temporal, L. Cysne-Finkelstein and L. L. Leon, Efficacy of a diarylheptanoid derivative against *Leishmania amazonensis. Mem. Inst.* Oswaldo Cruz, Rio de Janeiro, 2003. V: 98, 4: p. 553-555.

[60]   http://www.metabolome.jp/software/FlavonoidViewer/, last visited 27.10.08

[61]   http://breast-cancer-research.com/content/6/3/119/figure/F1,   last   visited 27.10.08.

[62]   A. D. Kinghorn et al., Induction of the Phase II enzyme, Quinone Reductase, by Withanolides and norwithanolides from Solanaceous Species. Mini-Reviews in Org. Chem., 2004. 1: p. 115-123.

[63]   http://www.mobot.org/MOBOT/research/APweb/top/glossaryi_p.html,
        last visited 05.01.09.

[64]   http://www.friedli.com/herbs/phytochem/flavonoids.html, last visited 27.10.08.

[65]   O. Kutuk, A.Pedreck, P. Harrison and H. Basagi, Pamanicin induces apoptosis in Jurkat Leukemia cells: Arole of JNK, p38 and caspase activation. Apoptosis, 2005. V: 10, 3: p.597-609.

VITA

Bahar Öndül was born in 1983, Ankara, Turkey. She had been to Karşıyaka Anadolu Lisesi for high school education. She received her BSc. degree in 2006, in Bioengineering at Ege University. During the undergraduate studies she worked in Biochemistry Centrum Hiedelberg for internship and was an exchange student in Lund University. She worked as an teaching and reseach assistant at Koç University during 2006–2008 with TUBITAK schoolarship. She pursued MSc. degree in Computational Science and Engineering in Koç University. Contact; e-mail: ondulb@yahoo.com.