In Silico Drug Design Studies of Bioactive Cannabinoid and [60]Fullerene Derivatives

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)
submitted to the Department of Biology, Chemistry and Pharmacy of
Freie Universität Berlin

by

SERDAR DURDAGI
from Tuzluca, Turkey
January, 2009

1st Reviewer: Prof. Hartmut Oschkinat

2nd Reviewer: Prof. Thomas Mavromoustakos

Date of defence: 04 May 2009

Abstract

The relationship between conformations of bioactive molecules with their pharmacological profiles has been well established. Only the unique biologically active conformation of a drug molecule can bind to the active site of the receptor. In this thesis, conformational analysis of bioactive compounds at various environments will be discussed. Two categories of molecules were investigated; cannabinoid (CB) analogues and [60]fullerene derivatives. The major structural characteristics of these molecules are: (i) amphiphilicity and (ii) existence of flexible and rigid pharmacophoric segments. Their flexible segments constitute a challenging field for conformational analysis exploring of putative bioactive conformations.

In case of CBs, a set of novel Δ^8-tetrahydrocannabinol (Δ^8-THC) and cannabidiol (CBD) analogues were subjected to three-dimensional quantitative structure-activity relationships (3D-QSAR) studies using comparative molecular field analysis (CoMFA), and comparative molecular similarity indices analysis (CoMSIA) methodologies. The high active compound C-1'-dithiolane Δ^8-THC analogue AMG3 at the data base was selected as template molecule. Using molecular modeling techniques such as Monte Carlo (MC), molecular dynamics (MD) and grid scan analysis, the putative bioactive conformation of AMG3 *in solution* was determined. This conformer was used as a template, and CB1 and CB2 pharmacophore models were developed. The availability of homology models of CB1 and CB2 receptors based on rhodopsin has allowed the conformational analysis studies of AMG3 at the binding site of the receptor. Derived low energy conformers of AMG3 at the receptor site have been compared with its *in solution* conformations. The steroelectronic properties of binding cavities of a receptor model are directly related to the performed molecular model coordinates. In the presented thesis, a homology modeling study based on β2-adrenergic receptor for both CB1 and CB2 receptors was also performed and results were compared with rhodopsin based homology models. Similar binding sites of CB1 and CB2 receptors using rhodopsin based models have been generated using the β2-adrenergic based receptors. The QSAR models were re-generated using putative bioactive conformers of AMG3 at the binding site of the CB1 and CB2 receptors. Relative contri-

butions of steric/electrostatic fields of the 3D QSAR/CoMFA and CoMSIA pharmacophore models have shown that steric effects govern the bioactivity of the compounds, but electrostatic interactions also play an important role. The comparison of derived QSAR models has shown that increasing the complexity level of calculations (mimicking more accurately the biological conditions) was positively affected the obtained statistical result. The optimal QSAR partial least square (PLS) analysis was used as an input in the *de novo* drug design studies and these simulations provided novel CB analogues with enhanced predicted binding affinities.

In case of fullerene derivatives, a series of experimentally reported as well as computationally designed monoadducts and bisadducts of [60]fullerene analogues have been used in order to analyze the binding interactions between fullerene based inhibitors and human immunodeficiency virus type I aspartic protease (HIV-1 PR), employing docking studies. MD simulations of ligand-free and the inhibitor-bound HIV-1 PR systems complemented the above studies and provided proper input structure of HIV-1 PR in docking simulations. The obtained results revealed a different orientation of the β-hairpin flaps at these two systems. In inhibitor bound system, the flaps of the enzyme are pulled in toward the bottom of the active site (the closed form) while, in ligand-free system flaps shifted away from the dual Asp25 catalytic site and this system adopts a semi-open form. The structural analysis of these systems at catalytic and flexible flap regions of the HIV-1 PR through the simulations, assisted in understanding the structural preferences of these regions, as well as, the adopted orientations of fullerene derivatives within the active site of the enzyme. The reported most active fullerene analogue in the data base has been used as template and 3D QSAR models were derived. Based on obtained contour plots and derived PLS analysis, *de novo* drug design studies were performed in order to propose novel analogues with enhanced binding affinities. Such structures may trigger the interest of medicinal chemists for synthesizing novel HIV-1 PR inhibitors possessing higher bioactivity, considering the urgent need for new anti-HIV drugs.
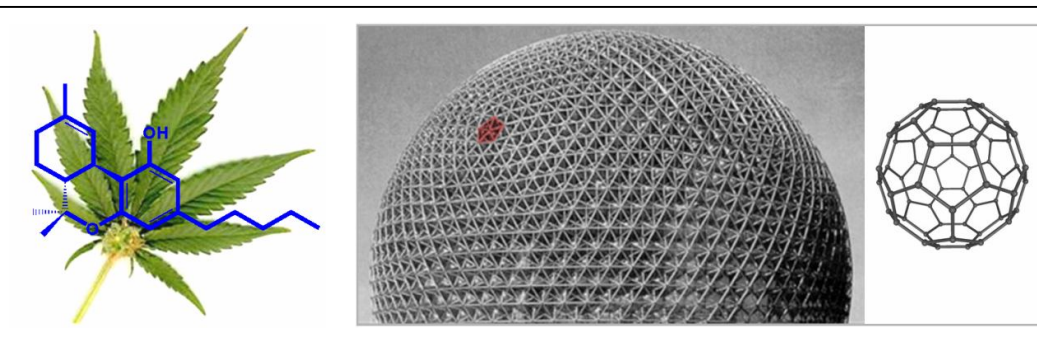
Contents

Chapter 1. Introduction

1.1 Motivation

Drug design is an iterative process which begins with identification of a compound that displays an interesting biological profile and proceeds until the activity profile is characterized by minimum undesirable side effects and chemical synthesis is optimized.[1] One of the main issues faced currently by the pharmaceutical industry is finding appropriate ligands for a given target protein and ensuring that they are highly specific for that target. Finding a therapeutic compound that binds selectively to a target receptor is not an easy task in the laboratory. Thus, this problem can be handled by using a combination of computer simulations together with laboratory work. Thanks to the wide spread availability of high performance computers and new computational techniques, receptors and their binding interactions with ligands can be defined at a molecular level with high accuracy.

The physicochemical and biological properties of a molecule depend on the conformations that it can adopt. Conformational analysis is the study of the conformations of a molecule and their influence to its chemical, physical or biological properties.[2] The modern conformational analysis studies were initiated by D. H. R. Barton[3], who showed that the reactivity of substituted cyclohexanes was influenced by the equatorial or axial nature of the substituents. The development of conformational analysis was enhanced due to the advancement of analytical techniques (e.g., infrared (IR) spectroscopy, nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography) which allow to explore the conformational properties of a molecule.

One of the frequently problems faced in drug design is to find the conformation of a molecule that adopts when it fits its target binding site (bioactive conformation). The relationship between the conformations of bioactive molecules with their pharmacological profiles has been well established. Only a unique conformation of a drug molecule can bind to the binding site of the receptor. The knowledge of the conformations of a ligand at the binding site of a receptor assists in the rational approach to drug design. One might suggest that the most stable conformation *in solution* is likely to be the bioactive conformation since the molecule is most likely to be in that conforma-

tion. However, the bioactive conformation of a ligand at the active site of the receptor is not necessarily identical with the lowest energy conformation, *in solution*. This is attributed to the fact that favourable binding interactions of the molecule with its target can lead a complex stabilization when the molecule adopts conformation that appears *in solution* with higher energies.[4] Nonetheless, a very high energy conformation that is excluded from the population of conformations *in solution* cannot be biologically active. Moreover, the investigation of the optimal conformation of a bioactive compound also plays an important role in three dimensional quantitative structure-activity relationships (3D QSAR) studies, because the output from the constructed 3D QSAR models is directly related with the alignment of the molecules in data set, based on template conformer.

The aim of applying 3D QSAR studies is to derive indirect binding information from the correlation between the biological activity of a training set of molecules and their 3D structures. The importance of steric and electrostatic characteristics is revealed by aligning structurally similar analogues using pharmacophoric features as structural superimposition guides. In contrast to the rigid molecules, defining the bioactive conformation of flexible ligands is more complex and therefore the alignment procedure of 3D QSAR is one of the most difficult steps. The combination of molecular modeling techniques and 2D NMR spectroscopy can be used to obtain the low energy conformation of a potent ligand in the data set which will serve as a template compound in the construction of QSAR models. In order to determine the linear correlation coefficients between actual versus calculated binding affinities, statistical analysis of the data can be used. The derived 3D QSAR models help to predict binding affinity values of ligands prior to their synthesis. The structure and binding affinity relationships of compounds can be graphically plotted and used to explain different steroelectronic requirements of ligands for binding to the receptor site. Contour results can be used as pilot models for proposing novel analogues before their synthesis.

The easiest way to analyze a bioactive conformation is to study the X-ray crystal structure of ligand-bound target structure. The structure of the protein/ligand complex can then be analyzed using computational methodologies and conformation of the

ligand can be identified. However, not all proteins can be easily crystallized (e.g., membrane proteins), thus other methodologies have to be used in order to identify the active conformation. If the X-ray structure of target is unknown, a homology model target can be created by molecular modeling based on known X-ray structure of a protein from same family, thus binding sites may be constructed to aid the drug design process. If one of the active compounds for a specific target is a rigid molecule which has one possible conformation, it can be used as template for more flexible molecules. The geometry of the pharmacophore (atoms and functional groups required for a specific pharmacological activity, and their relative positions in space) can be determined for the rigid molecule, and flexible molecules can be compared with this rigid molecule in order to find conformation which will place the important binding groups in the same relative geometry.[4] If a fully rigid molecule is not available to act as a template, it may be possible to match up different ligands that have a rigid part somewhere in their skeleton. Energy minimization methods play a crucial role in the conformational analysis.[4] If possible, it is desirable to identify all low energy conformations on the energy surface, but the number of minima may be so large that is not practical to pick up the conformer that has lowest energy. Under such circumstances population analysis can be performed using statistical mechanics. Solvation effects may also be important to include in the calculations of intramolecular energy. Several computational techniques (e.g., molecular dynamics (MD) simulations, Monte Carlo (MC) calculations) can be used as part of a conformational search strategy. More detailed information on conformational search strategy will be given in the "Theoretical Background" chapter of the thesis.

When no experimental structural information is available, molecular modeling techniques can be used to reveal the bioactive conformations of the ligand. These techniques include: (i) geometry optimization calculations, in gas phase; (ii) geometry optimization calculations employing a continuum model which simulates the biological medium; (iii) rotational energy barrier calculations; (iv) MC simulations; (v) MD simulations; and (vi) 3D QSAR models. Additional molecular modeling techniques can be used when the X-ray structure or homology models of a receptor are available: More specifically; *in silico* docking; and MD simulations of ligand at the active site of

the receptor can be performed. If the X-ray structure or a homology model of the target protein is available, the investigation of a bioactive conformation of molecule becomes more realistic and can be obtained with greater credibility.

1.2 Investigated Systems

In the present work, conformational analysis of bioactive compounds at various environments will be discussed. Two categories of molecules were studied: (i) cannabinoid (CB) analogues and (ii) [60]fullerene derivatives. The major structural characteristics of these molecules are: (i) amphiphilicity and (ii) existence of flexible and rigid pharmacophoric segments. Their flexible segments constitute a challenging field for conformational analysis exploring of putative bioactive conformations.

1.2.1 CBs

CB agonists have been suggested to have potential therapeutic uses such as neuroprotective, analgesics, apetite stimulants, anti-emetics, anti-glaucoma agents, and for the treatment of diseases associated with inappropriate retention of aversive memories (e.g., post-traumatic stress disorders and phobias).[5] The pharmacological activity of CBs is mediated by G protein coupled receptors (GPCRs) CB1 and CB2. On binding of agonists, GPCRs become activated, presumably by conformational changes in the transmembrane (TM) domain. The CB1 receptor is localized primarily in the central nervous system (CNS), reflecting its most abundant GPCR prevalence in brain. An interesting feature of CB1 is its ability to be activated by structurally different classes of molecules, thus increasing the possibility of multiple activated forms of the receptor. Although detectable at exceedingly low levels in brain, CB2 receptors are expressed mainly by immune cells and mediate immune responses, inflammatory and neuropatic pain.[6]

Presently known CB analogues show susceptibility towards enzymatic hydrolysis and/or not have CB1/CB2 receptor selectivity. There is considerable interest to design CB analogues possessing selectivity, high affinity and metabolic stability. Such ana-

logues may supply favorable response with fewer undesirable side effects and higher metabolic stability.[7]

Identification of the binding conformations of CB agonists within the binding site of the receptor is of great interest not only to understand the key binding interactions between the amino acid residues and the ligand, but also provide insight into the molecular mechanism of receptor activation.[8]

A set of novel synthesized Δ^8-tetrahydrocannabinol (Δ^8-THC) and cannabidiol (CBD) analogues were subjected to 3D QSAR studies using comparative molecular field analysis (CoMFA)[9], and comparative molecular similarity indices analysis (CoMSIA)[10] methodologies in order to propose new selective and high affinity CB analogues. The high active compound C-1'-dithiolane Δ^8-THC analogue (–)-2-(6a,7,10,10a-tetrahydro-6,6,9-trimethylhydroxy-6H-dibenzo[b,d]pyranyl)-2-hexyl-1,3dithiolane (AMG3), (Figure 1.1i) was selected as the template molecule. Using combination of several molecular modeling techniques such as molecular mechanics (MM) and quantum mechanics (QM) geometry optimization calculations, MD simulations, MC calculations and grid scan analysis, the putative bioactive conformation of AMG3 *in solution* was determined. This conformer was used as the template structure and CB1 and CB2 pharmacophore models were developed.

The availability of homology models of CB1 and CB2 receptors based on bovine rhodopsin, allowed the conformational analysis studies of AMG3 at the binding sites of CB receptors. Firstly, *in silico* docking simulations were performed using template ligand at the CB1 and CB2 receptors and cluster analysis was accompanied to the obtained binding poses. Secondly, the derived best binding pose of protein/ligand complex structure has been used as an input coordinate, and MD simulations have been performed in the presence of membrane bilayers. Trajectory analysis from MD simulations of generated snapshots clarified the favored adopted low energy conformers of template compound. The derived low energy conformers of AMG3 at the binding site of the receptor have been compared with those produced *in solution*. QSAR models were re-generated using putative bioactive conformers of AMG3 at the binding site of the CB1 and CB2 receptors. Generated QSAR models *in solution* and at the binding

site of the CB1 and CB2 receptors were compared to define environment that more closely resembles the bioactive conformer.

The derived optimal QSAR partial least square (PLS) analysis of CB models were used in the *de novo* drug discovery program for the predictions of novel compounds with enhanced predicted binding affinities.

Since the steroelectronic properties of binding cavities of a receptor model are directly related to the performed molecular model coordinates, a homology modeling study based on β2-adrenergic receptor for both CB1 and CB2 receptors was also performed and results were compared with former rhodopsin based homology models.

1.2.2 [60]Fullerene Derivatives

In the last few years, many interesting biological applications of [60]fullerene derivatives have started to be investigated due to their promising biological activities such as DNA photocleavage, human immunodeficiency virus type I aspartic protease (HIV-1 PR) inhibition, apoptosis and neuroprotection.[11,12] Indeed, unique spherical shape of [60]fullerene may be envisaged as fitting the hydrophobic cleft that is often characterizing the target structures. When the cleft hosts the guest molecule and the intermolecular steroelectronic interactions are sufficiently strong, inhibitory effect may be expected.

The inhibition of HIV-1 PR by fullerene analogues has been demonstrated by Friedman *et al.*[11,13] and the complexation of HIV-1 PR with fullerene compounds has been supported by molecular modeling studies. These studies showed that the fullerene derivatives can be perfectly accommodated inside the binding pocket of HIV-1 PR. However, the binding affinity (K_i) values of "first generation" fullerene inhibitors were not significant ($K_i \sim 10^{-6}$ M). Thus, further structural investigation is required in order to propose new HIV-1 PR/fullerene complexes with better binding affinity.

A series of experimentally reported as well as computationally designed monoadducts and bisadducts of [60]fullerene analogues have been used in order to analyze the binding interactions between fullerene based inhibitors and HIV-1 PR enzyme employing docking studies. MD simulations of ligand-free and the inhibitor-bound HIV-1 PR systems complemented the above studies and provided proper input structure of HIV-1 PR in docking simulations. The structural analysis of these systems at catalytic and flexible flap regions of the HIV-1 PR through the simulation, assisted in understanding the structural preferences of these regions, as well as, the adopted orientations of fullerene derivatives within the active site of the enzyme.

The reported most active [60]fullerene in the data base ((3'-Phenyl-3'-(α-hydroxybenzyl)-1,2-cyclopropano]buckminsterfullerene), (Figure 1.1ii)) was used as template and 3D QSAR/CoMFA and CoMSIA models were derived. Based on contour plots and PLS analysis from the models, *de novo* drug design studies were performed in order to propose novel analogues with enhanced binding affinities. Such structures may trigger the interest of medicinal chemists for novel HIV-1 PR inhibitors possessing higher bioactivity.
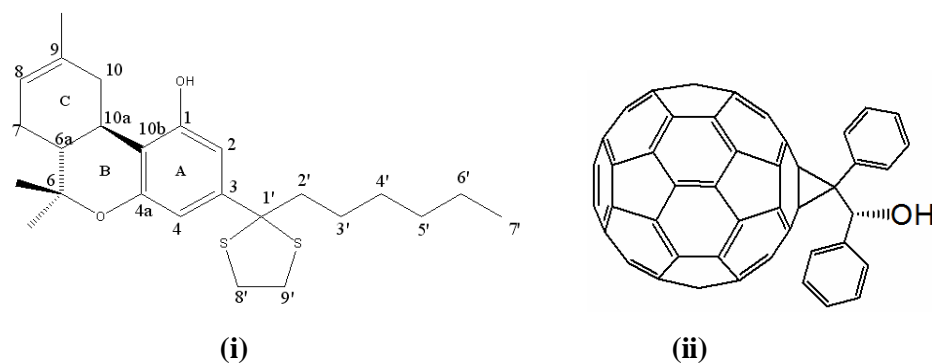
Figure 1.1 (i) Molecular structures of potent bioactive CB analogue AMG3; and **(ii)** an anti-HIV-1 PR bioactive fullerene analogue.

1.3 The Aim of the Study

The aim of this study is to shed some light on the following questions addressed in the CB and fullerene fields.

(i)   Can 3D QSAR studies help to design putative bioactive novel CB and fullerene analogues with higher binding affinities and metabolic stabilities?

(ii)   Can 3D QSAR studies help to design, in particular, selective novel CB derivatives for CB1 and CB2 receptors?

(iii)   Is there any connection between biological activity of fullerene derivatives and the flap motion of the HIV-1 PR enzyme?

(iv)   How conformational changes at the catalytic site of the HIV-1 PR affect the biological activity of the fullerene derivatives?

(v)   How conformations of a template ligand derived using different environments affect the 3D QSAR models?

1.4 Outline of the Thesis Structure

The thesis structure is organized as follows. Chapter 1 introduces the reader to the subject and aims of this Ph.D. dissertation. Chapter 2 provides the "theoretical background" of some key concepts necessary to understand the discussed obtained results. In the Chapter 3, details of computational techniques used in the calculations are discussed. Chapter 4 provides discussion for the used strategies in the rational drug design. In the Chapter 5, a thorough conformational analysis of potent CB ligand AMG3 using several molecular modeling techniques, its binding interactions with CB1 and CB2 receptor models, comparative 3D QSAR studies of CB analogues, homology modeling calculations of CB receptors and *de novo* drug design studies of CB ligands are discussed. In the Chapter 6, conformational analysis and binding interactions of [60]fullerene derivatives at the binding pocket of HIV-1 PR, 3D QSAR and *de novo* drug design studies are discussed. A summary of the main results is given in the last chapter (Chapter 7). Concluding remarks for the impact of this dissertation are also highlighted.

Chapter 2. Theoretical Background
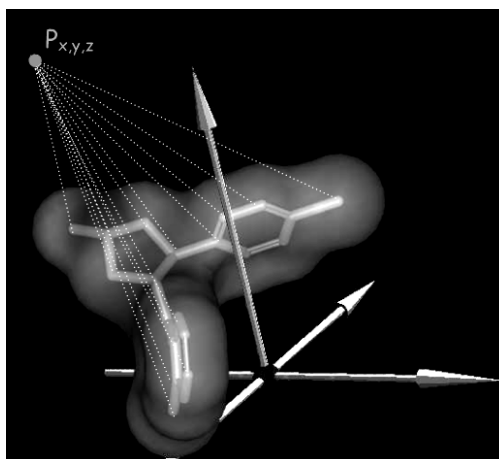
2.1 Some Useful Definitions for Many-body Systems

Approaches used in computational chemistry can be divided to two broad parts, empirical and quantum approaches. Empirical approaches (i.e., MM) use simple models of harmonic potential, electrostatic interaction, and dispersion forces for basic comparisons of energetics and geometry optimization. MM methods are extremely fast and are able to handle very large systems, such as enzymes.[2] Quantum approaches are roughly divided into semi-empirical methods and non-empirical (or *ab initio*) methods. Semi-empirical methods are the approximate methods in which parameters involved in the equations are taken from experiment, some are neglected, and some others are estimated by fitting to experimental data. Like MM methods, semi-empirical methods use experimentally derived parameters; and like *ab initio* methods, they are basically quantum mechanical (QM) in nature. The main difference between semi-empirical and *ab initio* methods is the extensive use of approximations and parameters optimized with respect to experimental data by the former approach. This allows semi-empirical methods to reduce the computational cost, while the computed results, in general, provide useful data.[14]

2.1.1 QM Calculations

The QM approach postulates the fundamental principles and then uses these postulates to deduce experimental results. For the definition of the state of a system in QM, the function of the coordinates of particles is referred as the wave function or state function Ψ. In general, the state changes with time, thus for one-particle, one-dimensional system, $\Psi=\Psi(x, t)$. The wave function contains all possible information about a system.[14]

Suppose there is a single particle (e.g., an electron of mass m) which is moving through space (given by a position vector $r = x\mathrm{i} + y\mathrm{j} + z\mathrm{k}$) under the influence of an external potential ϑ. In order to find the future state of a system from the knowledge of its first state, an equation is needed that tells how the wave function changes with

time (t).[15] Schrödinger's time dependent equation describes the particle by a wave function[15] $\Psi(r, t)$:

$$\{-\frac{\hbar^2}{2m}(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2})+\vartheta\}\Psi(r,t)=i\hbar\frac{\partial\Psi(r,t)}{\partial t}$$

(where, $\hbar$ = h/2π, h is the Planck's constant and i^2 = -1)

When the external potential ϑ is independent of time, then the wave function can be written as the product of a spatial part and a time part; $\Psi(r,t) = \psi(r)T(t)$. In many applications of QM, the potential is considered as independent of time, thus the time-dependent Schrödinger equation can be written in the more familiar, time-independent form[2,14,15]:

$$\{-\frac{\hbar^2}{2m}\nabla^2+\vartheta\}\Psi(r)=E\Psi(r)$$

where, E is the energy of the particle and $\nabla^2=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}$.

Usually, $-\frac{\hbar^2}{2m}\nabla^2+\vartheta$ is abbreviated as Hamiltonian operator H. Thus, Schrödinger equation is reduced to $H\Psi=E\Psi$. In order to solve the Schrödinger equation for many-body systems, Hartree-Fock (HF) and density functional theory (DFT) are the commonly used approaches in the QM calculations. In the HF approximation, instead of calculating repulsions between electrons in the system explicitly, repulsions are calculated between one electron and the average field of all of the other electrons. In the DFT approximation, the total electron density is decomposed into one-electron densities, which are constructed from one-electron wave functions.[16,17]

2.1.2 Semi-empirical Calculations

The greatest proportion of the computational time at the *ab initio* calculation is invariably spent calculating and manipulating integrals. In order to reduce the computational effort, the most obvious way is to neglect or approximate some of these integrals. Semi-empirical methods consider only valance electrons of the system and the

core electrons are subsumed into the nuclear core. The key point in semi-empirical methods is the overlap matrix S (in Roothaan-Hall equation FC=SCE), which is approximated by the identity matrix I.[18] Therefore, all diagonal elements of the overlap matrix are equal to one and all off-diagonal elements are zero. Thus, the Roothaan-Hall equation FC=SCE becomes FC=CE (F represents the Fock matrix, is a sum of one- and two-electron contributions, C is the molecular orbital coefficients and E is the energy levels).[18]

2.1.3 MM Calculations: Empirical Force Field Models

Generally, most of the systems that need to be studied using molecular modeling are too large for QM calculations. QM calculations include the electrons in a system, although some of them may be ignored (e.g., semi-empirical calculations), a large number of particles must still be considered, thus the calculations are time-consuming.[2] In the MM (also known as force field methods) not the electronic motions but only the function of nuclear positions of the system is included in the calculations of the energy of the system.[2] Thus, in contrast to *ab initio* methods, MM is used to compute molecular properties (e.g., geometrical properties, relative stability of conformers) which do not depend on electronic effects.[2]

Today, many of the MM force fields use relatively simple four components of the intra- and inter-molecular forces within the system. Energetic penalties are associated with the deviation of bond lengths, bond angles and torsion angles from their reference or equilibrium values. More sophisticated force fields may include additional terms, but they invariable contain these four components[2,16,17,19]:

$$\vartheta(\mathrm{r}^N) = \sum_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{V_n}{2}(1 + \cos(n\omega - \gamma))$$

$$+ \sum_{i=1}^{N}\sum_{j=i+1}^{N}(4\varepsilon_{ij}[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}})$$

where, $\vartheta(\mathrm{r}^N)$ denotes the potential energy, which is a function of the positions (r) of N particles (e.g., atoms). The first term represents the interaction between pairs of

bonded atoms. This is modeled by a harmonic potential that leads to an increase in energy as the bond length (*l*) deviates from the reference value. The second term is associated with bond angle deformations modeled using a harmonic potential. The third term represents the torsional potential and models the energy changes relatively to bond rotations. The fourth term is associated with the non-bonded atoms and is calculated between all pairs of atoms (*i* and *j*) that are in different molecules or in the same molecule but separated by at least three bonds. This term is usually modeled using a Lennard-Jones potential for van der Waals interactions and Coulomb potential for electrostatic interactions.

2.2 Continuum Representations of the Solvent

Most of the chemical processes take place in a solvent, thus it is important to consider how the solvent affects the behaviour of the system. In some cases, the solvent merely acts as a 'bulk medium', but it can still significantly affect the solute behaviour, with dielectric properties of the solvent often being particularly crucial.[2] In such cases, it would be useful not to have explicit solvent molecules in the system, to enable us to concentrate on the behaviour of the solute(s). The solvent acts as a perturbation on the gas-phase behaviour of the system. This is the purpose of the 'continuum' solvent models.[2]

2.2.1 Solvation Free Energy

The solvation free energy (ΔG_{sol}) is the free energy change to transfer a molecule from vacuum to solvent phase. The ΔG_{sol} can be considered to have three components[2]:

$$\Delta G_{sol} = \Delta G_{elec} + \Delta G_{vdw} + \Delta G_{cav}$$

where, ΔG_{elec} is the electrostatic component, ΔG_{vdw} is the van der Waals interaction between the solute and solvent. ΔG_{cav} is the free energy required to form the solute cavity within the solvent. This component comprises the entropic penalty associated with the reorganization of the solvent molecules around the solute together with the work done against the solvent pressure in creating the cavity.[2]

2.2.2 Electrostatic Contributions to the Free Energy of Solvation

Born derived an expression for the electrostatic contribution of the ΔG_{sol} by placing a charge within a spherical solvent cavity.[2] In the Born model[20], ΔG_{elec} of an ion equals to the work done to transfer the ion from vacuum to the medium. This in turn is equal to the difference in the electrostatic work to charge the ion in two environments. The work to charge an ion in medium of dielectric constant ε equals to $q^2/2\varepsilon a$, where q and a are the charge on the ion and the radius of the cavity, respectively. Thus, ΔG_{sol} is the difference in the work done in charging the ion in the dielectric and *in vacuo*[2]:

$$\Delta G_{elec} = -\frac{q^2}{2a}(1-\frac{1}{\varepsilon})$$

In many cases, solvent effects can be incorporated into a force field model. It is possible to study larger systems with the empirical models, in which case it is necessary to include dielectric properties of both solute and solvent. The generalized Born equation has been widely used to represent the ΔG_{elec} contribution to the ΔG_{sol}.[2] The model includes a system of particles with radii a_i and charge q_i. The total electrostatic free energy of this system is given by the sum of the Coulomb energy and the Born free energy of solvation in a medium of relative permittivity[2] ε:

$$G_{elec} = \sum_{i=1}^{N}\sum_{j=i+1}^{N}\frac{q_i q_j}{\varepsilon r_{ij}} - \frac{1}{2}(1-\frac{1}{\varepsilon})\sum_{i=1}^{N}\frac{q^2}{a_i}$$

2.3 Computer Simulations Methods

2.3.1 Energy Minimization Method

The aim of energy minimization is to find a set of coordinates representing a molecular conformation such that the potential energy of the system is at a minimum. This can be formally stated as follows: given a function f which depends on one or more independent variables $x_1, x_2, x_3,...x_i$, find the values of those variables where f has a minimum value.[2,21] At the minimum, the first derivative of the function with respect to each variable is zero and second derivatives are all positive ($\frac{\partial f}{\partial x_i}=0$; and $\frac{\partial^2 f}{\partial x^2_i}>0$).

Information provided by energy minimization calculations in some cases can be sufficient to predict accurately the properties of a system. If low energy conformations of a system on an energy surface can be identified then statistical mechanics techniques can be used to derive a partition function from which thermodynamic properties can be calculated. However, this is possible only for small molecular assemblies in the gas phase. Computer simulation methods assist to study large systems and predict their properties through the use of techniques that consider small replications of a macroscopic system with manageable numbers of atoms.[2] Simulations generate a time-dependent behavior of these small replications in such a way that accurate values of structural and thermodynamic properties can be obtained with a feasible computation time.

2.3.2 Molecular Dynamics (MD) Method

The essence of the MD method is in the numerical integration of Newton's second law relating the mass and acceleration of an atom in the system to the gradient of the potential energy function and its associated force field.[19] Thus, an approximate velocity for the atom can be computed with the given acceleration for a given period of time and changes in atomic coordinates can be determined. MD is a deterministic method, that is the state of the system at any future moment can be predicted from its current state.[2] For this, continuous nature of potentials requires the equation of motion to be integrated by breaking the calculations into a series of very short time steps (δt, usually in fs time level). Finite difference techniques are used to generate MD trajectories with continuous potential models. At each time step, the forces on the atoms are computed and combined with the current positions and velocities to generate new positions and velocities a short time ahead ($t + \delta t$).[2,19] There are many algorithms for integrating the equations of motion, using finite differences and they assume that the positions and dynamic properties of the atoms consisting the system under study can be calculated as Taylor series expansions[2]:

$$r(t+\delta t) = r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t) + \frac{1}{6}\delta t^3 b(t) + \frac{1}{24}\delta t^4 c(t) + ...$$

$$v(t+\delta t) = v(t) + \delta t a(t) + \frac{1}{2}\delta t^2 b(t) + \frac{1}{6}\delta t^3 c(t) + \frac{1}{24}\delta t^4 d(t) + ...$$

where, v is the velocity (the first derivative of the positions with respect to time), *a* is the acceleration (the second derivative), *b* is the third derivative, and so on. The Verlet algorithm is one of the most widely used algorithm for the integrating the equations of motion in MD simulations. It uses the positions and accelerations at time t, and the positions from the previous step r(*t-δt*), in order to calculate the new positions, r(*t+ δt*). The system is followed for user defined time, taking care that the temperature and pressure remain at the required values, and coordinates as a function of time are written to an output file at regular intervals. Thus, MD simulation generates trajectory files that describe how the dynamic variables change through simulation. Usually, simulation time in the MD calculations is between hundreds of pico-second (ps) and a few nano-second (ns) level. In the last years, MD simulations are extensively used to investigate the conformational properties of flexible molecules.

2.3.3 Monte Carlo (MC) Method

MC is another computer simulation method, while the low energy conformations of a system are connected to the time, in a MD simulation; in a MC simulation, each conformation depends only to the predecessor and not upon any other conformations previously visited.[19] The MC technique derives conformations randomly and uses a special set of criteria to decide whether or not to accept each new conformation. These criteria ensure that the derived conformation is equal to its Boltzman factor $\exp\{-\vartheta(\mathrm{r}^N)/k_B T\}$, where $\vartheta(r^N)$ is calculated using the potential energy function.[2] In MC method, each new conformation of the system is generated by randomly rotating the bonds. The energy of the new system is then calculated using the potential energy function, and if the energy of the new system is lower than the energy of its predecessor then the new conformation is accepted. If the energy of the new system is higher than its predecessor, then the Boltzmann factor of the energy difference is calculated with $\exp\{-(\vartheta_{new}(\mathrm{r}^N) - (\vartheta_{old}(\mathrm{r}^N))/k_B T\}$ and compared with a generated random number between 0 and 1.[2] If the Boltzmann factor is greater than the random number then the new conformation is accepted, if not, then initial conformation is retained for the next move.[2] Indeed, numbers generated from random number generator are not truly random because the same sequence of numbers should always be generated when the

program is run with the same initial conditions. Usually, the linear congruential method is used for generating random numbers; each number in the sequences generated by taking the previous number, multiplying by a constant (multiplier, a) , adding a second constant (increment, b), and taking the remainder when dividing by the third constant (modulus, m); ($\xi[i] = MOD\{(\xi - 1) \text{ x } a + b), m\}$).[2] If the constants are chosen carefully, the method generates all possible numbers between 0 and m-1. The linear congruential method generates integral values, which can be converted to real numbers between 0 and 1 by dividing by m.[2]

There are some difficulties applying the MC simulations to flexible molecules. It has been found that, even small movements away from an equilibrium bond length, cause a large increase in the energy. One of the widely used methods to overcome this problem is to freeze out some of the internal degrees of freedom, such as bond lengths and bond angles.[2,16,19]

2.4 Methods for Exploring Conformational Space

2.4.1 Grid Search

In the grid search, all of the rotatable bonds are identified in the system and these bonds are then systematically rotated through 360° using a fixed increment (during the search bond lengths and bond angles remain fixed).[2] Then, every derived conformation is subjected to energy minimization in order to obtain low energy conformations. For example, in the alanine dipeptide (Figure 2.1) conformational energy surface can be obtained by grid search algorithm. If it is assumed that the amide bonds adopt *trans* conformations then only two torsional angles ψ and φ of the backbone are flexible. Then, the energy is the function of these two variables and can be represented as contour maps.[2] These contour plots in amino acids is also known as Ramachandran map, after G. N. Ramachandran who found that the amino acids were restricted to a limited range of conformations. A representative Ramachandran map (for HIV-1 PR enzyme which has hundreds of ψ and φ torsion angles) is shown in Figure 2.2, where points lie on the axes indicating N- and C- terminal residues for each subunit. The most al-

lowed region of Ramachandran space is colored blue, and partially allowed regions are colored green.
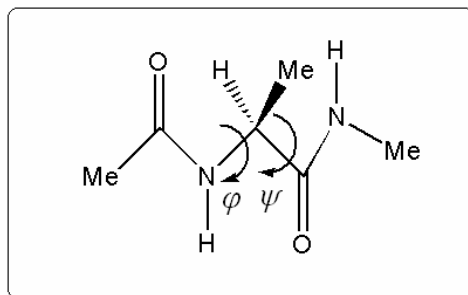
Figure 2.1 The structure of alanine dipeptide.
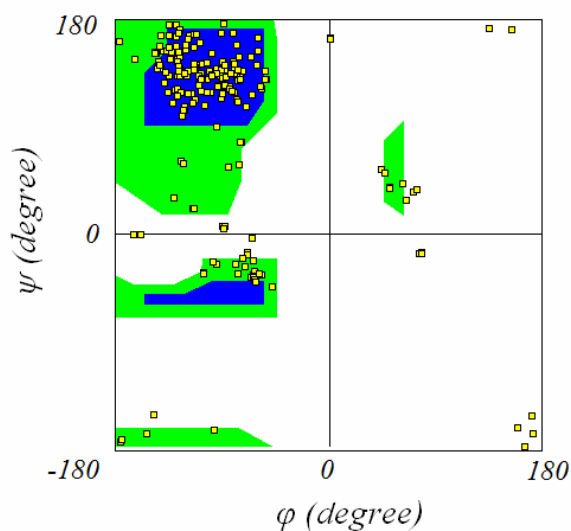
Figure 2.2 Ramachandran plot of experimentally derived structure of HIV-1 PR enzyme (pdb, 1AID).

2.4.2 Random Search Methods

In the random search method, conformational space can be explored by changing either the atomic Cartesian coordinates or the torsion angles of rotatable bonds. At each iteration, a random change is performing to the current conformation and the new structure is then refined using energy minimization. If this conformation has not been obtained previously, it is stored. A new starting conformation is then chosen for the next iteration and the cycle starts again. The procedure continues until a given number

of iterations have been performed. The systematic search ends when all possible combinations of bond rotations are covered, however in the random search, there is no natural endpoint.[2]

2.4.3 Distance Geometry in NMR

2D-NOESY (Nuclear Overhauser effect spectroscopy), 2D-ROESY (rotating-frame Overhauser effect spectroscopy) and 2D-COSY (correlated spectroscopy) are mainly used experimental methods in the conformational analysis problems.[2] NOESY gives the information about the distance between atoms which are close together in space but may be separated by many chemical bonds. The strength of the NOESY signal is inversely proportional to the sixth power of the distance, thus it is possible to calculate approximate values for the distance between the pairs of atoms. ROESY is an experiment in which homonuclear NOE effects are measured under spin-locked conditions and usually used when the NOE is small for any internuclear distance or mixing time. COSY experiments are used to provide information for the atoms which are covalently separated by one to three bonds.[2] Usually, 2D-NMR spectroscopy and molecular modeling calculations are combined for the elucidation of low energy conformers of a molecule.

2.4.4 Exploring Conformational Space by Simulation Methods

MC and MD simulations can be performed to explore the conformational space of molecules. There are some difficulties in using MC method for flexible molecules, as it has been discussed in section 2.3.3. A common strategy for a conformational space problem, solved by MD simulations is to perform the simulations at a very high temperature (in some cases physically unrealistic temperatures). A high-temperature MD simulation may be able to overcome high energy barriers, thus explore conformational space.

2.4.5 Clustering Algorithms

Many of the conformational analysis algorithms can generate conformations that are very similar. In this case, it is desirable to select a representative small set of conformations with a subsequent analysis (e.g., cluster analysis). Cluster analysis groups similar objects, from which the representatives can be picked up. Cluster analysis measures the similarity between the pairs of objects. In comparison of conformations of a molecule, the root mean square deviation (RMSD) would be an obvious measure to use[2]:

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^{N_{atoms}} d_i^2}{N_{atoms}}}$$

where, N_{atoms} is the number of atoms and d_i is the distance between the coordinates of atom i in the two structures. Another method is to measure similarities of torsional angles between the structures[2]:

$$d_{ij} = \sum_{m=1}^{N_{tot}} (w_{m,i} - w_{m,j})$$

where, $w_{m,i}$ is the value of torsional angle m in conformation i. N_{tot} is the total number of torsional angles.

2.5 The Use of Molecular Modeling and Cheminformatics Methods in the Design of New Molecules

2.5.1 Molecular Docking

Molecular docking or (*in silico* docking) is a method which predicts the preferred conformations of one molecule to a second one when bound to each other to form a stable complex. Molecular docking can be considered as a dynamic procedure where the correct geometry of a "*key*" is sought which will open the "*lock*".

The docking problem involves several degrees of freedom, for example there are six translational and rotational degrees of freedom of a molecule relative to other as well as the conformational degrees of freedom of each molecule.[2,14,16] Many algorithms have been developed to solve the docking problem. These can be grouped according to the number of degrees of freedom that they ignore. The simplest algorithms use two molecules as rigid bodies, thus explore only six degrees of translational and rotational freedom (rigid docking). In flexible docking, the conformational degrees of freedom need to be taken into consideration. Most of the current docking programs treat only ligand as flexible and receptor is considered rigid (flexible docking). Conformational search methods have been incorporated at some stage in the docking calculations. For example, MC simulation is used for the changing of internal conformation of ligand. The binding energy of ligand at the binding site of the receptor is calculated by MM methods. The ideal docking method would allow flexibility to both ligand and receptor to explore their conformational degrees of freedom. Thus, the most 'natural' way to incorporate flexibility to ligand and receptor is via MD simulations to the ligand/receptor complex. However, these calculations are computationally very demanding and, in practice, MD is used for refinement purposes of docking modes. Thus a successful initial docking pose has to be obtained before MD is applied.

2.5.2 Scoring Functions for Molecular Docking

Most of the docking programs are able to generate a large number of docking poses. Some of them have very high energy and clash with protein, therefore are neglected. The rest of the solutions are ranked using some scoring functions. Usually, scoring functions attempt to approximate the ligand-receptor binding free energy. The free energy of binding can be calculated with several simulation techniques but these calculations are very time-consuming and too slow to be of value in docking calculations. More approximate and faster methodologies have to be used. In contrast to the free energy perturbation methods, these algorithms compute the free energy of binding as an additive equation of various components[2,22]:

$$\Delta G_{bind} = \Delta G_{solvent} + \Delta G_{conf} + \Delta G_{int} + \Delta G_{rot} + \Delta G_{t/r} + \Delta G_{vib}$$

where, $\Delta G_{solvent}$ is the contribution of the solvent effects arising from the interactions between the solvent and the ligand, protein and intermolecular complex. ΔG_{conf} is contribution of conformational changes in the protein and in the ligand (usually, docking algorithms assume a rigid receptor, thus ΔG_{conf} depends only on the conformational changes of ligand). ΔG_{int} is the free energy, due to ligand-protein interactions. ΔG_{rot} (entropic contribution) is the free energy loss because of freezing internal rotations of protein and the ligand. This penalty function is simply calculated by assuming three states per rotatable bonds (*trans* and *gauche*±) of equal energy, thus leading to a free energy loss of RT ln3 (~0.7 kcal/mol) per rotatable bond. $\Delta G_{t/r}$ is the penalty function for the loss of translational and rotational free energy caused by the association of two-bodies (ligand and receptor) to give a single body (the intermolecular complex). ΔG_{vib} is the free energy of the system due to changes in vibrational modes (calculations of this term is very difficult and do not affect significantly the total sum, thus, they are usually ignored). There is a considerable discussion for each of the terms in above equation in literature, and for some of them, there may be a number of different ways for their estimation. However, many of these approaches are not suitable for the docking calculations, because they require a high computational cost.[2] Despite the simplicity of scoring functions, such functions rate well in comparisons of different functional forms. One of the interesting scoring function methods was suggested by Böhm. In this approach, a simple relationship between the free energy of binding and a variety of parameters as well as fast calculation has been considered[2,23]:

$$\Delta G_{bind} = \Delta G_0 + \Delta G_{H-bonds} \sum_{H-bonds} f(\Delta r, \Delta \alpha) + \Delta G_{ionic} \sum_{ionic} f(\Delta r, \Delta \alpha) + \Delta G_{lipo} \mid A_{lipo} \mid + \Delta G_{rot} N_{rot}$$

where, ΔG_0 is a constant term and it is independent from the system. It was interpreted as corresponds to $\Delta G_{t/r}$ term, in the previous equation which shows the overall change in translational and rotational free energy. $\Delta G_{H-bonds}$ is the contribution from an ideal hydrogen bond and multiplied by a penalty function $f(\Delta r, \Delta \alpha)$, which accounts for large deviations of a hydrogen bond from ideal geometry (assumed as 1.9 Å and 180°). ΔG_{ionic} is the contribution from an unperturbed ionic interactions. ΔG_{lipo} is the

contribution from lipophilic interactions, which are assumed to be proportional to the lipophilic contact surface between ligand and receptor A_{lipo}. ΔG_{rot} is the loss of free energy due to freezing a rotatable bond in the ligand upon binding, and it is multiplied by a number of rotatable bonds in the ligand (N_{rot}).

2.5.3 3D QSAR Studies

2.5.3.1 Structure-Activity Relationships (SAR)

Molecular properties are coded by molecular structure. Compounds with similar structures often tend to have similar pharmacological activity. However, they usually exhibit differences in potency, undesirable side effects and in some cases different binding affinities. These structurally related differences are commonly referred to as SAR.

SAR can be defined as "the relationship between chemical structure and pharmacological activity for a series of compounds".[4,24] Traditional SAR studies are usually carried out by making large numbers of analogues of the lead compound and testing them for biological activity at a specific target. A SAR study of a lead compound and its analogues can be used to determine the segments of the lead compound that are crucial for its binding affinity, that is, its pharmacophore as well as its unwanted side effects. This information can be used subsequently to design a new drug that has increased affinity and fewer unwanted side effects than the existing drug (optimize its SAR). Traditional SAR investigations are useful tools in the search for new drugs, however they are expensive in both personnel and materials. Thus, a number of attempts have been made to improve the traditional SAR studies.

2.5.3.2 QSAR

The success of the SAR approach to drug design depends not only on the knowledge and experience of the drug design group but also may related with luck. QSAR is an attempt to remove the luck factor from drug design by establishing a relationship in the form of a mathematical equation between biological activity and the measurable

physicochemical parameters of a drug that represents its properties such as lipophilicity, shape and electron distribution, which have major effects on the activity.[4,24] Quantitative structure-property relationships (QSPR) are also used, particularly when a specific property other than biological activity is considered. If an equation is formed, then a medicinal chemist could determine from the equation the value of parameter, and hence the structures, that would optimize the activity. These predictions allow medicinal chemists to make a more informed choice as to what analogues to design and synthesize. Obviously, this could considerably cut down the cost of drug development. The relationship between these numerical properties and the activity is described by a general equation; $\upsilon = f(p)$, where υ is the activity (usually defined as log (1/C), where C is the concentration of the compound required to produce standard response in a given time), and p is the molecular descriptor (i.e., structure-derived properties of the molecule). These properties that influence the activity of a drug are quite diverse, the major ones being lipophilicity, steric effects and electronic effects. The parameters commonly used to represent these properties are partition coefficients for lipophilicity, Taft steric constants for steric effects, and Hammett σ constants for electronic effects.[4,24]

2.5.3.2.1 Partition Coefficients (P)

A drug has to pass through the biological membranes in order to reach its site of action. Consequently, the relative solubilities of a drug in the aqueous medium and lipids are crucial in the transport of that drug to its site of action, especially at the aqueous medium/lipid interface. P measures the distribution of a compound between two immiscible solvents, thus attempts have been made to correlate the activities of drugs with their lipid/water partition coefficients. It is not easy to measure P *in situ*, so the less accurate model systems (e.g., organic solvent/aqueous solution) are used; $P =$ [drug in the organic phase] / [drug in the aqueous phase]; (usually, water or a phosphate buffer at the pH of blood (7.4) is the most commonly used aqueous phase and *n*-octanol is the most commonly used organic phase).[4,24] The nature of the relationship obtained depends on the range of P values for the used compounds. If this range is

small the results may be expressed as a straight line equation having the general form[2]:

$$\log (1/C) = k_1 . \log P + k_2$$

where, k_1 and k_2 are the constants.

Over large ranges of P values, the graph of log (1/C) versus log P often has a parabolic form with a maximum value (log P^o). Thus, there is an optimum balance between the aqueous and lipid solubility for maximum activity. Below P^o the drug will be reluctant to enter the lipid bilayer, whereas, above P^o the drug will be reluctant to leave from the lipid bilayer.[4] Therefore, analogues that have P values near the P^o are likely to be the most active and worth for further investigation. Hansch showed that many of these parabolic relationships could be represented reasonably accurately by equations of the form[24]:

$$\log (1/C) = -k_1 (\log P)^2 + k_2 \log P + k_3$$

An alternative way to express the Hansch equation is to use a lipophilic substituent constant (π). This is the logarithm of the P of a compound with substituent X relative to a parent compound in which the substituent is hydrogen[24]:

$$\pi = \log (P_X/P_H)$$

Thus,

$$\log (1/C) = -k_1 (\log \pi)^2 + k_2 \log \pi + k_3$$

π can be used as an alternative to P, when dealing with a series of analogues in which only the substituents are different.

2.5.3.2.2 Electronic Parameters

The distribution of the electrons in a drug molecule will have a considerable influence on the distribution and activity of the drug. As a general rule, non-polar and polar drugs in their unionized form are usually more readily diffused through membranes than polar drugs and drugs in their ionized forms.[4,24] Once the drug reaches the ligand binding site at the receptor the distribution of electrons in its structure will determine the type of electrostatic bonds between drug and receptor.[4,24] The distribution of electrons within a molecule depends on electron-donating and electron-withdrawing

groups found in the molecule.[24] Louis Hammett observed that the dissociation constants (K) of aromatic acids are influenced by the electronic properties of the substituents on the phenyl ring. For example, benzoic acid is weakly ionized in water[24]:
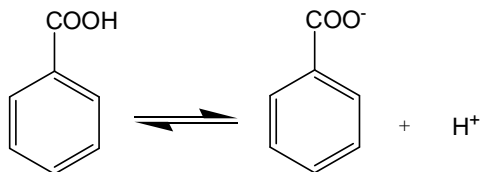
The K values of substituted benzoic acids indicate that electron-withdrawing groups (e.g., nitro group) increase the dissociation, while electron-donating groups (e.g., ethyl) decrease it. A similar effect exists for other equilibrium such as substituted phenyl acetic acids (Table 2.1)[24]:
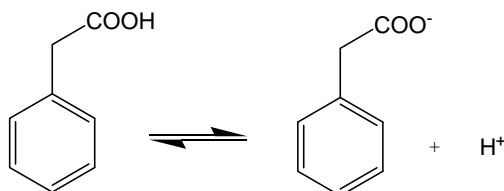
When plotting the quantity log (K/K_0) for benzoic acids on the x axis, where K and K_0 are the constants for unsubstituted and substituted compounds, respectively, and corresponding values measured for the same set of substituents in phenylacetic acids on the y axis, Hammett obtained a straight line. Because of the association between dissociation constants and free energies ($\Delta G = -2.3RT\log K$) this phenomenon is known as the linear free energy relationship.[24]

Benzoic acid			Phenylacetic acid		
R	K	log (K/K_o)	R	K	log (K/K_o)
H	6.2×10^{-5} (K_o)	0	H	5.2×10^{-5} (K_o)	0
NO_2	37.1×10^{-5}	0.78	NO_2	14.1×10^{-5}	0.43
Et	4.4×10^{-5}	-0.15	Et	4.2×10^{-5}	-0.09

Table 2.1 K and log (K/K_0) values for unsubstituted and substituted benzoic acids (left) and for same set of unsubstituted and substituted phenylacetic acids.

Thus, straight line on Figure 2.3 can be written as a linear equation, the Hammett equation[24]:

$$\log (K/K_0) = \rho \log (K/K_0);$$

$$\log (K/K_0) = \rho \sigma$$

where, ρ is related to a given scaffold (e.g., phenylacetic acids), and σ is a descriptor of a substituent and describes its influence on the K. The parameter ρ describes the magnitude of the effect a substituent can exert on the dissociation reaction of a given scaffold. As the distance between the substituent and the dissociated proton increases, its influence on the dissociation reaction decreases and so does the value of ρ. A negative or positive σ value indicates that substituent is acting as an electron-donor or electron-withdrawing groups, respectively. The value of σ varies depending on the position of the substituted group (i.e., ortho, meta, para) in the molecule.[24] A substituent may have an opposite sign of σ depending on its position on the ring because σ includes both inductive and resonance contributions of the electron distribution. For example, σ_m for the methoxy group of the *m*-methoxybenzoic acid is 0.12, however σ_p (*p*- methoxybenzoic acid) has the value of -0.27, because in the former case the electron distribution is dominated by the inductive contribution, whereas in the latter case it is controlled by the resonance effect.[24]
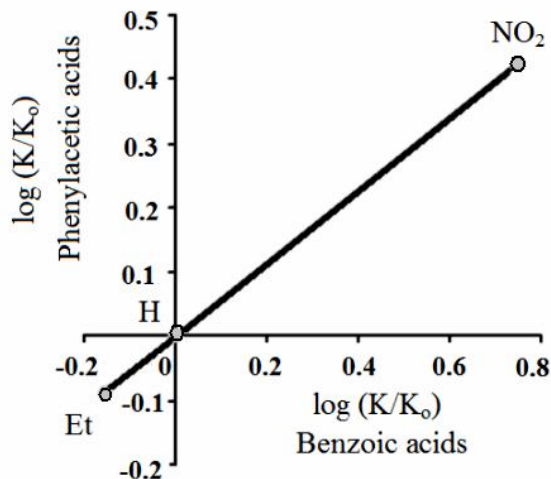

Figure 2.3 Plot of log (K/K_o) values of benzoic acids versus log (K/K_o) values of phenylacetic acids.

2.5.3.2.3 Steric Parameters

The Taft steric parameter (E_s) was the first attempt to show the relationship between a measurable parameter related to the shape and size (bulk) of a drug and the dimensions of the target site and activity of a drug.[24] Taft used the relative rate constants of the acid-catalysed hydrolysis of α-substituted methyl ethanoates in order to define his steric parameter, because it had been shown that the rates of these hydrolyses were almost entirely dependent on steric factors[24]:

$$E_s = \frac{\log k_{(XCH_2COOR)}}{\log k_{(CH_3COOR)}} = \log k_{(XCH_2COOR)} - \log k_{(CH_3COOR)}$$

where, k is the rate constant of the appropriate hydrolysis, when X = H, $E_s = 0$.
E_s is an experimental value based on rate constants for a given model reaction. The bulkier the substituent, the more negative the E_s.

Hansch postulated that the biological activity of a drug could be related to all or some of above stated factors by a simple mathematical relationships based on the general formula[4,24]:

log(1/C) = c_1(partition parameter) + c_2(electronic parameter) + c_3(steric parameter) + c_4

where, c_1, c_2, c_3 and c_4 are numerical coefficients obtained by feeding the data into a suitable computational statistical package. Parameters in the above equation are also known as molecular descriptors or simply descriptors. It must be noted that, several parameters different than above parameters (e.g., molecular weight, density, molar refractivity, etc.) can be used to correlate biological activity.

2.5.3.2.4 Deriving a QSAR Equation

The starting point for deriving a QSAR equation is the study table. It consists of a spreadsheet with molecules (data set) across the rows and molecular characteristics (biological activity, descriptors) down the columns. Typically, the first column indicates the molecular identification (e.g., compound number), the second column shows the activity of compounds in the data set, and subsequent columns present the values of the corresponding descriptors (Table 2.2). The most widely used method for deriv-

ing QSAR equations is linear regression, which uses the least-squares fitting to find the best combination of coefficients in the equation. The least-squares technique can be illustrated by using a simple case where the activity is a function of only one descriptor.[4] Therefore, the form of equation will be $y = mx + c$, where, y is the dependent variable (observations), (e.g., activity) and x is the independent variables (descriptors), (e.g., $\log P$). The objective of a regression analysis is to find the coefficients m and c that minimize the sum of deviations of the observed values from the fitted equation.[4,24]

The least-squares coefficients m and c in the linear regression equation are given by[2,4,24]:

$$m = \frac{\sum_{i=1}^{n}(x_i - <x>)(y_i - <y>)}{\sum_{i=1}^{n}(x_i - <x>)^2} ; \qquad c = <y> - m <x>$$

where, $<x>$ and $<y>$ are the mean values of the independent and dependent variables, respectively. The 'quality' of a simple linear regression equation is often reported as the squared correlation coefficient r^2 value. This value indicates the fraction of the total variation in the dependent variables that is explained by the regression equation. In order to determine r^2 value, the total sum of squares (TSS) of the deviations of the observed y values from the mean $<y>$ value is calculated together with the explained sum of squares (ESS), which is the sum of square deviations of the y values calculated from the model, $y_{calc.,i}$ from the mean. The $y_{calc.,i}$ is obtained by feeding the appropriate x_i value into the regression equation. Other common squared term is the residual sum of squares (RSS), which is the sum of squares of the differences between the observed and calculated y values[2,3,24]:

$$TSS = \sum_{i=1}^{N}(y_i - <y>)^2, \quad ESS = \sum_{i=1}^{N}(y_{calc.,i} - <y>)^2, \quad RSS = \sum_{i=1}^{N}(y_i - y_{calc.,i})^2$$

where, TSS is equal to the sum of ESS and RSS. The r^2 is then given by:

$$r^2 = (ESS/TSS) = (1 - (RSS/TSS))$$

The r^2 adopts values between 0 and 1; where, 0 indicates that none of the variations in the observations is explained by variation of independent variables in the equation and 1 indicates that all of the variation in the observed values can be explained.

	Data set	Property of interest	Descriptors			
	Compound	Activity	logP	Molecular Refractivity	Molecular Weight	Density
	1					
	2					
	3					
	4					
	n					

Table 2.2 An example of QSAR study table.

It is not always possible to correlate biological activities with a single descriptor (linear model with one descriptor). Given that biological action results from the combined influence of many factors, one can extend the QSAR model to multiple descriptors. Indeed, the observation that several parameters used simultaneously can lead to good models prompted the development of a method referred to as multiple linear regression (MLR).

In MLR, the activity is expressed as a linear combination of descriptors. In most cases, the obtained model with the minimum sum of deviations from the observations is not perfect and an error (e) is usually unavoidable. Thus,

$$y = \sum_{j=1}^{m} b_j x_j + e$$

where, y is the activity, x_j is the value of descriptor j and b_j its associated coefficient, e is the error.[4,24]

Another quantity that is commonly reported in the QSAR studies is the F statistics. The F value is the ratio of the explained mean square divided by the residual mean square. The F values are available in statistical tables at different levels of confidence;

if the calculated value is greater than this tabulated value, then the equation is significant at that particular confidence level.[24] It must be noted that the F value depends upon the number of independent values in the equation and the number of data points.[24] As the number of data points increases and/or the number of independent variables falls, the value of F which corresponds to a particular confidence level also decreases. Because, the aim is to obtain a model to be able to explain a large number of data points with an equation containing as few variables as necessary; such an equation would be expected to have a greater predictive power. The F-test employs the F-distribution to test whether the r^2 obtained from the MLR analysis significantly differs from 0. The larger the F-value, the larger the probability that r^2 significantly differs from the 0. This is formally taken into account via the number of degrees of freedom associated with each parameter. An MLR is associated with N-1 degrees of freedom, because the fitted line always passes through the means of the dependent and independent variables. The TSS is always associated with N-1 degrees of freedom. If there are p independent variables (descriptors) in the equation, then there are N-p-1 degrees of freedom associated with the RSS and p degrees of freedom associated with the ESS. Therefore, the explained mean square equals to ESS divided by p; and the residual mean square equals to RSS divided by N-p-1, and so F is given by[24]:

$$F = \frac{\text{ESS}}{p} \frac{N-p-1}{\text{RSS}}$$

One of the easiest ways to validate a QSAR model is to calculate the standard error or standard deviation (SE or SD), which is calculated as the average square deviation of each number (the residuals) from the mean. The smaller the SD, the higher the quality of the model[24]:

$$SD = \sqrt{\frac{\sum (y_{obs.} - y_{calc.})}{n-p-1}}$$

where, $y_{obs.}$ and $y_{calc.}$ are the observed and calculated activities, respectively, n is the number of data points and p is the number of descriptors used.

2.5.3.2.5 Cross-validation

The indices of r^2 and F can be generated to evaluate the quality of the obtained QSAR models. However, these parameters basically only tell us about the ability of the QSAR model to reproduce the data from which it was derived and not its aptitude to predict the activities of new compounds.[4,24] Two methods can be used to estimate the predictive power of a QSAR model. The first method is known as the test set method and consists of partitioning the initial data into two sets, a preferred strategy when a large data set of compounds is available. The initial data set is divided into two parts; the first one (training set) is usually used to build a QSAR model and the second one (test set) to validate this model. The second method is known as the cross-validation method, it is preferred when the data set is too small. In this method, the data set is divided to N equal parts, N-1 parts are used to build the model which is then used for the remaining Nth part to predict the activities of the corresponding molecules. The procedure is repeated until all activities of all compounds have been predicted independently. The most common form of the cross-validation method is the leave-one-out method, in which each data value is left out and a model derived using the remainder of the data. A value then can be predicted for the data left out and can be compared with the true observed value and this procedure is repeated for every data point. This procedure permits the calculation of cross-validated r^2 (r_{cv}^2) or q^2.

The q^2 is computed by analogy with r^2, the difference being the use of the predicted sum of squares (PRESS) rather than the RSS in the numerator. PRESS is calculated as the difference between the measured activity and the predicted activity for the test set[2,4]:

$$q^2 = 1 - \frac{PRESS}{\sum_{i=1}^{n}(y_i - <y>)^2}; \qquad PRESS = \sum_{i=1}^{n}(y_{pred,i} - y_i)^2$$

2.5.3.2.6 Principal Components Regression (PCR)

MLR can not deal with data sets, where the variables are highly correlated and/or where the number of variables exceeds the number of data points. Two methods are commonly used to deal with such situations; PCR and partial least squares (PLS). In

PCR, the variables are subjected to principle component analysis (PCA), and then regression analysis is performed using the first few principle components. PCA is a commonly used technique for reducing the dimensionality of a data set (dimensionality of a data set is the number of variables that are used to describe each objects). Consider the data shown in Figure 2.4, there is a high correlation between x and y values. If we were to define a new variable, $z = x + y$, then we could express most of the variation in the data as the values of this new variable z, which is called principal component. In general, the principal component (p) is a linear combination of the variables[2,4]:

$$p_i = \sum_{i=1}^{v} c_{i,j} x_i$$

where, p_i is the principal component and c_i is the coefficient of the variable x_j (there are v such variables). The first principle component of the data set corresponds to that linear combination of the variables that gives the best fit straight line through the data when they are plotted in the v-dimensional space. The second and subsequent principle components account for the maximum variance in the data not already accounted for the previous principle components. Each principle components corresponds to an axis in a v-dimensional space, and each principle component is orthogonal to all the other principle components. Indeed, in order to explain all of the variation in the data, it is usually necessary to include all the principle components (there can be as many principle components as there are dimensions in the data), however, in many cases only a few principle components are enough to explain a significant proportion of the variation in the data. The principle components are calculated using standard matrix methods. The first step is to calculate the variance-covariance matrix; if there are s observations (number of compounds), each of them contains v values (number of descriptors), and then the data set can be represented as a matrix $\mathbf{D}$ with v rows and s columns. The variance-covariance matrix $\mathbf{Z}$ is[2]:

$$\mathbf{Z} = \mathbf{D}^T \mathbf{D}$$

The eigenvectors of Z are the coefficients of the principle components. As the Z is a square symmetric matrix, its eigenvectors will be orthogonal.[2] The eigenvalues and their associated eigenvectors can be obtained by solving the secular equation $|\mathbf{Z} - \lambda \mathbf{I}| =$

0. The first principle component corresponds to the highest eigenvalue, the second principle component to the second highest eigenvalue, and so on.[2]
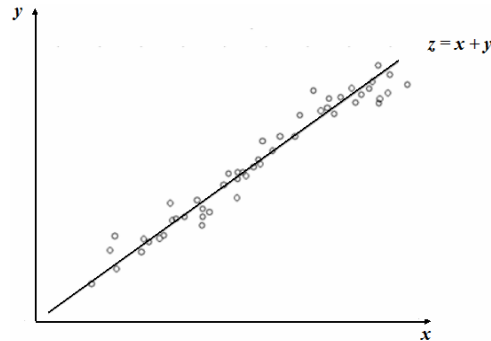

Figure 2.4 Most of the variance in this set of highly correlated data values can be explained in terms of a new variable z, $z = x + y$.

A general rule of thumb is that only those principle components that have eigenvalues which are greater than 1 should be used for inclusion in a PCR.

2.5.3.2.7 Partial Least Squares (PLS)

PLS is an alternative technique to PCR. The PLS expresses a dependent variable y in terms of linear combination of independent variables[2] x_i:

$$y = b_1 t_1 + b_2 t_2 + b_3 t_3 + \ldots + b_m t_m$$

where,

$$t_1 = c_{11}x_1 + c_{12}x_2 + c_{13}x_3 + \ldots + c_{1p}x_p$$
$$t_2 = c_{21}x_1 + c_{22}x_2 + c_{23}x_3 + \ldots + c_{2p}x_p$$
$$t_3 = c_{31}x_1 + c_{32}x_2 + c_{33}x_3 + \ldots + c_{3p}x_p$$
$$t_m = c_{m1}x_1 + c_{m2}x_2 + c_{p3}x_3 + \ldots + c_{mp}x_p$$

t values are called latent variables (or components) and are constructed in such a way that they form an orthogonal set. The PLS method is using as a tool for decreasing the variations of many correlated descriptors in a data to a few uncorrelated latent variables. The use of orthogonal linear combinations of the x values is very similar with

PCA. The major difference is that the latent variables in a PLS are constructed to explain not only the variation in the independent variables (*x*) but also to simultaneously explain the variation in the dependent variables (observations, *y*).

2.5.3.3 3D QSAR

3D QSAR is a method based on statistical correlation analysis enabling the comparison of 3D molecular forces (fields) produced in the vicinity of different compounds to find correlations between biological activities and these fields.[2] QSAR approaches aim to establish relationships between biological activities and chemical structure. In classical QSAR (2D QSAR) molecular properties are described by parameters that are not x,y,z dependent (e.g., log*P*, MR, E_s, σ, π, etc.), whereas in 3D QSAR, they are represented by a set of values of x, y, z functions, measured at many different locations in the space around the molecules. Thus, in 3D QSAR, there are many more descriptors than 2D QSAR. The comparative molecular field analysis (CoMFA) and the comparative molecular similarity indices analysis (CoMSIA) are two most popular approaches used in 3D QSAR studies.

2.5.3.3.1 CoMFA

CoMFA was first described by Cramer *et al.*, in 1988.[25] Prior to the CoMFA analysis, each compound in the data set has to be subjected to conformational analysis and it is presumed as the active conformation at the binding site of the receptor. These conformers must be superimposed on a selected template compound from the data set. The molecular fields (e.g., steric, electrostatic) surrounding each molecule are then calculated by placing appropriate probe groups at points on a regular lattice (in order to simplify calculations of the fields created around a molecule, the method consists of superimposing 3D lattice defining grid points regularly distributed in space, and calculating the interaction energy between the molecule and the probe at each grid point, using a potential energy function, see Figure 2.5). In order to test the presence of a field and to measure, it requires the use of probes with associated energy functions.[2] Usually a probe atom is employed which is placed at well selected points in the

space, to quantitatively measure the value of the field created by the molecule at the point considered. The probe must be of same type of the field to be measured (e.g., van der Waals probes for steric fields, charged probes for electrostatic fields). The electrostatic field is obtained by calculating the electrostatic interaction energy between the molecule and a probe at each grid point using Coulomb's law[2]:

$$E_{es} = \frac{1}{4\pi\varepsilon_0} \sum_{i=1}^{n} \frac{q_i q_p}{r_i}$$

where, q_i and q_p are the atomic charges for the ligand and the probe, respectively.

The steric field is obtained by calculating the van der Waals interaction energy between the molecule and the probe at each grid point using for example a Lennard-Jones potential[2]:

$$E_s = \sum_{i=1}^{N} \frac{A}{r_i^{12}} - \frac{B}{r_i^6}$$

where, A and B are specific parameters, different for each interacting pair of atoms.
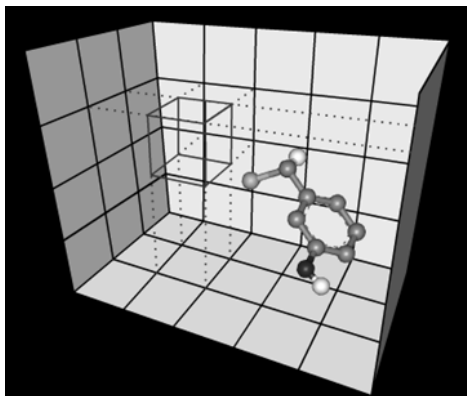
Figure 2.5 3D lattice defining grid points makes it possible to sample the space with a finite number of points of molecular interactions fields.

2.5.3.3.2 CoMSIA

CoMSIA was developed by Klebe *et al.*,[10] in 1994 and it is an extension of the CoMFA methodology. Both models are forms of QSAR and are based on the assump-

tion that changes in binding affinities of ligands are related to changes in molecular properties, represented by fields. CoMFA approach has been widely accepted and a useful technique in 3D QSAR, however, it is not without problems. In particular, both steric and electrostatic potential functions are very steep near the van der Waals surface of the molecule, causing rapid changes in surface descriptions, and they require the use of cut-off values so calculations are not done inside the molecular surface. In addition, a scaling factor is applied to the steric field, so both fields can be used in the same PLS analysis. Finally, changes in orientation of a set of superimposed molecules, relative to the calculation grid, can cause significant changes in CoMFA results due to steric cut-off values. In CoMSIA, similarity indices are calculated at regularly spaced grid points for the pre-aligned molecules. A comparison of the relative shapes of CoMFA and CoMSIA fields is shown in Figure 2.6. For the distance dependence between the probe atom and the ligand atoms a Gaussian function is used. Because of the different shape of Gaussian function, the similarity indices can be calculated at all grid points, both inside and outside the molecular surface. The equation used to calculate similarity indices is as follows[9]:

$$A_{F,k}^{q}(j) = \sum_{i} w_{probe,k} w_{ik} e^{-\alpha r_{iq}^{2}}$$

where, A is the similarity index at grid point q, summed over all atoms i of the molecule j under investigation, $w_{probe,\,k}$ is the probe atom with radius 1 Å and charge +1, w_{ik} is the actual value of the physicochemical property k of atom i, r_{iq} is the distance between the probe atom at grid point q and atom i of the test molecule, α is the attenuation factor and its optimal value is normally between 0.2 and 0.3.[26]
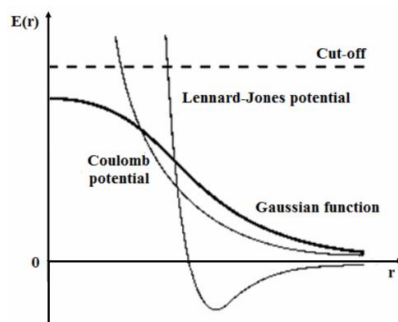

Figure 2.6 Shapes of various functions.

2.5.4 *De novo* Drug Design

In *de novo* design, the 3D structure of the receptor or the 3D pharmacophore is used to design new molecules. There are two basic types of *de novo* design algorithms; 'outside-in' and 'inside-out' methods. In the first class of methods, the binding site is first analyzed to determine where specific functional groups might bind tightly. These groups are connected together to give molecular skeletons, which are then converted to 'real' molecules. In the second class of methods, molecules are grown within the binding site under the control of an appropriate search algorithm, with each proposition being evaluated using an energy function.[2]

Chapter 3. Computational Details
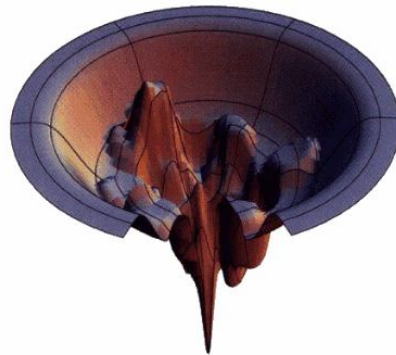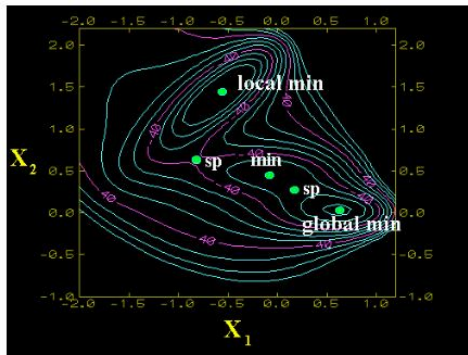
3.1 Minimization Methods

Energy is a function of the atomic coordinates of the system and software programs attempt to generate the xyz coordinates of atoms which correspond to a minimum energy. This is accomplished by minimization procedures and these techniques are iterative in which the atomic coordinates are altered from one iteration to next in order to minimize the energy. Commonly used optimization methods are:[2]

(i) The Steepest Descent (SD) Method. The SD method uses a first-order minimization algorithm and changes the coordinates of the atoms in the system gradually in order to reach closer and closer to minimum energy. In the SD method, r is defined as a vector of all coordinates of N atoms in the system. Then, the net force on each atom F and the potential energy at each iteration step t are calculated. New positions are calculated by[2]:

$$r_i^{t+1} = r_i^t + F(r_i) \qquad i = 1, 2, 3,.....N.$$

The geometry optimization process stops, when the specified predetermined threshold conditions by the user are fulfilled. The SD is often used for structures far from the minimum as a rough and introductory run followed by a subsequent minimization procedure employing a more advanced algorithm (e.g., conjugate gradient).[1]

(ii) The Conjugate Gradients (CG) Method. In the SD method, both gradients and direction of successive steps are orthogonal, however in the CG method, the gradients at each point are orthogonal but the directions are conjugate. New positions are calculated by[2]:

$$r_i^{t+1} = r_i^t + h_i^{t+1} \qquad i = 1, 2, 3,.....N.$$

where, $h_i^{t+1} = F(r_i^{t+1}) + \gamma_i^t . h_i^t$

and, a scalar constant $\gamma_i^t = \dfrac{F(r_i^{t+1}).F(r_i^{t+1})}{F(r_i^t).F(r_i^t)}$

Convergence properties of CG are superior to SD method.[2,19]

(iii) The Powell Method: Another first-order minimization algorithm is the Powell method and belongs to the CG family of optimization method. It applies a more efficient CG method to determine the descent direction and it is more tolerant to inexact line searches.[2,19]

(iv) The Newton-Raphson Method: This method does not use only first derivatives, but also the second derivatives in the minimization procedures. Second derivatives give information about the curvature of the function.[2,19]

3.1.1 Geometry Optimizations of Investigated Systems

The structures of the studied molecules were subjected to geometry optimization using a combination of the standard Tripos MM force field of the Sybyl molecular modeling package[27] (Powell energy minimization algorithm, Gasteiger-Huckel charges and 0.001 kcal/mol Å energy gradient convergence criterion). For the conformational analysis of template compounds used in 3D QSAR, *ab initio* B3LYP/6-31G* level QM calculations were also performed.[26,28-31]

3.2 MC Simulations

MC analysis of ligands has been carried out with the CHARMm force field of QUANTA package[32] (Powell energy minimization algorithm, Gasteiger-Huckel charges and 0.001 kcal/mol Å energy gradient convergence criterion) as well as the semi-empirical methods of AM1 and PM3 (SCF convergence criterion has been set to 10^{-6} as energy gradient convergence limit).[26,29,30]

3.3 3D QSAR/CoMFA and CoMSIA Settings

The steric and electrostatic field energies were calculated using the Lennard-Jones and the Coulomb potentials, respectively with a 1/r distance-dependent dielectric constant in all intersections of a regularly spaced (2 Å) grid. An sp^3 carbon atom with a radius of 1.53 Å and a charge of +1.0 was used as a probe to calculate the steric and

electrostatic energies between the probe and the molecules using the Tripos force field. The truncation for both the steric and electrostatic energies was set to 30 kcal/mol.[26-28,30]

3D QSAR/CoMFA and CoMSIA PLS Analysis and Validations: The initial PLS analysis was performed using the "leave-one-out" cross-validation method for all 3D QSAR analyses. A minimum column filtering value of 2.00 kcal/mol was set to improve the signal to noise ratio by omitting those grid points whose energy variation was below this threshold. In both CoMFA and CoMSIA analyses, the descriptors were treated as independent variables, whereas the pK_i values were treated as dependent variables in the PLS regression analyses in order to derive the 3D QSAR models. The final model (non-cross-validated conventional analysis) was developed from the model with the highest r_{cv}^2, and the optimum number of components was set equal to that yielding the highest r_{cv}^2. The non-cross-validated models were assessed by the conventional correlation coefficient r^2, standard error of prediction, and F values. For the creation of the CoMFA field, '*CoMFA standard*' scaling was selected, while in the case of CoMSIA '*none*' option was selected in the Sybyl.[26,28,29,31]

3.4 Molecular Docking

Flexible docking studies have been applied using FlexX program of Sybyl molecular modeling package.[27] FlexX is a flexible docking method that uses an incremental construction algorithm to place ligands into a binding site of a receptor.[23] The base fragment (the ligand core) is automatically selected and placed into the active site of the receptor using a pose clustering algorithm.[2] Then, the remainder of the ligand is built up incrementally from other fragments. Finally, placement of the ligand is scored on the basis of protein-ligand interactions and the binding energy is estimated and placements are ranked.

The aim of base placement is to allow favorable simultaneous interactions between ligand and protein by pose clustering method.[2] In this algorithm, the base fragment and receptor site are represented as a finite set of interaction points. A unique trans-

formation of base fragment into binding site is defined by superposing three interaction sites from the base fragment into three interaction sites on the receptor. Thus, the first step of the algorithm is to find all such compatible triangles.[2] All placements are clustered according to RMSD between sets of two placements, using the complete-linkage hierarchical cluster algorithm.[2] Basically, the two nearest clusters are merged into one, as long as the distance between them is lower than the predefined threshold. Conformational flexibility is taken into consideration by systematically generating low energy conformations of the ligand. Bond lengths and bond angles are held fixed at their input values and a set of up to twelve preferred dihedral angles is assigned according to records from a molecular fragment database. A grid box technique is used in the interaction energy calculations. Basically, the system is put inside a cubic grid, and to check a solution for steric overlap between receptor and ligand, only those receptor atoms in cubes within the van der Waals radius of the ligand atom and the largest receptor atom radius must be checked. For each ligand-receptor interaction, the interaction center on the ligand (l_i), the interaction center on the receptor (r_i) and energy contribution of the interaction (w_i), provides optimal geometry. For the superimposition routing, l_i are fitted onto r_i, thus $\Sigma w_i (l_i - r_i)^2$ is minimized. In order to guide the growth of the ligand, a numerical measure of "goodness of fit" must be determined in order to measure the strength of protein-ligand interactions.[2]

The physicochemical model behind FlexX can be divided into three parts: the analysis of the conformational space of the ligand, the model of protein-ligand interactions, and the scoring function. The scoring function of FlexX, developed by Böhm in order to rank the solutions, is an estimation of the free binding energy ΔG of the protein-ligand complex[23]:

$$\Delta G_{bind} = \Delta G_0 + \Delta G_{H\text{-}bonds} \sum_{H-bonds} f(\Delta r, \Delta \alpha) + \Delta G_{ionic} \sum_{ionic} f(\Delta r, \Delta \alpha) + \Delta G_{arom} \sum_{arom} f(\Delta r, \Delta \alpha) +$$

$$\Delta G_{lipo} \sum_{lipo} | A_{lipo} | + \Delta G_{rot} N_{rot}$$

$$f(\Delta r, \Delta \alpha) = f1(\Delta r) f2(\Delta \alpha)$$

$$f1(\Delta r) = \begin{cases} 1 & \Delta r \le 0.2 \text{ Å} \\ 1 - (\Delta r - 0.2)/0.4 & \Delta r \le 0.6 \text{ Å} \\ 0 & \Delta r > 0.6 \text{ Å} \end{cases}$$

$$f2(\Delta\alpha) = \begin{cases} 1 & \Delta\alpha \le 30^\mathrm{O} \\ 1 - (\Delta\alpha - 30)/50 & \Delta\alpha \le 80^\mathrm{O} \\ 0 & \Delta\alpha > 80^\mathrm{O} \end{cases}$$

where; $f(\Delta r, \Delta\alpha)$ is a scaling function that penalizes deviations from ideal geometry. ΔG_0 is a contribution to the binding energy that does not directly depend on any specific interactions with the protein. It may be considered as a reduction of binding energy due to overall loss of translational and rotational entropy of ligand. $\Delta G_{H\text{-}bonds}$ represents the contribution from an ideal hydrogen bond. ΔG_{ionic} describes the contribution from an unperturbed ionic interaction. ΔG_{arom} accounts for the interactions of aromatic groups and ΔG_{lipo} represents the contribution from lipophilic interactions. ΔG_{rot} describes the loss of binding energy due to freezing of internal degrees of freedom in the ligand. N_{rot} is the number of rotatable bonds that are immobilized in the complex. It is assumed that such interactions are proportional to A_{lipo}, the lipophilic contact surface between the protein and the ligand.

(i) Molecular docking studies for CB ligands at the active site of CB1 and CB2 receptors[30,31]: 3D models of the CB1 and CB2 receptors based on template rhodopsin from Tuccinardi *et al.*[33] has been used for the initial docking experiments, however, the critical binding site residues were determined from several molecular modeling studies of CB receptors (e.g., Tuccinardi *et al.*[33], Salo *et al.*[34] and Shim *et al.*[8]). The active site in the docking runs included all atoms within a radius of 6.5 Å around the critical amino acids for CB1 receptor: Phe174, Leu190, Lys192, Leu193, Gly195, Val196, Thr197, Phe200, Thr201, Pro251, Trp356, Leu359, Ser383, Cys386, Leu387 and for CB2 receptor: Leu108, Ser112, Pro168, Leu169, Trp194, Trp258. (The complete list of amino acid residues form the binding site is detailed in the Appendix). In addition docking simulations were repeated with 3D models of the CB1 and CB2 receptors produced using β2-adrenergic template receptor.

(ii) Molecular docking studies for fullerene analogues at the binding cavity of HIV-1 PR[28,29]: Since the X-ray structure of the HIV-1 PR complexed with fullerene based

inhibitors has not yet been reported, the initial structure was taken from the HIV-1 PR complexed with haloperidol derivative at 2.2 Å resolution (pdb code; 1AID).[35] The water molecules and the inhibitor were then removed and all hydrogens were added to the system. Ionization states for ionizable amino acid residues were assigned according to the standard pK_a values of amino acids. The geometry of the enzyme has been optimized by using the Tripos MM force field. The maximum number of conformers for each molecule was set to 30 and the top 10 lowest energy conformers were used in docking simulations and only the best docked complex was considered for further analysis. The default FlexX scoring function was used for the simulations. FlexX uses formal charges, which were turned on during the docking. The active site in the docking runs, included all atoms within a radius of 6.5 Å around the critical amino acids: Asp25, Asp25', Ile50, and Ile50'. (Full list of binding site residues have been detailed in the Appendix).

3.5 MD Simulations

The MD simulations were performed with Groningen Machine for Chemical Simulations (GROMACS) version 3.3.1 software package[36] using the gmx force field.[36] Simulations were run with periodic boundary conditions. In a truly macroscopic system, surface effects may not be so crucial in the properties of the system, because as a spherical system increases in size, its volume grows as the cube of the radius, while its surface grows as the square. However, in a typical simulation, computational resources inevitably constrain the size of the system to be so small that surface effects may play an important role in the properties of the system under study. Thus, modeling of a cluster may not tell one much about the behaviour of a macroscopic system. In order to eliminate surface effects from the computation, periodic boundary conditions can be used.[2] Periodic boundary condition are represented in Figure 3.1. Under periodic boundary conditions, the system which is modeled is assumed to be a unit cell in some ideal crystal (e.g., cube, hexagonal prism, orthorhombic). Periodic boundary conditions combined with the minimum image convention: only one, the nearest, image of each particle is considered for short-range non-bonded interaction terms. For the long-range electrostatic interactions this is not always accurate enough, and has to

and has to be incorporated with lattice sum methods (e.g., Ewald sum, particle mesh Ewald). In these methods, a particle interacts with all other particles in the simulation box and with all of their images in an infinite array of periodic cells.
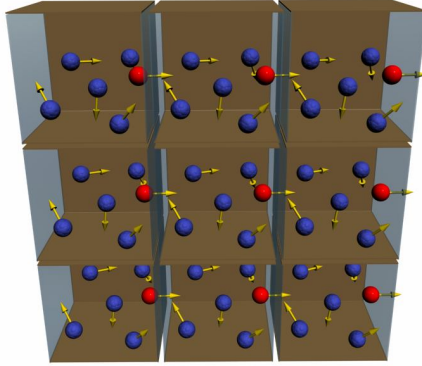

Figure 3.1 Periodic boundary conditions.

Because of heating due to external or frictional forces, it is necessary to control the temperature of the system. In MD simulations, Berendsen thermostat were used.[37] This algorithm mimics weak coupling with first-order kinetics to an external heat bath with given reference temperature T_0. Deviation of the system temperature from T_0 is slowly corrected by[38]:

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{T_0 - T}{\tau}$$

where, τ is the temperature coupling which is given by:

$$\tau = 2\frac{C_V \tau_T}{N_{df} k}$$

where, C_V is the total heat capacity of the system, k is Boltzmann constant, and N_{df} is the number of total degrees of freedom, τ_T is the user defined temperature coupling time constant and close to τ but not exactly equal because the kinetic energy change caused by scaling the velocities is partly redistributed between kinetic and potential energy, thus change in temperature is less than the scaling energy. Similarly with the temperature coupling, the system can also be coupled to a pressure bath. For the MD simulations of the systems, Berendsen barostat was applied.[37] The Berendsen algorithm rescales the coordinates and box vectors with a scaling matrix μ in every step,

which has the effect of a first-order kinetic relaxation of the pressure with given refer-ence pressure P_0[38]:

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \frac{P_0 - P}{\tau_P}$$

The scaling matrix μ is given by[38]:

$$\mu_{ij} = \delta_{ij} - \frac{\Delta t}{3\tau_P} \beta_{ij} \{P_{0ij} - P_{ij}(t)\}$$

where, β is the isothermal compressibility of the system.

All bonds were constrained using the linear constraint solver (LINCS) algorithm[39]. LINCS is an algorithm that resets bonds to their correct lengths after an unconstrained update. Visualization of the dynamics trajectories was performed with the visual mo-lecular dynamics (VMD) software package[40].

3.5.1 Details of the Performed Simulations

3.5.1.1 MD Simulations of CB

(i) In DPPC lipid bilayers (without receptor) MD simulations: The coordinates of AMG3 conformers were used as input at the PRODRG program[41] in order to obtain topologies which will be used in the MD simulations. DPPC lipid bilayer for the MD simulations was obtained from Dr. M. Karttunen (128 DPPC lipids and 3655 water molecules after 100 ns).[42,43] Simulations were run in the NPT ensemble at 300 K and 1 bar with periodic boundary conditions. Berendsen barostat and thermostat algo-rithms were used.[37] Electrostatic interactions were calculated using the particle mesh Ewald method.[44] Cut-off distances for the calculation of Coulomb and van der Waals interactions were 1.0 and 1.4 nm, respectively. Prior to the dynamics simulation, en-ergy minimization calculations were applied to the full system without constraints us-ing the SD method for 5000 steps with the initial step size of 0.01 Å (the minimiza-tion tolerance was set to 100 kJ/(mol.nm)). The system was then equilibrated via 250 ps simulations with a time step of 2 fs, subsequently a 2.5 ns simulations were per-formed at 300 K and 1 bar with a time step of 2 fs. All bonds were constrained using

the LINCS algorithm.[39] Origin 6.0 program was used for dihedral angles versus time plots and statistical calculations.[45] *(ii) Membrane associated receptor MD simulations:* The lipid used *in solution* calculations was employed here, however the lipid extended by 4x4x1 in xyz directions, respectively in order to have enough area of lipid for the protein merging. Same running parameters with *DPPC lipid bilayers (without receptor)* MD simulations were used. *(iii) MD simulations at the homology modeling step of the receptors:* Same parameters with above defined MD have been used. The convergence criteria have been tested by the potential energy versus time plot and showed sufficient convergence. Each receptor backbone structure from a trajectory is compared with reference (initial) structure in order to obtain the receptor backbone RMSD versus time plots (Figures A1 and A2, Appendix).

3.5.1.2 MD Simulations of HIV-1 PR and Fullerene Analogues

Canonical NVT ensemble at 300 K was used with periodic boundary conditions, and the temperature was kept constant by the Berendsen thermostat.[37] Electrostatic interactions were calculated using the particle mesh Ewald method.[44] Cut-off distances for the calculation of Coulomb and van der Waals interactions were 1.0 and 1.4 nm, respectively. Prior to the dynamics simulation, energy of the full system has been optimized without constraints using the SD integrator for 5000 steps. The system was then equilibrated via a 100 ps MD simulations at 300 K. Finally, a 2 ns simulation was performed with a time step of 2 fs. The convergence criteria have been tested by the potential energy versus time plot and showed sufficient convergence. The receptor backbone RMSD versus time plot has also been presented (Figure A3, Appendix).

3.6 *De Novo* Drug Design

LeapFrog algorithm under Sybyl was used in order to automatically generate a series of ligands for the binding pocket of a receptor.[27] Leapfrog is a second generation *de novo* drug discovery program for the design of potentially active compounds, using molecular evolution or electronic screening, by repeatedly making structural changes and then either keeping or discarding the obtained results, depending on the binding

energy results. There are two starting input options in order to generate site point probe atoms that will be used in the binding energy calculations, these are a pharmacophore model or a receptor structure. The charge of the site point probe atom is positive, negative or lipophilic and its value is compared with ±1.0: If the value is smaller than +1.0, it is lipophilic, if the value is bigger than +1.0, site points seek a negative atom and if the value is less than -1.0, site points seek a positive atom in the fragment. Binding energy calculations in LeapFrog were performed by steric, electrosteric and hydrogen bonding enthalpies of ligand cavity binding using the Tripos force field under Sybyl molecular modeling package (v. 6.8).[27] As with many *de novo* drug design programs, central operation of LeapFrog is the processing loop. In each pass through the loop, a type of move is selected randomly. If the move succeeds, a new structure is evaluated. A new ligand which passes evaluation is added to the pool of ligands available for the next move. The fragments used in JOIN, FUSE and BRIDGE are stored as a molecular data base in Sybyl. A hydrogen atom is chosen within the selected ligand, randomly; and a local energy check is performed on its cavity environment within a 3.0 Å radius. If steric interactions are not favorable over more than half of the environment volume, the hydrogen atom is sterically excluded. If the first chosen hydrogen is not accessible, another one is chosen, randomly until an accessible one is found. If no accessible hydrogens are found, the JOIN move fails. The FUSE move process is similar with JOIN; environment checks for steric accessibility are performed as JOIN move, however in a FUSE attempt, existence of a ring bond flanked by hydrogen in both ligand and fragment are required. Thus, FUSE move aims to fuse (usually rings) starting ligand and fragment from data base. The BRIDGE move attempts to bridge available fragments.[46-49]

LeapFrog studies for CBs. As an initial basic procedure of LeapFrog, site point probe atoms were generated using the receptor cavity as well as a pharmacophore model inferred from the PLS results options. Template compound AMG3 was selected as starting structure. Firstly, the OPTIMIZE module was used for the improvement in binding energy. Secondly, several moves such as JOIN, FUSE, BRIDGE and OPTIMIZE options were used after the initial run of 100 moves taken into account the synthetic difficulties. The derived ligands that had the best binding energy were used for repeating the cycle of 5000 moves.

LeapFrog studies for fullerenes. Same parameters used in CBs have been applied for the fullerenes. Monoadduct and bisadduct of [60]fullerene derivatives that have higher affinity with HIV-1 PR (molecules **1**, **3**, **23**, and **42**, Tables 6.1 and 6.2 in Chapter 6) were used as starting structures with restricting to only changes of pendant groups on the [60]fullerenes.

Chapter 4. Strategies in Computational Drug Design

In-silico drug design has supported pharmaceutical research for over three decades. The receptor-based and ligand-based approaches are two major drug design methodologies.[4,24] These strategies, in particular in combination with fragment-based lead discovery (identification of bioactive compounds by assembling small-molecule fragments) are considered to be complementary to high-throughput screening. A prerequisite for ligand-based approach is the availability of at least one bioactive reference structure.[4,24] Therefore, in ligand-based design, attempts are done to derive other compounds having a similar shape and electrostatic potential with the known bioactive compound in order to improve the biological activity. When an X-ray structure or 3D homology model of a receptor is known, then receptor-based approaches can be used to screen compound collections virtually. Usually, both ligand-based and receptor-based techniques may be applied simultaneously. Pseudoreceptor modeling fills the gap between these two approaches.[50] 3D QSAR techniques that map molecular features onto a field surrounding the spatial alignment of reference conformers are used in pseudoreceptor generation. The pseudoreceptor models attempt to capture the shape of the ligand binding pocket.[50] The 3D information of receptor (target) is not always available at the beginning of a drug design project. For example, it is very difficult to obtain the X-ray structure of GPCRs or other membrane-bound proteins. Currently, protein data bank (PDB) contains only about 30 entrie (out of 50,000 total entries) related to GPCR (rhodopsin and β2-adrenergic receptor structures). Thus these projects are limited to ligand-based design methods or structural information may be obtained from homology modeling studies based on available X-ray structure from same family.

Computational design of novel lead drugs is getting more and more popular and tends to substitute the classical approach.[51-53] There are many reasons that contributed to the preference of *in silico* design compared to the classical approach[53]: (i) The advancement of the computer science which leads to the construction of powerful and friendly used computers; (ii) the development of statistical packages that can utilize databases containing theoretical or experimental data which can be subjected to QSAR; (iii) the development of new techniques in the experimental procedures for characterizing proteins and biological targets (i.e. X-ray crystallography and NMR spectroscopy); (iv)

the increase in the knowledge of the molecular basis of drug action. Thanks to the advancements in enzymology, molecular biology, pharmacokinetics and pharmacodynamics; to mention only few representative fields that experienced huge developing steps; (v) the application of conformational search methodologies (e.g., MD and MC simulations); (vi) Adsorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) simulations have also contributed significantly in the drug development.

The most important characteristic of the rational drug design is to utilize in a positive way all known information of the system under study for developing a strategy for potential leads in drug discovery. This knowledge hopefully will lead to reduced human power cost, time saving and laboratory expenses. For the drug development new experimental and theoretical background is necessary. Usually, the first unsuccessful attempts serve as a feed back for development of new drugs.

Each specific system under study requires its own strategy in the computational drug design. Some strategies used in the current research activities for tackling problems with different information in the system will be given. In first system, there is no information available for the receptor binding site (system-A). Thus, no crystallographic data for the ligand bound to the receptor binding site is available in this system. In the second case, the X-ray structure of the receptor is not available, but homology models of the receptors have been performed using the available X-ray of a prototype receptor (system-B). Another case is when the X-ray structure of the receptor or ligand-bound receptor, associated with experimental ligand binding measurement tests, are available (system-C and system-D, respectively).

In order to develop more active novel compounds for system-A, strategy which is shown in Figure 4.1 can be used. Steroidal mustard esters for the antileukemic activity can be given as an example for the system-A.

CBs can be given as an example of system-B. Because of the absence of X-ray structure of CB receptors, led to their modeling based on the rhodopsin as well as β2-adrenergic receptors that constitute also members of GPCRs. The strategy followed

for designing novel leads was based on application of: (i) biophysical studies (ii) theoretical calculations combining with 2D NMR spectroscopy; (iii) *in silico* docking simulations; (iv) MD simulations at different environments; (v) 3D QSAR studies; (Figure 4.2).

Anti HIV-1 PR fullerene analogues design can be given as example for the system-C. In this case, the X-ray structure of the non-fullerene ligand bound HIV-1 PR receptor is known and biological ligand binding data for some analogues possessing different biological activity are available. The strategy applied in this case is shown in Figure 4.3.
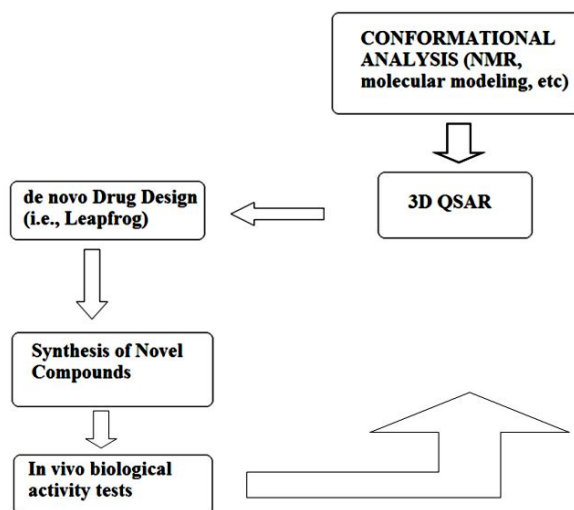
Figure 4.1 Strategy used to develop more potent bioactive molecules for molecules acting on a system-A type.
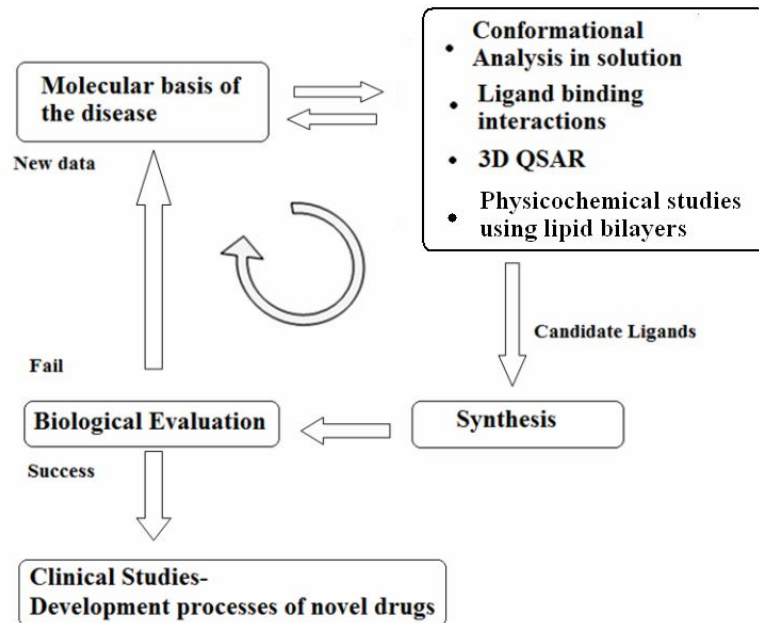
Figure 4.2 Strategy used for designing new leads in the system-B.
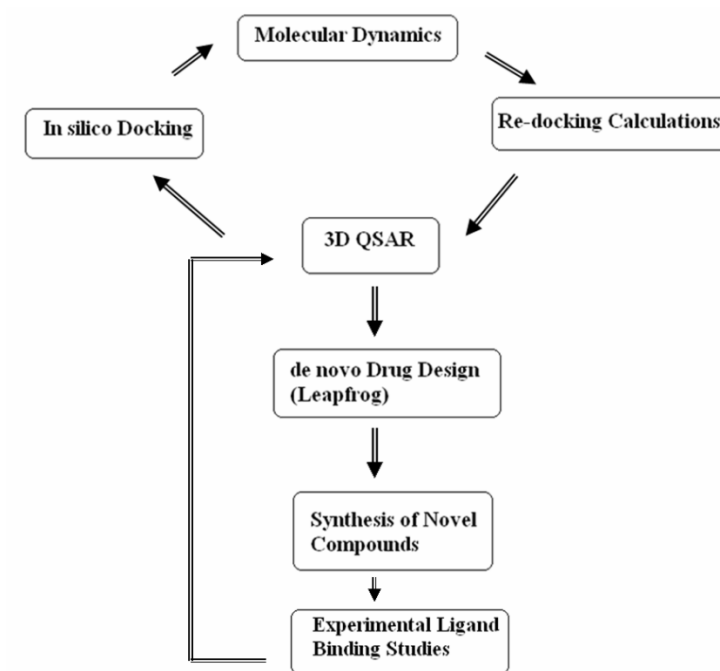
Figure 4.3 Strategy used for designing new leads in the system-C.

Renin inhibitors can be given as example for the system-D. The target of these compounds is the inhibition of the action of renin, an aspartic protease that evolves in the elevation of blood pressure through the renin angiotensin system. In this case, the X-ray structure of the renin with a variety of renin inhibitors is known. For this kind of systems, strategy shown in Figure 4.4 can be applied.
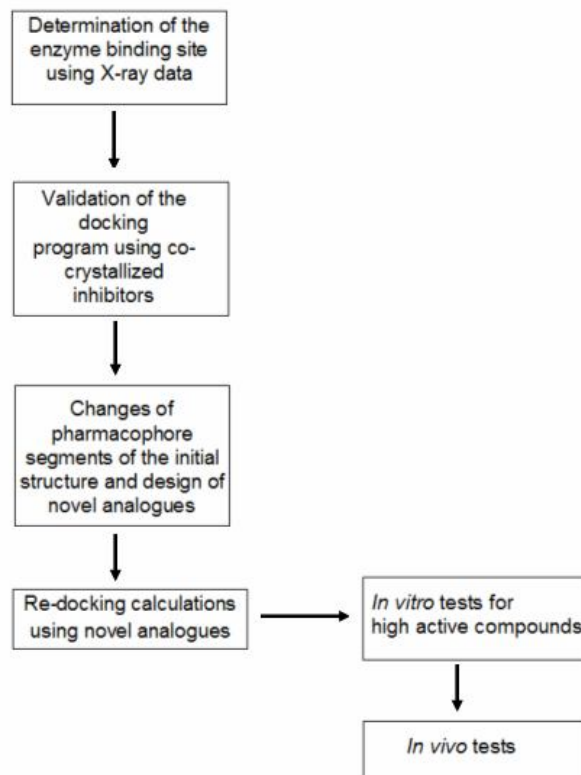
Figure 4.4 Strategy used for designing new leads in the system-D.

In the current thesis, examples of systems B and C have been studied extensively.

Given the atomic resolution structure of a target molecule, it should be possible to find novel molecules that bind to it, modulating its activity. Indeed, this is one of the main aims behind all receptor-based ligand design. Molecular docking is a key tool in the target-based approach. The promise of the docking method is that the structure of the receptor will provide a template for the design of novel ligands. The need to account for the dynamic behaviour of a target has long been recognized as a complicat-

ing factor in computational drug design field. Usually a single rigid receptor structure was chosen in order to reduce the computational cost. For example, if a large database of compounds is to be screened for the binding affinity against a target, then several conformers of each compound will be compared with each protein configuration. Although it is more accurate to use the above mentioned way, it is usually impractical, because of very high computational cost. In order to gain flexibility to receptor and ligand, one of the easiest ways is to combine MD simulations with docking calculations.

As it is mentioned, conformational analysis is a crucial step in computational drug design strategy. In the current thesis, the conformational analysis shown in Figure 4.5 was followed.
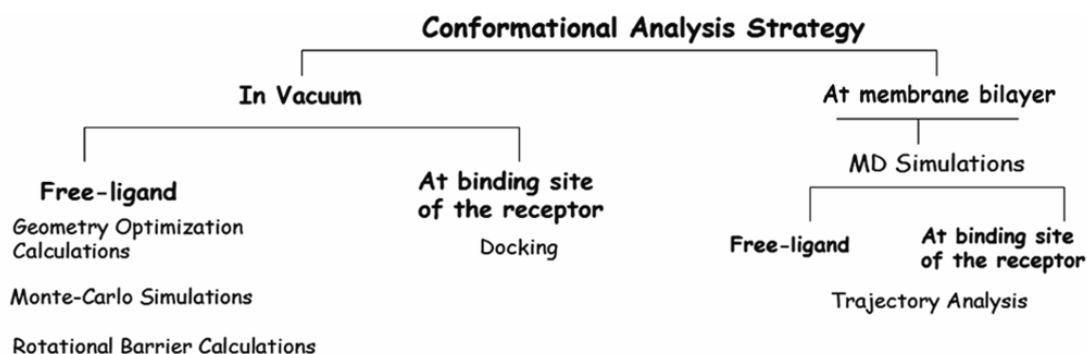
Figure 4.5 Performed conformational analysis strategy.

Chapter 5. Conformational Analysis, 3D QSAR, Homology Modeling, Molecular Docking and MD Simulations of CBs
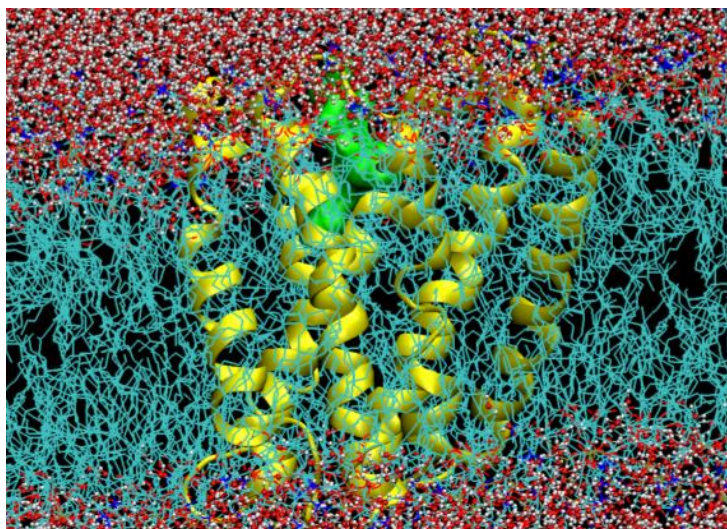
5.1 Introduction

Cannabis sativa L. is one of the oldest known medicinal plants and has been extensively used with respect to its psychotropic and pharmacological effects. Δ^9-tetrahydrocannabinol (Δ^9-THC), (Figure 5.1a) is the primary psychoactive constituent of cannabis and it was identified by Gaoni and Mechoulam in 1964.[54]

The pharmacological activity of CB ligands is mediated by two CB receptors: CB1[55-58] and CB2.[59] Both CB1 and CB2 receptors belong to the Class A, membrane-bound rhodopsin-like family of GPCR, possessing seven characteristic TM domains.[4-60,61] The CB1 receptor is abundant especially in the central nervous system (CNS)[62,63], peripheral tissues[64] and is assumed to be involved in the regulation of cognition, memory, motor activity, and the inhibition of transmitter release through its coupling to Ca^{2+} and K^+ channels. The CB2 receptor, on the other hand, is exclusively present in the tonsils and cells of immune system[65,66] such as B lymphocytes and macrophages. It is also found in the marginal zone of the spleen. The CB2 receptor is assumed to participate in the regulation of immune responses and inflammatory reactions and neuropatic pain. Pharmacological studies have shown that CBs possess many potential therapeutic applications including against cancer, AIDS, stroke, pain, obesity, cachexia and neuronal disorders such as multiple sclerosis, Huntington's chorea and Parkinson's disease, as well as reduction of blood ocular pressure in glaucomic patients.[5,6,67-74]

The recent studies showed that CB receptors work to block pain with a mechanism similar to the one which opiate receptors use when activated by the powerful painkilling drug morphine.[74] The CBs might provide an alternative to morphine, which can have serious side effects such as dependency, nausea and vomiting. [74]

CBs can be classified mainly into three categories: Natural (herbal) or classical CBs, endogenous CBs, and synthetic CBs. Natural CBs occur in the cannabis plant. THC, CBD (Figure 5.1b) and cannabinol (Figure 5.1c) are examples of natural CBs. Endogenous CBs are produced in the bodies of human and animals. An endogenous CB

ligand isolated from porcine brain by Mechoulam and Devane in 1992[75] and is identified as arachidonylethanolamide (anandamide), (Figure 5.1d). It binds to the CB1 and CB2 receptors with modest (K_i = 61 nM), and with low (K_i = 1930 nM) affinities, respectively.[75] It behaves as a partial agonist in the biochemical and pharmacological tests used to characterize CB activity.[67,76] 2-Arachidonyl glycerol (2-AG), (Figure 5.1e) is the second identified endogenous CB, and was isolated from brain and intestinal tissues by Sugiura and Mechoulam groups.[77,78] It was found to bind weakly to both of CB1 (K_i = 472 nM) and CB2 (K_i = 1400 nM) receptors. Prior to the discovery of CB receptors, a number of independent research laboratories and pharmaceutical industries developed a large number of synthetic CB ligands as pharmacological and biochemical probes for studying CB biology and also prototypes for developing new medications.[4,79,80] HU-210, CP55,940 and Nabilone are examples of such synthetic CB analogues (Figures 5.1f-h).[81] The discovery of endogenous ligands prompted further studies aimed at the elucidation of the chemical and pharmacological behavior of the CB1 and CB2 receptors and cannabinomimetic ligands. These studies pointed out that, in addition to classical CBs, other structurally different molecules may interact with the same receptors, inducing analogues responses.[68,82] Classical CBs possess two pharmacophores within the CB prototype that are important for cannabimimetic activity: a phenolic hydroxyl and a lipophilic side chain. The early SAR studies have been reviewed comprehensively by Thakur[67], by Khanolkar[6], by Razdan[83], and by Makriyannis and Rapaka[84]. Earlier literature reports[5,6,67,84,85] showed that the lipophilic alkyl side chain plays a crucial role in determining cannabimimetic activity and selectivity towards CB receptors, as well as pharmacological potency. The alkyl side chain fits into a hydrophobic pocket such that the chain is oriented nearly perpendicular to the aromatic ring A.[5,86-89] Analogues with alkyl side chains of less than five carbons have limited affinity for the CB1 receptor.[6,67] Extension of the five carbon chain by adding one or two carbons improves binding, while further extension is detrimental to binding due to steric hindrance.[5,6,67] Structural variations within this pharmacophore can result in analogues varying by up to three orders of magnitude in binding affinity for the CB receptor and in pharmacological potency. The structural modifications of the side chain produce high affinity ligands with either antagonist, partial agonist, or full agonist effects.[5]
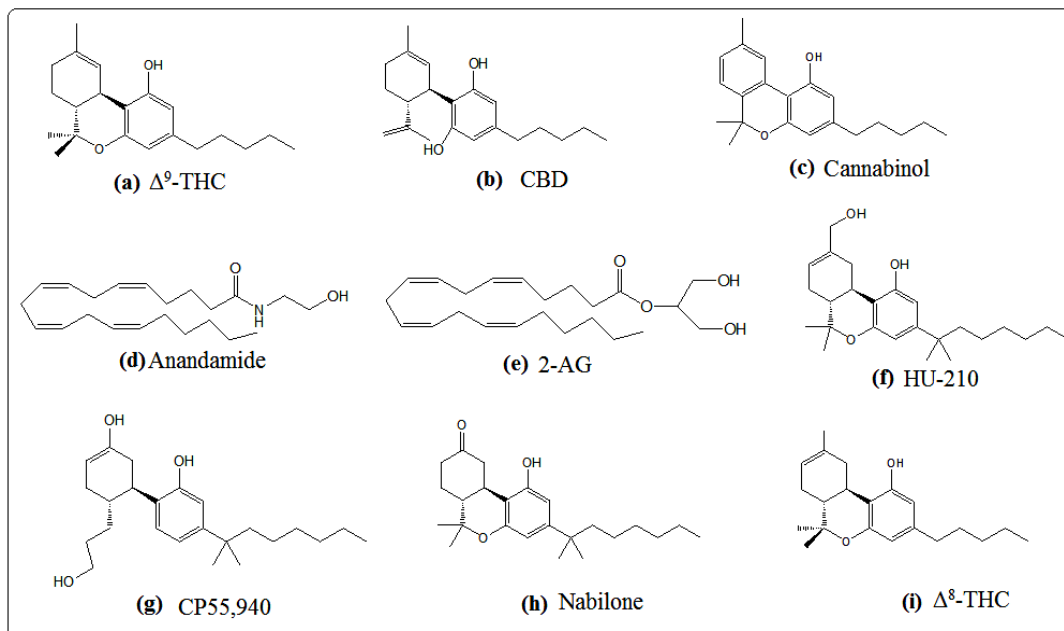
Figure 5.1 Chemical structures of **(a)** Δ^9-THC, **(b)** CBD, **(c)** Cannabinol, **(d)** Anandamide, **(e)** 2-AG, **(f)** HU-210, **(g)** CP55,940, **(h)** Nabilone, **(i)** Δ^8-THC.
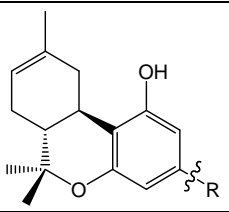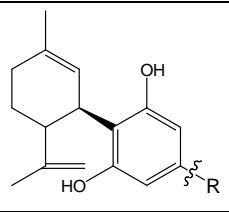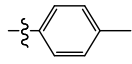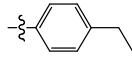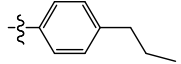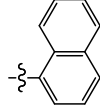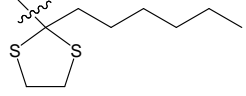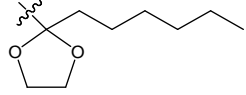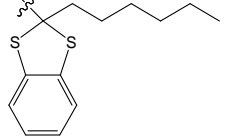
In order to improve the medicinal properties and eliminate or reduce untoward effects, medicinal chemists are designing, synthesizing and testing additional CB1 and CB2 ligands. One of main effort of our laboratory is to explore the pharmacophoric requirements of the alkyl side chain within the classical Δ^8-tetrahydrocannabinol (Δ^8-THC), (Figure 5.1i) and CBD (Figure 5.1.b) templates. Δ^8-THC has a very similar pharmacologic profile as Δ^9-THC, however it is chemically more stable. Several cannabinergic ligands possessing high affinities for both of CB receptors have been developed recently (Table 5.1). One of the most successful compounds that resulted from this work was the C1'-dithiolane analog AMG3; (**12** in Table 5.1) exerting binding affinity values of 0.32 and 0.52 nM for the CB1 and CB2 receptors, respectively.[85]

Hitherto, no direct observation of a CB ligand bound to a CB receptor using X-ray crystallography has been reported.[90] Therefore, active sites of these receptors have been postulated from many approaches, such as receptor binding analyses of a variety of CB derivatives using wild type and mutated receptor systems, molecular modeling analysis and 3D QSAR studies.[90-96] 3D QSAR/CoMFA and CoMSIA techniques[9,10] have been successfully used previously in 3D QSAR studies of CBs and other

ligands.[72,90,94,97-103] The present study uses 3D QSAR/CoMFA and CoMSIA analyses on novel CB analogues (Table 5.1)[85,104-107] with a wide variation of biological activity (4000, and 1200-fold variances in bioactivity for the CB1 and CB2 receptors, respectively). Thus studied compounds are characterized by subtle structural variations and a wide range of biological activities which constitute an ideal base for 3D QSAR studies.

The selection of the lowest energy conformation of the bioactive conformation of the template molecule and the superimposition of all molecules on template compound are the two most critical steps in the 3D QSAR studies, especially for CoMFA and CoMSIA methodologies. These steps not only affect the output of the analysis, but they also contribute to the design of novel molecules. Among the synthesized analogues shown in Table 5.1, AMG3 was selected as a template, because: (i) it has the highest binding affinity at the CB1 receptor and forth highest binding affinity at the CB2 receptor, in the dataset; (ii) preliminary results of the low energy conformers of AMG3 have been reported using a combination of 2D NMR spectroscopy and molecular modeling techniques.[89]

The knowledge of the receptor structure is not a prerequisite for 3D QSAR analysis, however, the availability of its crystal structure or 3D model facilitates the structure alignment, and can provide statistically more reliable models.[108,109] 3D homology models of the CB1 and CB2 receptors were constructed by Shim et al.,[8] Salo et al.[34] and Tuccinardi et al.[33] with a molecular modeling procedure using the X-ray structure of bovine rhodopsin[110] as the initial template and taken into account the available site-directed mutagenesis data. These groups studied different CB classes, however, they found them to interact at an active site with similar homologies. Therefore, in addition to the conformational analysis studies of template compound *in solution*, conformational analysis studies have been also carried out at the binding sites of the receptors. Although, the more recent 3D receptor models were used for the molecular docking studies, the active site residues are determined considering both three models mentioned above.

Compound No.	R	K_i for CB1 (nM)	K_i for CB2 (nM)	Compound No.	Ki for CB1 (nM)	Ki for CB2 (nM)
1		95.49	71.81	2	638.1	374.4
3		119.0	51.70			
4		57.77	107.80			
5		11.73	9.39	6	753.5	221.6
7		27.90	25.20	8	255.0	105.0
9		8.26	3.86	10	319.0	110.7
11		168.0	103.0			
12		0.32	0.52	13	136.0	50.40
14		0.52	0.22			
15		56.90	257.0			

No.	Structure			No.		
16		1.80	3.60			
17		32.30	19.70			
18		0.45	1.92			
19		47.60	39.30	**20**	1265.0	230.0
21		22.00	-			
22		0.83	0.49			
23		0.44	0.86	**24**	58.68	99.23
25		1.27	0.29	**26**	666.4	32.87
27		0.71	1.03	**28**	189.0	63.30
29		21.70	83.70			
30		2.17	3.30			

Table 5.1 Molecular structures and binding affinity K_i values of CB analogues used as the training set to construct 3D QSAR/CoMFA and CoMSIA models.[85,104-107]

In order to include solvent and temperature effects to the conformational analysis of the template compound more sophisticated calculations have been performed both *in solution* and at the membrane associated receptors using MD simulations. The obtained conformations of template compound have been compared and for each conformer 3D QSAR models were created and statistical analysis calculations were performed. The effect of the increase of biological resemblance of a system to the derived statistical results from constructed QSAR models was investigated.

Reported low energy conformations of Δ^8-THC analogues using NMR spectroscopy and molecular modeling studies *in solution* differed.[89,90,102,111] The derived conformations differed in the alkyl side chain or in the ABC tricyclic segments (e.g., conformers **A**-**D** in Figure 5.2). For example, in both conformers **A** and **B**, proposed for classical Δ^8-THC analogues, and shown in the Figure 5.2, the alkyl chain adopts an orthogonal orientation relative to the horizontal plane of ring A, however B and C rings have different geometries (e.g., ring B has half chair-like and boat-like forms, in conformers **A** and **B**, respectively). In conformer **C**, the alkyl chain has been extended away from the ABC tricyclic segment while in conformer **D** the alkyl chain has been wrapped towards the tricyclic part (Figure 5.2).

Conformational studies of AMG3 using a combination of 1D and 2D NMR spectroscopies as well as molecular modeling techniques, showed that, the alkyl side chain adopts a perpendicular orientation relative to the horizontal plane of ring A of AMG3.[89] The conformation of the flexible l',l' dimethylheptyl side chain was also analyzed using a combination of theoretical studies and NMR experiments for classical CB (-)-9-nor-9β-hydroxy(dimethylheptyl)-hexahydrocannabinol and nonclassical CBs CP47,497, CP55, 244 and CP55, 940 by Xie *et al.* [86,112,113]. Results showed that the alkyl side chain is almost perpendicular to the horizontal plane of the ring A.
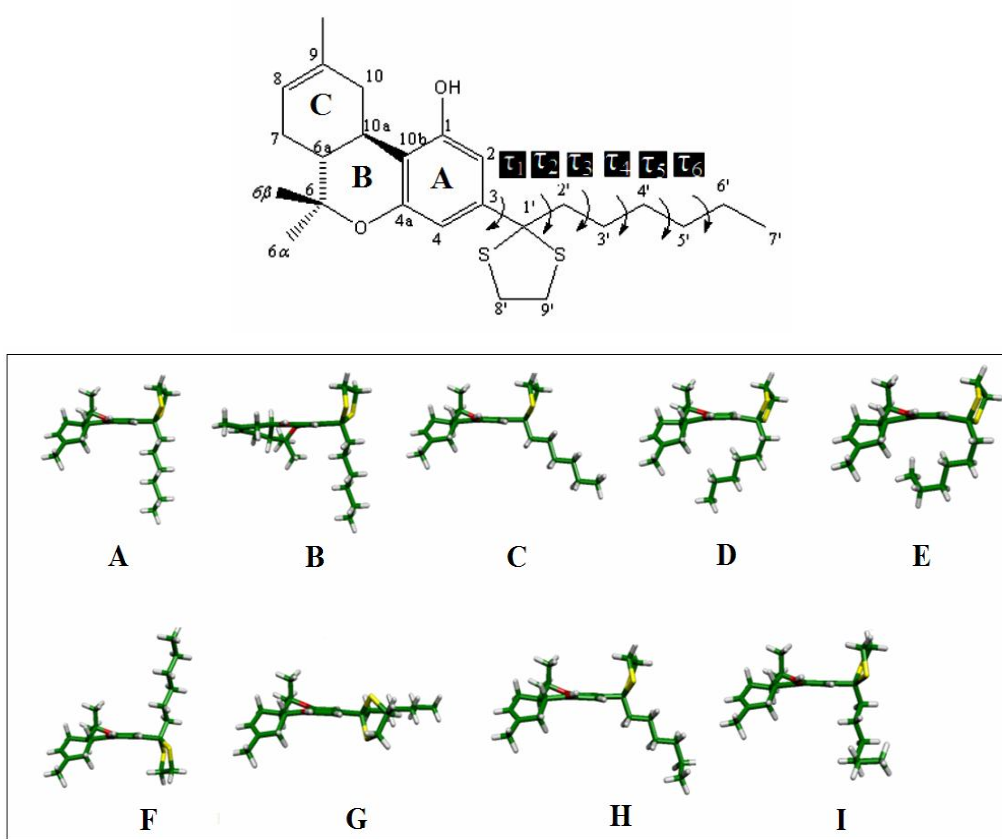
Figure 5.2 (top) Molecular structure of AMG3 and; **(bottom)** its derived 3D low energy conformers **A-I**. Dihedral angles of the alkyl side chain are assigned on the top structure (τ_1, C2-C3-C1'-C2'; τ_2, C3-C1'-C2'-C3'; τ_3, C1'-C2'-C3'-C4'; τ_4, C2'-C3'-C4'-C5'; τ_5, C3'-C4'-C5'-C6'; τ_6, C4'-C5'-C6'-C7').

CBs are predicted to exert their biological action in the TM3-TM7 helices of GPCRs.[5,114] Reported experimental results suggest that biological activity of CBs can be related with two sequential criteria: (i) 'proper' topology and orientation of the drug in the membrane bilayer; (ii) diffusion and 'appropriate' fit of the drug in the receptor.[115] CBs are lipophilic molecules and are considered to first interact with the lipid microenvironment that surrounds the membrane-associated protein and then diffuse laterally at the active site of the receptor.[116] Therefore, it is important to understand the conformational properties of a drug molecule into the lipid microenvironment. Membranes do not lend themselves in a detailed analysis of their structural and dynamical properties by means of a single physicochemical method because of their

complexity and instability.[115] Thus, it is preferable to combine several experimental and/or computational methods seeking to obtain molecular information on the interactions of drugs with membranes.[115] Biophysical studies using different techniques (solid-state NMR spectroscopy, X-ray diffraction, Raman spectroscopy, IR spectroscopy, differential scanning calorimetry (DSC), etc.) in combination with molecular modeling studies assist in the determination of drug-membrane interactions and the role of membrane in the putative bioactive conformation of the drug molecule.

The effect of the cannabinomimetic drug AMG3 on the thermotropic and structural properties of dipalmitoyl-*sn*-glycero-3-phosphorylcholine (DPPC) liposomes have been studied by X-ray diffraction and DSC methodologies by Mavromoustakos *et al.*[117] AMG3 was found to efficiently fluidize domains of the lipids in the L_{β}' gel phase and to perturb the regular multibilayer lattice. In the liquid crystalline L_{α} phase, AMG3 was also found to cause irregularities in packing, suggesting that the drug induces local curvature. Biophysical studies by Makriyannis *et al.*[115] have also provided detailed information for the topography, the stereochemistry, and the dynamic properties of the CB ligand-membrane interactions by applying neutron diffraction, solid state NMR, DSC and small angle X-ray spectroscopy of Δ^8-THC.[118-120] In these studies, THC assumes an 'awkward' orientation in the bilayer with the long axis of its tricyclic system being perpendicular to the bilayer chains, while its aliphatic side chain orients parallel to the chains of the membrane phospholipids.

Generated 3D QSAR models were used in *de novo* drug design program and allowed us to propose novel ligands with higher predicted binding affinity values. In order to determine the linear correlation coefficients between actual (measured) versus calculated binding affinities, PLS statistical analyses of the data were used. CoMFA and CoMSIA contour plots were used to explain different structural requirements for CB binding to the CB1 and CB2 receptors. Derived contour plots can be used as pilot models for testing the designed novel analogues before their synthesis.

5.2 Results and Discussion

5.2.1 Conformational Analysis of AMG3 in Vacuum and in Lipid Bilayer Environments (without Receptor)

5.2.1.1 Selection of Low Energy Conformers of AMG3 Using MC Studies

Low energy conformers of AMG3 are derived using MC conformational search analysis. The application of MC analysis, which allows full angular window specification and random change of dihedral angles of rotatable bonds, derived 1000 conformers of AMG3. These conformers were clustered into eight different groups based on the dihedral angle criterion. The lowest energy conformers from each cluster have been selected for further analysis. The number of conformers for each of the eight clusters has been shown in the Table 5.2. Cluster analysis showed that conformer **A** (it has a perpendicular orientation of alkyl side chain relative to long axis of the ABC tricyclic segment) and wrapped conformer **D** (wrapped conformations are defined as those conformations adopting *gauche* τ_3 and τ_4 dihedral angles at the alkyl chain) have highest and smallest group member populations, respectively. Among eight lowest energy conformers, three conformers (**A**, **C**, and **D**) of AMG3 in Figure 5.2 are identical with those reported conformations[89,90,102,111] using experimental results and/or molecular modeling techniques, while other five conformations (conformers **E-I** of AMG3 in Figure 5.2) differed. Conformer **B** was not obtained by MC cluster analysis.

5.2.1.2 Geometry Optimization Calculations

MM geometry optimization has been applied using Sybyl molecular modeling package.[27] According to MM calculations, conformer **B** has significantly high total energy ($\sim$8.8 kcal.mol^{-1}) due to its high torsional energy, while other conformers have a similar energy plateau within the range of $\sim$2.3 kcal.mol^{-1} (Table 5.2). The high energy conformer **B** of AMG3 has been transformed to conformer **A** at the *ab initio* B3LYP/6-31G*[121,122] level optimizations. QM calculations show that all low energy conformers are almost isoenergetic (maximum total energy differences between the

conformers is ~2.5 kcal.mol^{-1}, Table 5.2). In order to examine the solvent effect over geometrical properties of conformers, the vacuum medium has been modified to amphiphilic environment. The dielectric constant ε was set to 45, in order to simulate an amphiphilic environment, which mimics physicological conditions and therefore it is

Conformer	Number of conformers in each obtained cluster by MC analysis	MM Relative Energy (kcal.mol^{-1})					QM Geometry Optimization (B3LYP/6-31G*), Relative Energy (kcal.mol^{-1})
		Torsional Energy	VDW Energy	Electrostat. Energy	Other Contributions[a] to MM Energy	Total Energy	
A	188	5.520	-3.821	-5.820	6.621	2.499	0.10
B	-	9.147	-2.674	-5.936	8.211	8.748	0.10
C	147	5.550	-3.055	-5.817	6.688	3.366	1.26
D	42	6.412	-5.857	-5.696	7.647	2.508	2.56
E	102	6.493	-6.877	-5.676	8.157	2.097	0.00
F	156	5.528	-3.877	-5.782	6.558	2.427	0.00
G	92	6.180	-2.894	-5.693	6.773	4.366	1.57
H	134	5.699	-3.013	-5.788	7.120	4.018	2.03
I	139	5.783	-4.094	-5.726	7.507	3.470	1.48
Average[b]	111	6.257	-4.018	-5.770	7.254	3.723	0.93

[a] Other contributions include bond stretching energy, angle bending energy, out of plane bending energy. [b] Average results do not include conformer **B** for QM relative energies, because conformer **B** is transformed to conformer **A** when QM geometry optimization is applied.

Table 5.2 Number of conformers of AMG3 in each derived cluster by MC analysis and comparison of the total energies of conformers using MM and QM methods.

appropriate for investigating biological structures.[123] It is observed that the effect of continuum model (ε = 45) to the analyzed conformers compared to gas phase is very limited as it is depicted by the small RMSD value between the conformers in gas phase and in continuum model. Dihedral angles of the alkyl side chain segment of all conformers applying full geometry optimization with MM and QM methods in gas phase and in continuum model are presented in Table 5.3.

5.2.1.3 Rotational Energy Barrier Calculations

In order to characterize the conformational flexibility properties of AMG3, rotational energy barriers were estimated using torsional grid scan analysis with semi empirical method PM3.[124] Six rotatable torsional angles of AMG3, shown in Figure 5.2 and defined in its figure legend, were analyzed. The analysis is initiated with τ_1 and the energetically lowest structure (optimal dihedral angle) is used for the next torsional angle analysis. Rotation around dihedral angles τ_4-τ_6 showed similar energy profiles and rotational energy barriers are found to be ~4 kcal.mol^{-1}. Their optimal dihedral angle was found to be ~180° and the local minima were observed at ~60° and ~300°. Rotation around dihedral angles τ_1 and τ_2 showed a rather more complex energy profiles due to the presence of the tricyclic ring segment. Rotation around dihedral angle τ_3 showed the largest rotational energy barrier (~6 kcal.mol^{-1}) and its optimal dihedral angle value was found to be ~180° (Figure 5.3).

5.2.1.4 MD Simulations of Conformers in Lipid Bilayer

Computer simulations in general and MD simulations in particular, are of increasing importance in revealing details of molecular motions as well as structural and microscopic properties of the solution, which are difficult to measure experimentally.[125] Heating increases the kinetic energy of the system which after equilibration at the given temperature overcomes any energy barriers close to the initial energy minimum. In order to examine the environmental effects over the structures, MD simulations were performed for all examined conformers with the Gromacs 3.3.1 software package[36] in DPPC membrane bilayer environment. MD simulations were used in order to (i) further study the conformational space of AMG3 and (ii) explore the possibility of interconvertion between conformers in amphiphilic environments.

MD in DPPC bilayer. Initial positions of the conformers in DPPC bilayer have been constructed according to experimental findings.[115] Alkyl chain of conformers have been inserted through the alkyl chains of the lipid (parallel orientation with lipid bilayer chains).[115] The tricyclic ABC segment of the AMG3 was localized close to the

head groups of DPPC bilayers and orienting their long axis perpendicular to the bilayer plane (Figure 5.4). MD simulations analysis have shown that last four defined torsional angles (τ_3-τ_6) at the alkyl side chain of AMG3 are *all trans* with some perturbations around these angles for all of the analyzed conformers. Therefore, wrapped conformers (e.g., **D** and **E**) and conformers which have *gauche* dihedral angles within τ_3-τ_6 (e.g., **H** and **I**) do not keep their initial values and turn to *trans* dihedral angles. (Torsional angle screening throughout the simulations for wrapped conformers **D** and **E** have been shown in Figure A4, Appendix). The dihedral angle τ_1 shows more resistance to be transformed to another torsional angle through simulations. The dihedral angle τ_2 is the most flexible torsional angle in the alkyl chain and adopts *gauche*± and *trans* dihedral angles, thus it determines the characteristics of the conformers. MD simulations of AMG3 in lipid bilayer environment produced three more low energy conformers in addition to previous ones. These conformers are called **J**, **K** and **L** and are shown in Figure 5.5 (their adopting dihedral angles of alkyl chain are presented in their corresponding figure legend). Trajectory analysis results showed that stable conformers have *gauche*± for τ_1, *gauche*± and *trans* for τ_2 and *all trans* conformations for τ_3-τ_6 dihedral angles throughout the simulations. These stable conformers are **A**, **C**, **F**, **J**, **K** and **L**, and are considered for further investigations.

These stable conformers form two favored plane angles which are are ~90° and ~140°. The first plane is defined by ring A (Figure 5.2), and the second plane is defined by the alkyl chain (Figure 5.6). Formed plane angles of favored conformations after MM and QM geometry optimizations and their relative energies are compared in the Table 5.4. Relative energy differences are used for clarity and are defined on lowest total energy of conformers calculated by MM and QM methods. The lowest energy conformers (**J** for QM and **K** for MM methods) are arbitrarily assigned with zero potential energy and relative energies of other conformers are calculated. Both MM and QM calculations showed that, all favored conformations derived from MD simulateons in lipid bilayer have similar relative energies; however conformers which form perpendicular plane angle between tricyclic ring and flexible segments have slightly lower energy (~1 kcal/mol) than corresponding plane angle of ~140°.

CONF.	τ₁ (degree)				τ₂ (degree)				τ₃ (degree)				τ₄ (degree)				τ₅ (degree)				τ₆ (degree)				RMSD	
	MM ε=1	MM ε=45	QM ε=1	QM ε=45	MM ε=1	MM ε=45	QM ε=1	QM ε=45	MM ε=1	MM ε=45	QM ε=1	QM ε=45	MM ε=1	MM ε=45	QM ε=1	QM ε=45	MM ε=1	MM ε=45	QM ε=1	QM ε=45	MM ε=1	MM ε=45	QM ε=1	QM ε=45	MM	QM
A	73.8	73.2	63.3	65.3	57.8	57.8	61.1	61.3	178.9	179.0	181.1	181.7	179.7	179.7	180.5	180.2	180.0	180.0	180.2	180.8	180.0	180.0	180.1	180.0	0.25	0.75
B	78.1	78.1	63.3	65.3	56.7	56.8	61.1	61.3	178.5	178.5	181.1	181.7	179.3	179.3	180.5	180.2	180.1	180.1	180.2	180.8	180.0	180.0	180.1	180.0	0.01	0.75
C	82.0	82.0	81.5	77.2	181.0	181.0	184.0	176.5	179.8	179.8	185.3	182.2	180.0	180.0	180.6	180.6	180.0	180.0	180.5	179.6	180.0	180.0	179.9	180.0	0.00	3.86
D	55.4	54.8	62.6	63.1	65.0	65.2	65.5	65.3	293.9	294.3	279.2	278.9	191.4	191.3	185.0	185.0	181.9	181.8	179.4	179.4	180.3	180.2	179.4	179.5	0.13	0.25
E	55.7	55.2	61.4	59.0	68.9	69.1	66.7	68.6	300.8	301.2	281.5	285.5	191.0	190.9	186.8	186.2	176.5	176.4	176.5	177.7	63.9	63.7	65.9	66.0	0.12	2.08
F	258.0	257.6	250.5	244.1	57.9	58.0	60.9	60.8	178.6	178.6	180.8	180.7	179.7	179.7	180.1	180.2	180.0	180.0	180.1	180.0	180.0	180.0	180.0	180.0	0.17	2.61
G	187.4	187.0	187.2	187.1	176.9	177.0	175.6	175.6	176.6	176.5	177.8	177.8	179.5	179.5	179.6	179.5	179.9	179.9	179.5	179.5	180.0	180.0	180.1	180.1	0.17	0.06
H	81.9	82.2	86.9	83.0	181.1	181.1	182.7	184.1	179.9	179.9	184.8	185.5	180.5	180.6	180.7	181.0	185.7	185.7	183.2	183.5	296.3	296.4	294.3	294.4	0.14	1.71
I	103.1	103.7	116.1	128.3	302.9	302.9	298.0	297.3	182.6	182.6	178.9	178.0	186.5	186.6	184.2	181.8	302.0	302.1	296.6	295.4	302.3	302.3	297.2	297.3	0.25	5.11

Table 5.3 Values of dihedral angles corresponding to the alkyl chain part for low energy conformers of AMG3 derived by applying full geometry optimization with MM (left) and QM (right) with B3LYP/6-31G* level of theory. After optimization with QM, conformer **B** was transformed to conformer **A**, therefore they have identical dihedral angles for the corresponding method.

<table>
<tr><td>(i)</td><td>(ii)</td><td>(iii)</td></tr>
</table>

<table>
<tr><td>(iv)</td><td>(v)</td><td>(vi)</td></tr>
</table>
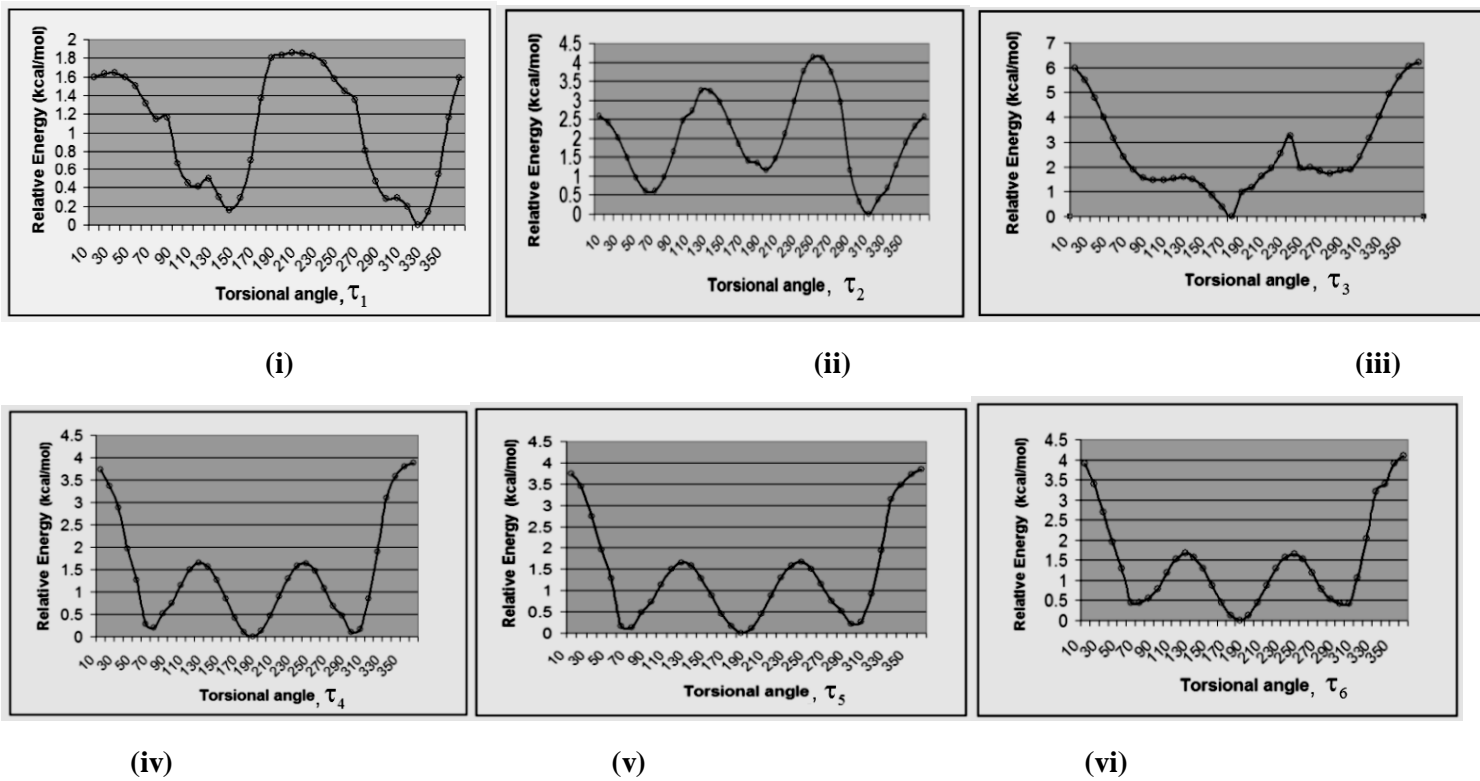
Figure 5.3 Rotational energy barriers of AMG3 after applying grid scan analysis with consecutive optimization of dihedral angles (i) τ_1, (ii) τ_2, (iii) τ_3, (iv) τ_4, (v) τ_5 and (vi) τ_6. In figure, relative energies (differences of initial and final values of total energies with rotation) have been used instead of total energies for clarity.
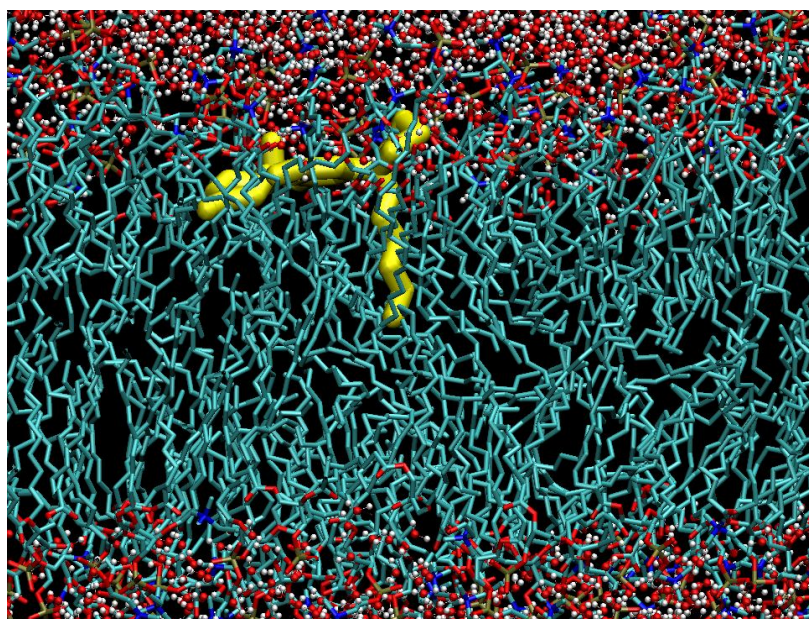
Figure 5.4 MD simulations of AMG3 conformers in the lipid bilayer environment which consists of DPPC bilayers and water molecules.
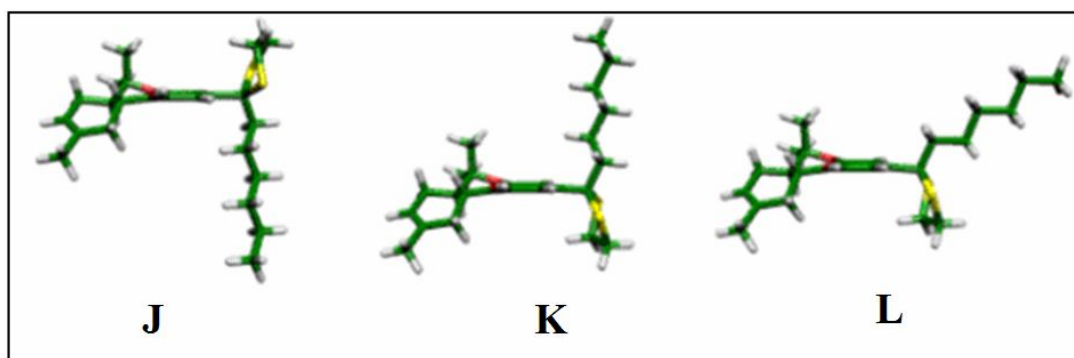
Figure 5.5 MD simulations in lipid bilayer produced three additional low energy conformations of AMG3 (named **J**, **K** and **L**) in addition to the presented conformers. Their dihedral angles of alkyl chain τ_1-τ_6 were derived after applying MM and QM geometry optimizations in gas phase. These values are 103.9^o, 302.1^o, 181.2^o, 180.2^o, 180.0^o and 180.0^o, respectively for conformer **J**; 282.8^o, 302.2^o, 181.4^o, 180.3^o, 180.0^o, and 180.0^o, respectively for conformer **K**; and 274.6^o, 179.3^o, 180.2^o, 180.0^o, 180.0^o and 180.0^o, respectively for conformer **L**.
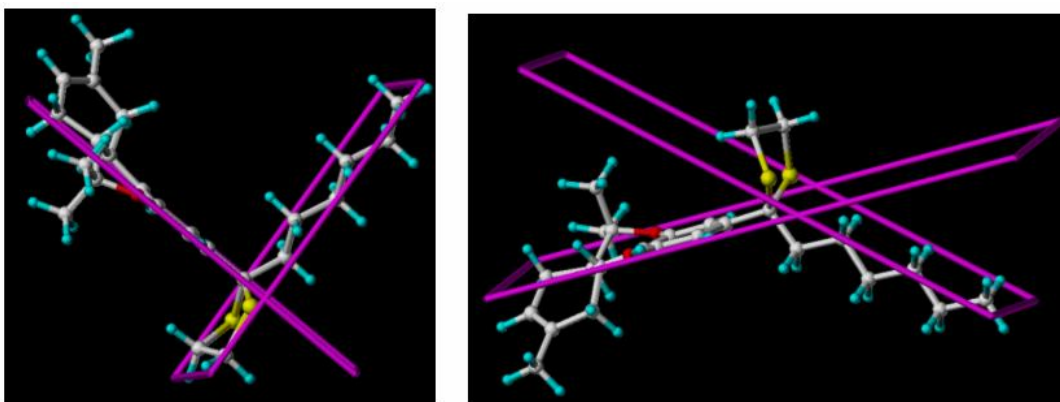
Figure 5.6 Two favored plane angles between rigid and flexible segments of AMG3; ~90° (on the left) and ~140° (on the right) derived using MD simulations in lipid bilayer environment.

Conformer	MM-Plane Angle (degree)	QM-Plane Angle (degree)	MM-Relative Energy (kcal/mol)	QM-Relative Energy (kcal/mol)
A	87.0	89.0	0.16	0.16
J	87.5	87.5	0.23	0.00
F	88.8	86.6	0.09	0.15
K	89.4	91.4	0.00	0.60
C	137.9	135.5	1.03	1.40
L	141.4	137.9	0.81	1.30

Table 5.4 Plane angles between rigid and flexible segments and relative energies of favored conformations of AMG3 in lipid bilayer. In table, relative energies (differences of total energies based on lowest total energy of conformers calculated by MM and QM methods; the lowest energy conformer of AMG3 with MM and QM methods are found to be **K** and **J**, respectively) have been used instead total energies for clarity.

5.2.1.5 3D QSAR/CoMFA Results

The existing experimental findings combined with our molecular modeling studies assisted to obtain the template conformation of AMG3 for the constructions of 3D QSAR models.

Both MM and QM calculations resulted to favored conformers in lipid bilayer environment which perpendicular plane angle exists between ABC tricyclic ring and flexible alkyl chain segments. Initially, these conformers (**A**, **J**, **F** and **K**) have been used as template conformer and also compounds in the data base have been adopted with corresponding template compound conformation, subsequently statistical tests have been performed. The template conformer which showed optimum statistical results (conformer **A**) has been used for further analysis.

Several variations in the alignment schemes by superimposing the similar pharmacophoric features were considered. C_1, C_2, C_3, C_4, C_{4a}, C_{6a}, C_7, C_{10}, C_{10a}, C_{10b} and the oxygen atoms in the template ligand AMG3 (Figure 5.2) were selected for the structural superimposition processes. The alignment of the molecules was based on atom-by-atom superimposition of selected atoms, which is common in all compounds. The criteria applied for the selection were: (i) overlap of the putative biologically relevant pharmacophore groups (with minimum RMS) and (ii) form of statistically significant 3D QSAR/CoMFA and CoMSIA models. In order to build 3D QSAR/CoMFA and CoMSIA models for the binding affinity at the CB1 and CB2 receptors, a set of 30 CB analogues for the CB1 receptor and 29 CB analogues for the CB2 receptor were subjected to the cross-validated PLS analyses (Table 5.1). Figure 5.7 illustrates the superimposition of CB analogues used as the training set to construct CoMFA and CoMSIA models.
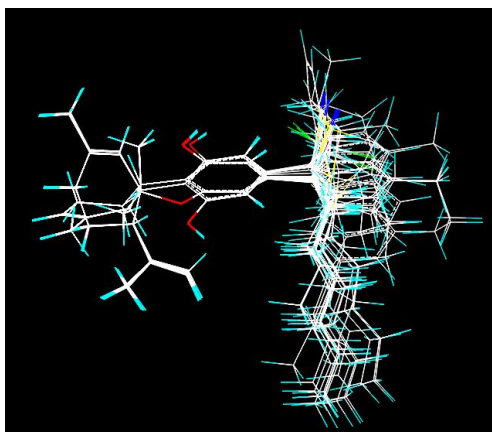
Figure 5.7 Structural alignments of the CB derivatives in the training set for constructing 3D QSAR/CoMFA and CoMSIA pharmacophore models.

Table 5.5 shows the derived statistical results using CoMFA technique for CB1 and CB2 pharmacophore models. The CoMFA study based on the selected lowest energy conformer of template ligand which was determined by a combination of experimental and molecular modeling studies, gave r_{cv}^2 values of 0.784 and 0.572 for CB1 and CB2 receptors, respectively. The non-cross-validated PLS analysis yielded an r^2 of 0.981 and 0.972; and the estimated standard errors were 0.173 and 0.187 for CB1 and CB2 receptors, respectively (Table 5.5). Therefore, the CoMFA-generated 3D QSAR models for the binding affinities of CB analogues at CB1 and CB2 receptors have a very good cross-validated correlation. Table 6 shows the relationship between the experimental and CoMFA-calculated pK_i (-$logK_i$) values of the non-cross-validated analyses for CB1 and CB2 receptors. Linearity of the plots at Figure 5.8 shows very good correlations between observed and predicted affinities for CoMFA models which developed in study for the binding affinities of CB ligands for the CB1 and CB2 receptors.

	CoMFA/CB1	CoMFA/CB2
Number of compounds in training set	30	29
r_{cv}^2	0.784	0.572
r^2	0.981	0.972
Standard error of estimate	0.173	0.187
F	197.531	127.260
Relative contributions of steric/electrosatic fields	0.640:0.360	0.632:0.368
Number of optimal components	6	6

Table 5.5 Statistical results obtained by 3D QSAR/CoMFA models for CB1 and CB2.

The contour maps were used to create a 'negative' matrix in the place of the unknown active site and variations of the used ligands can be generated as long as they fit better into the 'imaginary' active site. Figure 5.9 shows the steric-electrostatic contour maps of the CoMFA models for CB1 and CB2 receptors. The individual contributions from the steric and electrostatic favored and disfavored levels are fixed at 80% and 20%, respectively. The

CoMFA contours of the steric maps are shown in yellow and green colors, and those of the electrostatic contour maps are shown in red and blue colors. Greater values of "bioactive-measurement" are collected with: more bulky groups near the green colored contours; less bulky groups near the yellow colored contours; more positive charge near the blue colored contours, and more negative charge near the red colored contours.

Compound	CB1 CoMFA model		CB2 CoMFA model	
	pK_i(observed)	pK_i(predicted)	pK_i(observed)	pK_i(predicted)
1	7.02	7.16	7.14	7.25
2	6.20	6.18	6.43	6.50
3	6.92	7.09	7.29	7.19
4	7.24	7.13	6.97	7.13
5	7.93	7.86	8.03	7.98
6	6.12	6.20	6.65	6.64
7	7.55	7.69	7.60	7.87
8	6.59	6.63	6.98	7.03
9	8.08	8.09	8.41	8.27
10	6.50	6.56	6.96	7.03
11	6.77	6.66	6.99	6.95
12	9.49	9.34	9.28	8.96
13	6.87	6.89	7.30	7.39
14	9.28	9.40	9.66	9.68
15	7.24	7.17	6.59	6.62
16	8.74	8.74	8.44	8.44
17	7.49	7.61	7.71	7.68
18	9.35	9.50	8.72	9.05
19	7.32	7.33	7.41	7.51
20	5.90	5.80	6.64	6.43
21	7.66	7.73	-	-
22	9.08	8.99	9.31	9.04
23	9.36	9.13	9.07	9.12
24	7.23	6.88	7.00	7.16
25	8.90	9.19	9.54	9.35
26	6.18	6.53	7.48	7.19
27	9.15	9.11	8.99	9.29
28	6.72	6.58	7.20	7.26
29	7.66	7.52	7.08	6.93
30	8.66	8.54	8.48	8.39

Table 5.6 Observed and predicted pK_i (by CoMFA models) values at the training set for CB1 and CB2 receptors.
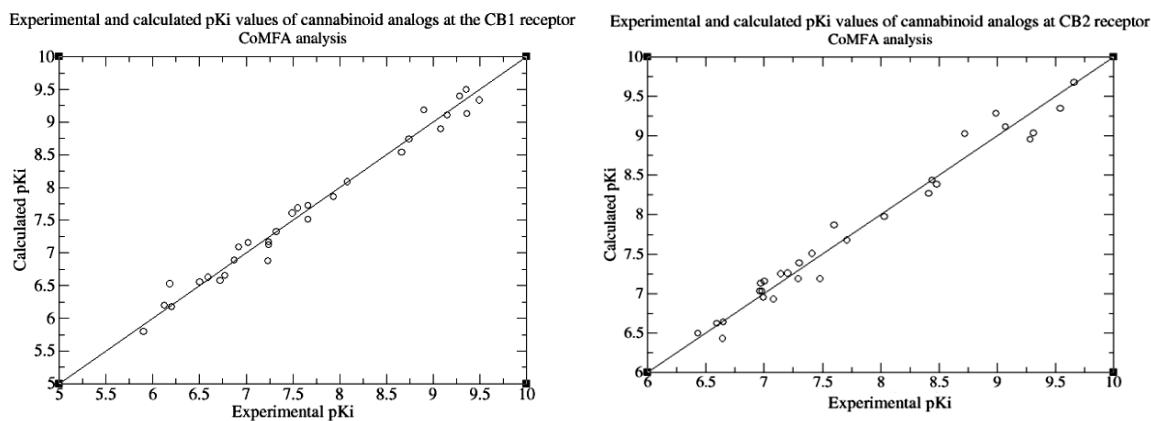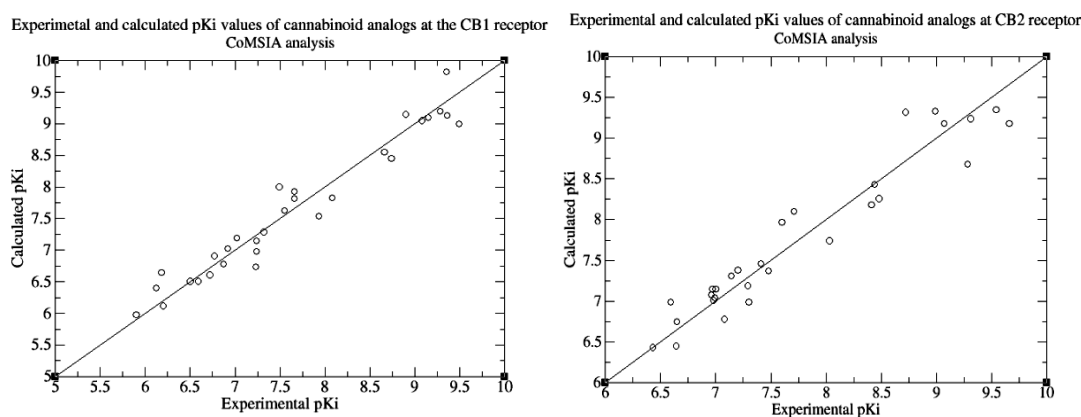
Figure 5.8 Plots of corresponding CoMFA-predicted and experimental values of binding affinity (given as pK$_i$) of CB analogues in the training set at the CB1 (on the left) and CB2 (on the right) receptors, respectively.

5.2.1.6 3D QSAR/CoMSIA Results

Several combinations of stereoelectronic fields (steric, electrostatic, H-bond donor, H-bond acceptor and hydrophobicity fields) of 3D QSAR/CoMSIA models were obtained from the compounds in the data set. The optimal 3D QSAR/CoMSIA model was derived by the combination of steric and electrostatic potential fields. This model based on the selected lowest energy conformer of template ligand AMG3, gave r_{cv}^2 values of 0.746 and 0.625 for the CB1 and CB2 receptors, respectively (Table 5.7). The non-cross-validated PLS analysis yielded an r^2 of 0.944 and 0.912, and the estimated standard errors were 0.296 and 0.324 for CB1 and CB2 receptors, respectively (Table 5.7). Relationship between the CoMSIA-predicted and experimental pK$_i$ values of the non-cross-validated analyses for CB1 and CB2 receptors have been shown in Table 5.8 and Figure 5.10.

Figure 5.11 shows the steric-electrostatic contour maps of the CoMSIA models for the CB1 and CB2 receptors. Since the CB analogues used in the training set differ mainly in the C1' position and the tricyclic part of Δ^8-THC or the CBD skeleton, the contour plots place more emphasis to these regions.
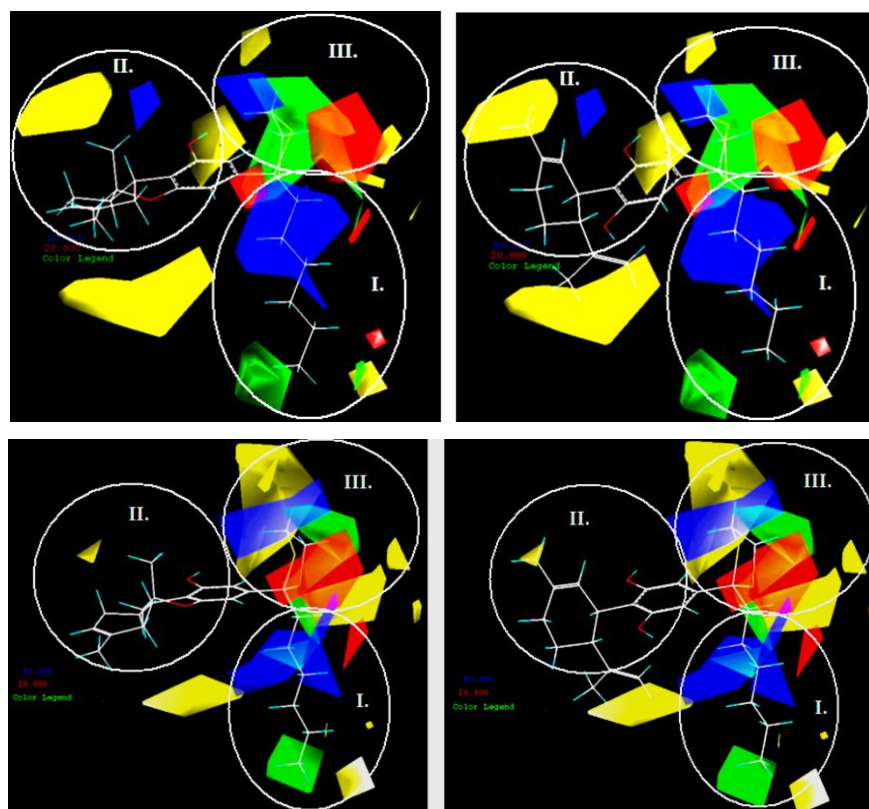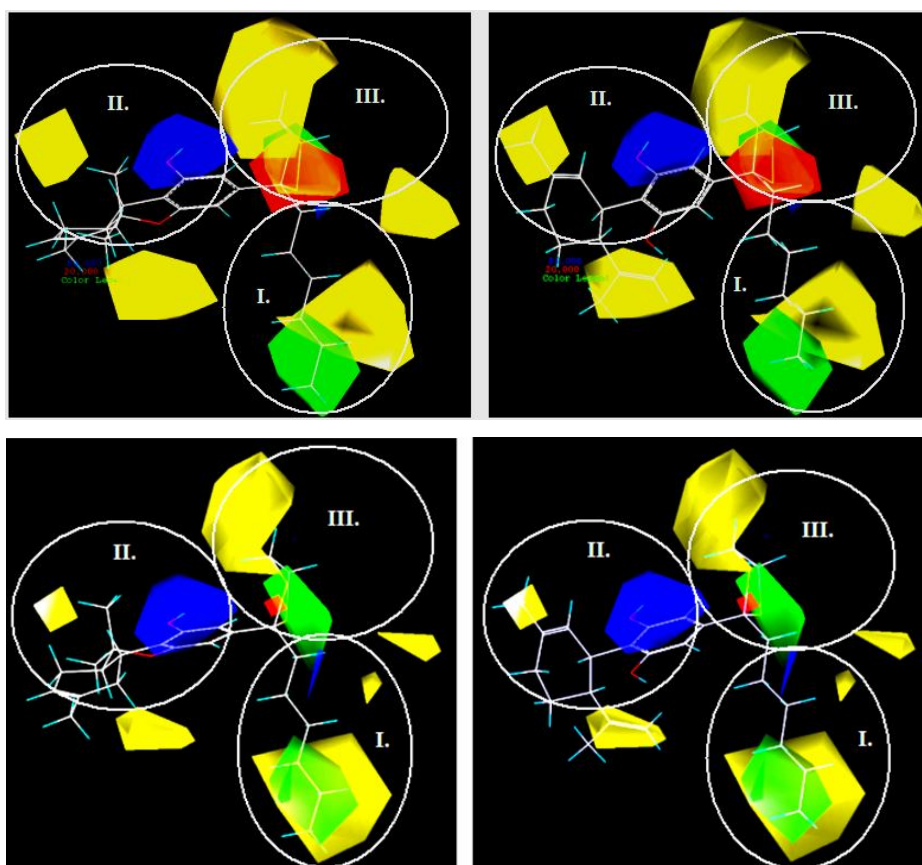
Figure 5.9 (top) CoMFA contour maps of template compound **12** (AMG3) (on the left) and its corresponding CBD analogue **13** (on the right) for the CB1 model. (Regions I, II, and III show contour maps around alkyl side chain, tricyclic part and α-face of C1' of ligand, respectively). **(bottom)** CoMFA contour maps of template compound **12** (on the left) and its corresponding CBD analogue **13** (on the right) for the CB2 model.

	CoMSIA/CB1	CoMSIA/CB2
Number of compounds in training set	30	29
r_{cv}^2	0.746	0.625
r^2	0.944	0.912
Standard error of estimate	0.296	0.324
F	65.031	47.855
Relative contributions of steric/electrosatic fields	0.890:0.110	0.918:0.082
Number of optimal components	6	5

Table 5.7 Statistical results obtained by 3D QSAR/CoMSIA models for CB1 and CB2.

5.2.1.7 Discussion

When correlations are sought using reported data, one must take into account (i) large variability in testing procedures and (ii) uncertainties related to enantiomeric purities of synthetic molecules. A careful examination of published data[5,84,85] identifies essential molecular fragments contributing to 'cannabimimetic activity'. One of them is the aliphatic C3-alkyl side chain; the role of this pharmacophore is important for hydrophobic interactions with the site(s) of action. There is an established SAR which indicates longer side chains are correlated with more potent CBs.[84,126] Decreasing the length of the n-pentyl side chain of Δ^9-THC by two carbons, reduces potency by 75%, and extension of the five carbon atom chain by adding one or two carbons favors binding, while further extension is detrimental. Interestingly, analogues with substituents, e.g. CH_3, C_2H_5, Cl, or I in the ortho position to the phenolic hydroxyl, retain substantial biological activity, however the para substitution produces inactive analogues.[84,126] Accordingly, para substituents prevent the side chain from orienting to a southern direction with respect to the phenolic hydroxyl group, resulting in decreased CB activity. On the other hand, ortho substitution allows such an orientation.[84] Thus, the orientation of the alkyl side chain plays an important role in the determination of biological activity. A significant degree of conformational restriction can be imposed upon the alkyl side chain either by the introduction of a double bond or a new cyclic ring fused to the aromatic ring A, leading to variations in biological responses.[67] Khanolkar *et al.*[6] presented a series of Δ^8-THC analogues, in which the *n*-heptyl side chain was restricted by a C2-C3 cyclohexyl ring, and showed that the alkyl side chain pointing downwards has a 18-fold higher binding affinity for the CB1 receptor and a 3-fold higher binding affinity for the CB2 receptor than the respective analogue in which the alkyl side chain orients laterally.

Compound	CB1 CoMSIA model		CB2 CoMSIA model	
	pK_i (observed)	pK_i (predicted)	pK_i (observed)	pK_i (predicted)
1	7.02	7.19	7.14	7.31
2	6.20	6.12	6.43	6.43
3	6.92	7.03	7.29	7.19
4	7.24	6.98	6.97	7.15
5	7.93	7.54	8.03	7.74
6	6.12	6.40	6.65	6.75
7	7.55	7.63	7.60	7.97
8	6.59	6.51	6.98	7.01
9	8.08	7.83	8.41	8.18
10	6.50	6.51	6.96	7.08
11	6.77	6.91	6.99	7.04
12	9.49	9.00	9.28	8.68
13	6.87	6.89	7.30	6.99
14	9.28	9.20	9.66	9.18
15	7.24	7.15	6.59	6.99
16	8.74	8.45	8.44	8.43
17	7.49	8.03	7.71	8.10
18	9.35	9.82	8.72	9.32
19	7.32	7.33	7.41	7.46
20	5.90	5.98	6.64	6.45
21	7.66	7.95	-	-
22	9.08	9.05	9.31	9.24
23	9.36	9.13	9.07	9.18
24	7.23	6.74	7.00	7.15
25	8.90	9.19	9.54	9.35
26	6.18	6.65	7.48	7.37
27	9.15	9.10	8.99	9.33
28	6.72	6.61	7.20	7.38
29	7.66	7.82	7.08	6.78
30	8.66	8.55	8.48	8.26

Table 5.8 Observed and predicted pK_i (by CoMSIA models) values at the training set for CB1 and CB2 receptors.

The CB1 and CB2 receptors belong to the same receptor family and exhibit a 44% sequence homology, which rises to 68% in the TM domains, an area thought to be involved in ligand recognition.[5] Because of this high degree of homology, it is not surprising that binding affinities for CB1 and CB2 receptors are correlated. Figures 9 and 11 show the field contributions to the binding affinity among the CBs and provide a visualization of both steric and electrostatic interactions at the receptor site. The result demonstrates the importance of the hydrophobic components of the CBs with cannabimimetic activity and is

Figure 5.10 Plots of corresponding CoMSIA-predicted and experimental values of binding affinity (given as pK$_i$) of CB analogues in the training set at the CB1 (on the left) and CB2 (on the right) receptors, respectively.

consistent with other studies. The CoMSIA results are in agreement with the CoMFA results. The contour maps resulted by applying CoMFA and CoMSIA methodologies demonstrate that there are similar and different structural requirements for optimum ligand binding at the CB1 and CB2 receptors. Derived 3D contour maps of CoMFA and CoMSIA models were investigated in the three distinct regions:

Alkyl side chain-Molecular segment I: The green colored contours along the left side of the end of the alkyl chain show that bulky groups enhance the binding affinity for the CB1 and CB2 receptors in both CoMFA and CoMSIA models (Figures 5.9 and 5.11). For example, the presence of adamantane, phenyl, *t*-butyl, isopropyl, or cyclopentyl groups in this region is expected to enhance CB1 and CB2 receptor binding affinities. There are large yellow colored contours on the right side of the end of the alkyl side chain in the CB1 and CB2 CoMSIA models (Figure 5.11), and small areas for the corresponding CB1 and CB2 CoMFA models (Figure 5.9) showing the existence of sterically unfavorable fields (the areas in which steric bulk is predicted to decrease binding). Thus, the orientation of the alkyl side chain plays an important role in determining biological activity. This result confirms the previous published reports.[84,104]

Figure 5.11 (top) CoMSIA contour maps of template compound **12** (AMG3) (on the left) and its corresponding CBD analogue **13** (on the right) for the CB1 model. (Regions I, II, and III show contour maps around alkyl side chain, tricyclic part and α-face of C1' of ligand, respectively). **(bottom)** CoMSIA contour maps of template compound **12** (on the left) and its corresponding CBD analogue **13** (on the right) for the CB2 model.

Compounds **12**, **14**, **16**, **18**, **22**, **23**, **25**, and **27** show high activity but low selectivity for the CB1 and CB2 receptors attributed to their fit in the hydrophobic subsite of both receptors.[6] An optimal interaction is observed when a lipophilic group is attached to C1' position. The CB1 receptor appears insensitive to isosteric groups attached to the C1' position whereas the CB2 receptor shows a higher preference for the smaller dioxolane five-membered ring rather than the dithiolane ring or more hydrophobic cyclopentyl analogues.[67]

ABC ring-Molecular segment II: The yellow colored contour at the α-face of the C-ring in the Δ^8-THC analogues (Figures 5.9 and 5.11, on the left) indicates areas in which steric bulk is predicted to decrease binding. However, in the case of CBD analogues this area fits on the C9-methyl group (Figures 5.9 and 5.11, on the right). Bulky groups localized between molecular segments I and II are expected to reduce the binding affinities of CB analogues at both CB1 and CB2 receptors. In these regions, the steric interactions affect differently the binding affinities of Δ^8-THC and CBD analogues for the CB1 and CB2 receptors in both the CoMFA and CoMSIA models. In Δ^8-THC analogues, a sterically unfavorable area (yellow colored contour) is located between the regions I and II (Figures 5.9 and 5.11, on the left). In the case of CBD analogues, because of the different structural orientation of the bicyclic segment, this area fits on the methyl and propenyl groups (Figures 5.9 and 5.11, on the right). If the binding affinity value of Δ^8-THC analogues and their respective CBD analogues is compared, CBD analogues usually have lower binding affinities than their corresponding Δ^8-THC analogues. For example, the template compound **12**, has 425-fold and 97-fold higher binding affinities than its respective CBD analogue **13**, for CB1 and CB2 receptors, respectively. This can be explained by different topographical requirements for the Δ^8-THC and CBD derivatives at the cyclic ring segment. The CB1 receptor is more sensitive than the CB2 receptor to this different structural orientation, because in this region the sterically unfavorable area (yellow colored contour) is larger at the CB1 model (Figures 5.9 and 5.11).

α-face of C1'-Molecular segment III: Sterically unfavorable contour (yellow colored) is localized in the vicinity of ring A (Figures 5.9 and 5.11). Therefore, the existence of bulky groups in this molecular segment results in the decrease of the binding affinity as it is confirmed by compounds **15** and **16**. Figure 5.12 shows the steric-electrostatic CoMSIA contour maps of compound **15** for CB1 and CB2 receptors, respectively. The contour maps show that the increased binding affinity and pharmacological potency are associated with bulky (green colored contours) and negatively charged groups (red colored contours) in the α-face of C1', (Figures 5.9 and 5.11). The presence of such groups (e.g., C_6H_5OH,

$C_6H_5CF_3$, $C_6H_5CCl_3$, $C_6H_5CI_3$, C_6H_5I, etc.) in this region, are expected to enhance CB1 and CB2 receptor binding affinities.

The electrostatic contour maps that were correlated with the predicted potency were seen in the α-face of C1' (molecular segment III) for the both of CoMFA and CoMSIA models and in the middle of the alkyl side chain (molecular segment I) predominantly in the CoMFA models. Results show that in the α-face of C1' and in the middle of the alkyl side chain, ligands may interact with corresponding electropositive and electronegative atoms of CB1 and CB2 receptors, respectively (Figures 5.9 and 5.11).

In order to test the predictive ability of the obtained CoMFA and CoMSIA models, 20 other Δ^8-THC analogues have been added to the training set for the CB1 model and 13 other Δ^8-THC analogues have been added to the training set for the CB2 model. (Binding affinities have been taken from reported values in the literature[6,67]. Binding affinities of 7 CB analogues have been measured only for the CB1 receptor). The same CoMFA and CoMSIA settings and PLS analyses with initially derived models have been performed for the re-constructed CoMFA and CoMSIA models. Compound **12** has been used as a template and same atoms in the CoMFA and CoMSIA models have been selected for the structural superimposition processes for re-constructed models.
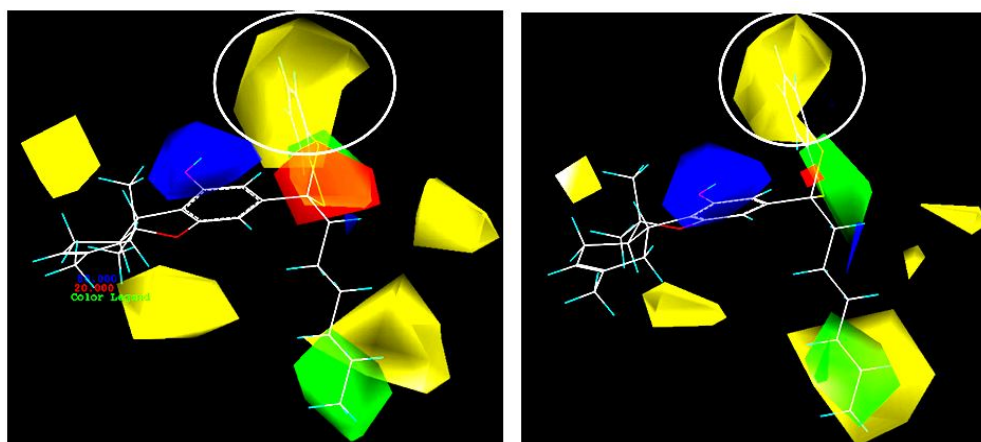
Figure 5.12 CoMSIA contour maps of **15** for the CB1 (on the left) and the CB2 models (on the right).

Table 5.9 shows the added Δ^8-THC analogues from literature[6,67] to the training set. Figure 5.13 illustrates the structural superimposition of CB analogues for the re-construction of derived CoMFA and CoMSIA models.

In order to re-build 3D QSAR models for the binding affinity at the CB1 and CB2 receptors, a training set of 50 CB analogues for CB1 receptor, and 43 CB analogues for CB2 receptor were included in the cross-validated PLS analyses. Table 5.10 shows the cross-validated and non-cross-validated r^2 values in all CoMFA analyses. The CoMFA study, based on selected lowest energy conformer of AMG3, gave cross-validated r^2 values of 0.770 and 0.614 for CB1 and CB2 models, respectively. The non-cross-validated PLS analysis yielded an r^2 of 0.955, and 0.926 and the estimated standard errors were 0.242 and 0.289 for CB1 and CB2 models, respectively. Thus, the CoMFA re-generated 3D QSAR models for the binding affinities to the receptors CB1 and CB2 has a very good cross-validated correlation. Figure 5.14 shows the relationship between the calculated and experimental pK_i values of the non-cross-validated analyses for CB1 and CB2 receptors. Figure 5.15 shows the steric-electrostatic contour maps of the re-generated CoMFA models for CB1 and CB2 receptors.

Table 5.11 shows the derived statistical results from re-constructed CoMSIA analysis. The CoMSIA study, based on AMG3 as template, gave cross-validated r^2 values of 0.740 and 0.572 for models CB1 and CB2, respectively. The non-cross-validated PLS analysis yielded an r^2 of 0.891, and 0.883 and the estimated standard error was 0.369 for both CB1 and CB2 models. Figure 5.16 shows the relationship between the calculated and experimental pK_i values of the non-cross-validated analyses for CB1 and CB2 receptors. Figure 5.17 shows the steric-electrostatic contour maps of the re-obtained CoMSIA models for CB1 and CB2 receptors.

Compound No.	R	K_i for CB1 (nM)	K_i for CB2 (nM)
31		14.00	-
32		10.90	-
33		3.90	-
34		2.70	-
35		141.0	-
36		22.30	58.60
37		6.80	52.0
38		0.34	0.39
39		0.57	0.65
40		0.94	0.22
41		67.60	85.90
42		297.0	23.60
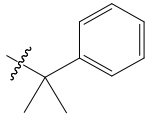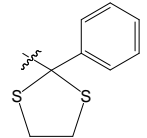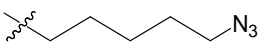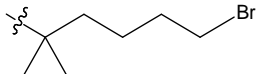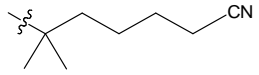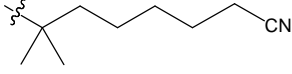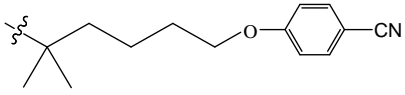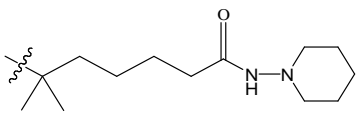
43		12.30	0.91
44		17.30	17.60
45		19.00	-
46		0.43	-
47		0.20	2.90
48		1.75	1.10
49		1.50	1.10
50		4.50	3.20

Table 5.9 Molecular structures and binding affinity, K_i values of CB analogues that added to training set (30 Δ^8-THC and CBD analogues) to re-construct CoMFA and CoMSIA models.[6,67]
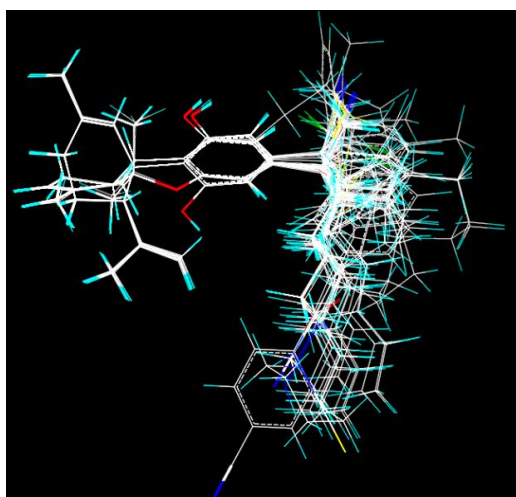
Figure 5.13 Structural superimpositions of the compounds for the re-construction of CoMFA and CoMSIA models.

	CoMFA/CB1	CoMFA/CB2
Number of compounds in training set	50	43
r_{cv}^2	0.770	0.614
r^2	0.955	0.926
Standard error of estimate	0.242	0.289
F	152.826	90.517
Relative contributions of steric/electrostatic fields	0.606:0.394	0.599:0.401
Number of optimal components	6	5

Table 5.10 Cross-validated and non-cross-validated analyses at the CB1 and CB2 receptors using the re-obtained CoMFA models, based on the AMG3 CB analogue used as the template.
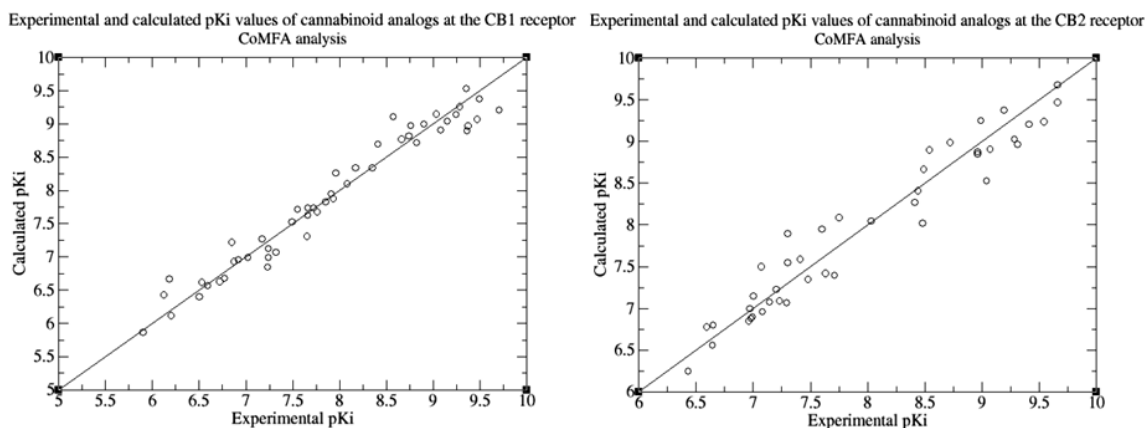
Figure 5.14 Plots of corresponding re-obtained CoMFA-calculated and experimental values of binding affinity (given as pK_i) of CB analogues in the training set at the CB1 (left) and CB2 (right) receptors, respectively.

The results from re-obtained models did not significantly modify the initially obtained models. The re-constructed 3D QSAR/CoMFA and CoMSIA models for the binding affinities to the CB1 and CB2 receptors have a very good cross-validated correlation. Although there are minor differences between initial and re-constructed CoMFA and CoMSIA models, the overall emerging picture is consistent. The main topographical requirements in the re-constructed CoMFA and CoMSIA models confirm the initially obtained models for the CB1 and CB2 receptors. The predictive ability of the initial model has been tested with added compounds, and it was shown that the model is able to accurately predict them as true unknowns.
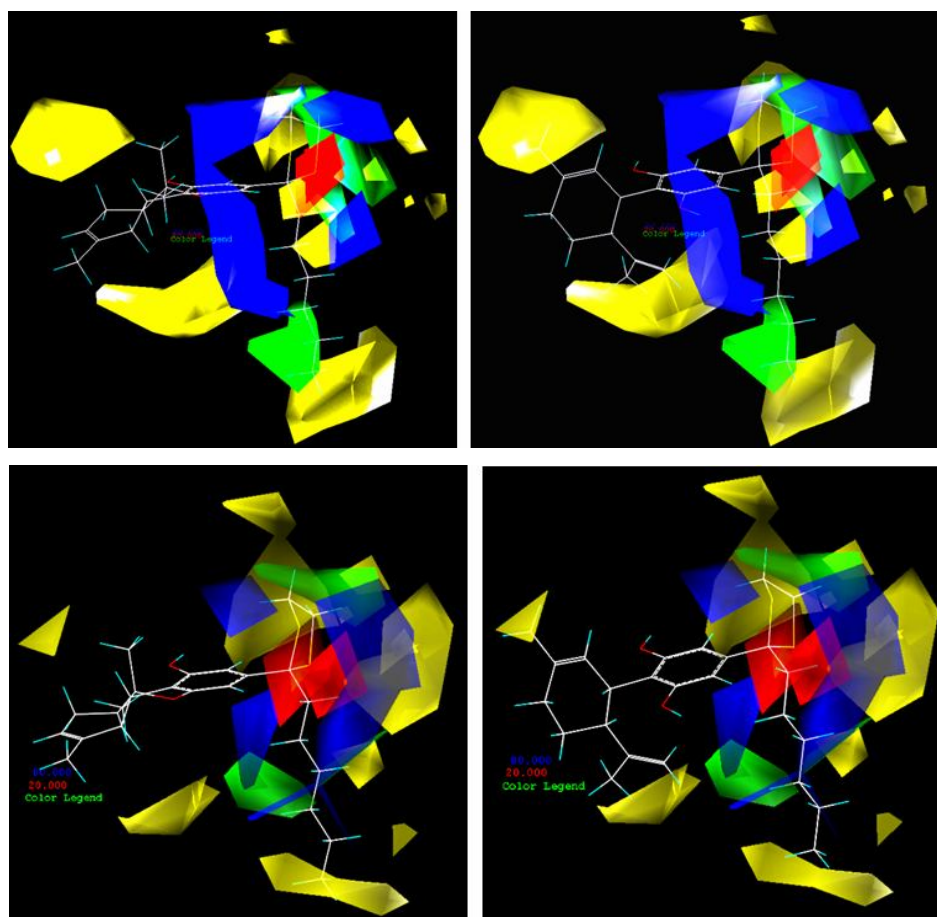
Figure 5.15 (top) CoMFA contour maps of template ligand **12** (AMG3, left) and its corresponding CBD analogue **13** (right) for the re-obtained CB1 model. **(bottom)** CoMFA contour maps of **12** (left) and its corresponding CBD analogue **13** (right) for the re-obtained CB2 model.

	CoMSIA/CB1	CoMSIA/CB2
Number of compounds in training set	50	43
r_{cv}^2	0.740	0.572
r^2	0.891	0.883
Standard error of estimate	0.369	0.369
F	71.836	44.096
Relative contributions of steric/electrostatic fields	0.947:0.053	0.921:0.079
Number of optimal components	5	6

Table 5.11 Cross-validated and non-cross-validated analyses at the CB1 and CB2 receptors using the re-obtained CoMSIA models, based on the AMG3 CB analogue used as the template.
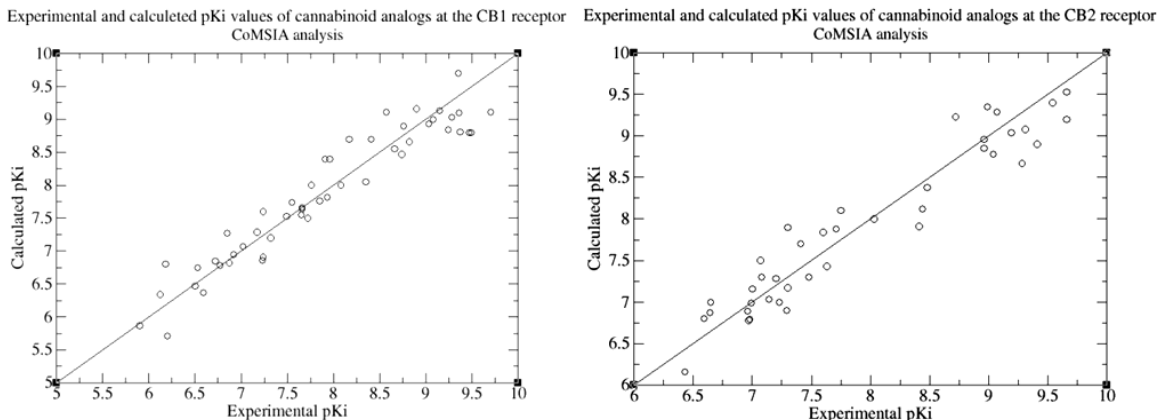
Figure 5.16 Plots of corresponding re-obtained CoMSIA-calculated and experimental values of binding affinity (given as pK_i) of CB analogues in the training set at the CB1 (left) and CB2 (right) receptors, respectively.
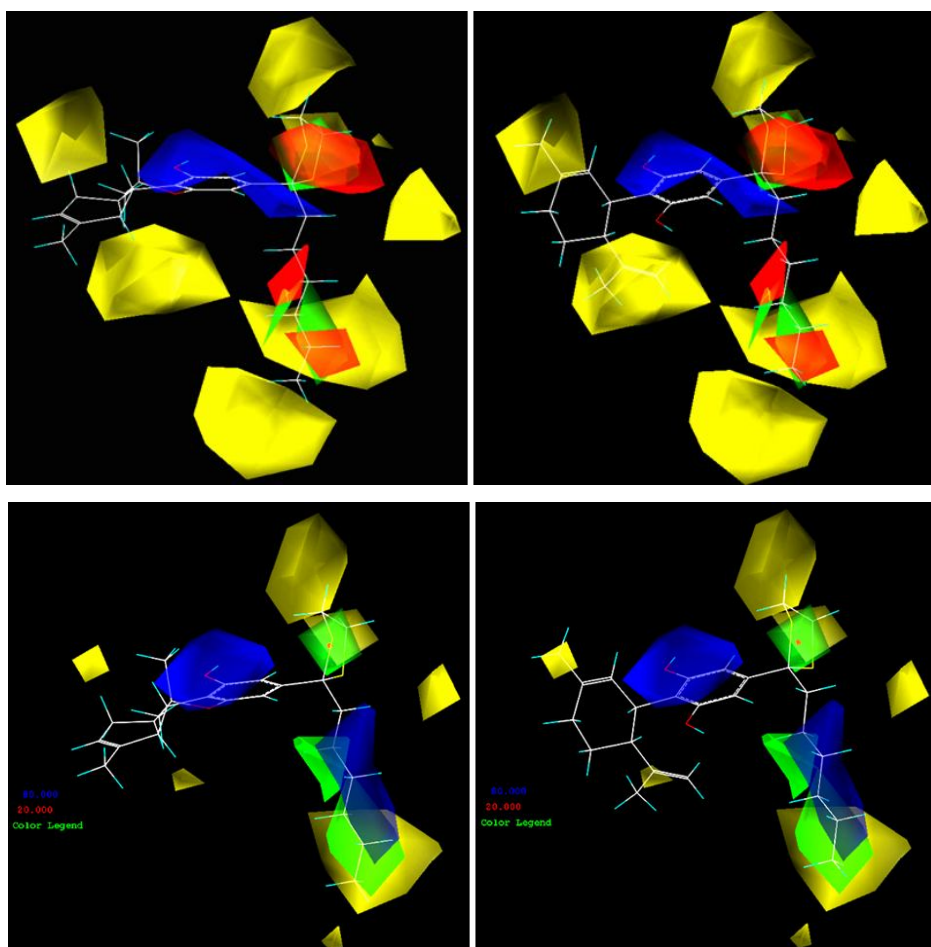
Figure 5.17 (top) CoMSIA contour maps of template ligand **12** (AMG3, left) and its corresponding CBD analogue **13** (right) for the re-obtained CB1 model. **(bottom)** CoMSIA contour maps of **12** (left) and its corresponding CBD analogue **13** (right) for the re-obtained CB2 model.

5.2.2 Conformational Analysis of AMG3 at the Binding Site of the Receptor

5.2.2.1 Molecular Docking Studies

3D models of the CB1 and CB2 receptors were constructed by several groups (e.g.; Salo *et al.*,[72] Shim *et al.*[8] and Tuccinardi *et al.*[33]) with a molecular modeling procedure, using the X-ray structure of bovine rhodopsin[110] as the initial template and taken into account the

available site-directed mutagenesis data. 3D models of CB receptors were obtained from Tuccinardi *et al.*[33] and used in the *in silico* docking simulations; however, the critical amino acids for the CB binding are determined considering reported CB models studied mentioned above. In addition, ligand binding pockets of the receptors have been obtained by the Biopolymer module of Sybyl molecular modeling package.[27] Figure 5.18 shows the proposed binding pockets for the CB1 and CB2 receptors. Two binding pockets in the CB1 and five binding pockets in the CB2 receptors have been determined. The found largest cavities of the receptors include same positions of the critical amino acids reported in the literature. Since conformers **A**, **C**, **F**, **J**, **K** and **L** are found as the most favored stable conformers through MD simulations *in solution* (DPPC bilayer environment without receptor), flexible docking has been employed to these six conformers using FlexX docking algorithm of Sybyl molecular modeling package.[27] The mean, the highest and the lowest values of the best 30 binding scores for the each complex of the CB1 and CB2 receptors and AMG3 conformers and the standard deviations between the scores are presented in Table 5.12. Among the conformations, the conformer **C** of AMG3 shows the best binding complex with the active site residues of both CB1 and CB2 receptors. Figure 5.19 (top) shows the localization of AMG3 ligand at the binding site of the CB1 (left) and CB2 (right) receptor models. Core of TM3-TM7 helices mainly participate to the binding cavities. Figure 5.19 (bottom) shows the interactions of binding site residues with the AMG3. The bioactive CB ligand AMG3 stabilizes its interactions with the active site through non-bonding van der Waals interactions with the non-polar surfaces of the active site residues of CB1 receptor (i.e., Lue193, Phe200, Thr201, Ile247, Pro251, Thr283, Trp356, Leu360, Val387) and CB2 receptor (i.e., Phe117, Leu194, Leu255, Trp258).
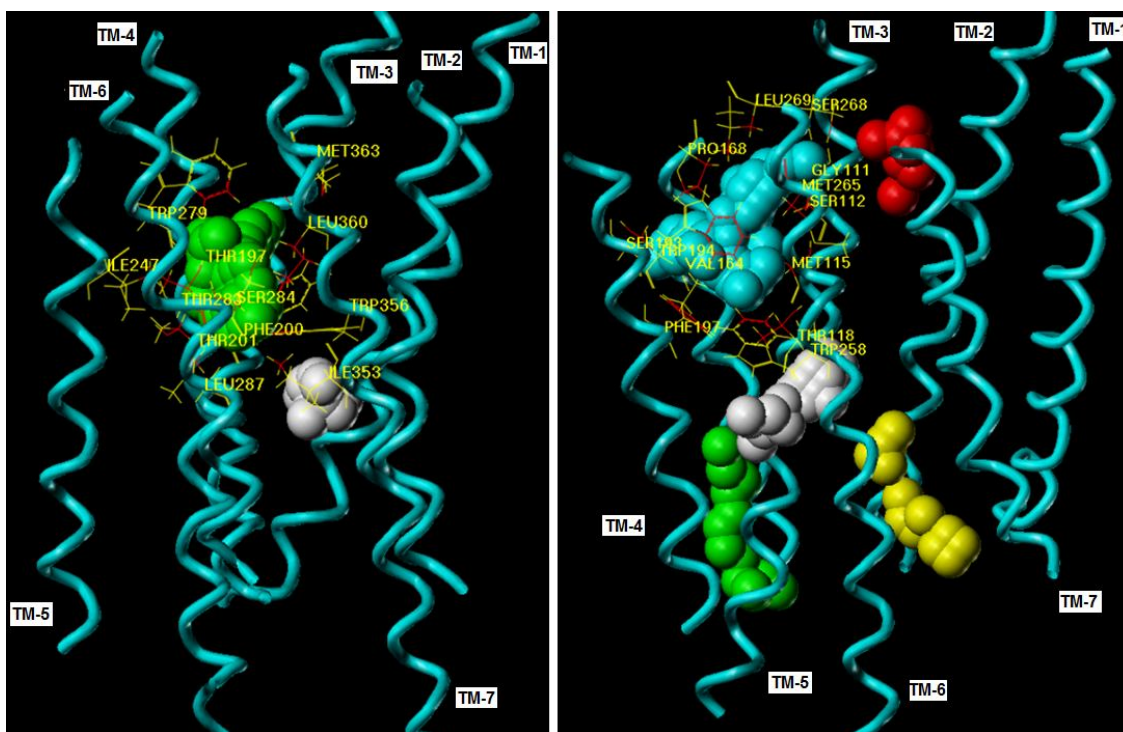
Figure 5.18 Proposed binding pockets for CB1 (left) and CB2 (right) receptor models.

The main characteristic of the low energy conformers of AMG3, both *in solution* and at the active site of the receptor is the high flexibility of the alkyl side chain. This is eminent by the low energy barriers observed in the various low energy rotamers of the molecule. It is noticed that the CB1 receptor has two available binding pockets (named as S1 and S2) for the accommodation of CB ligands (Figure 5.20i). S1 and S2 binding pockets constitute two cavities of ~7 Å and ~10 Å depths, respectively. They can both accommodate the alkyl side chain segment of CB ligands. Our findings are in accordance with previous reports[5,6,67], which show that extension of the five carbon atom chain (~7 Å) of THC by one or two carbon atoms (~10 Å) improves binding, while further extension (>10 Å) is detrimental due to steric hindrance. However, CB2 receptor has only one ligand binding pocket (Figure 5.20ii). Population analysis of docking modes for both CB1 and CB2 receptors showed that conformer **C** has highest propensity to bind at the active site of the CB1 and CB2 receptors.
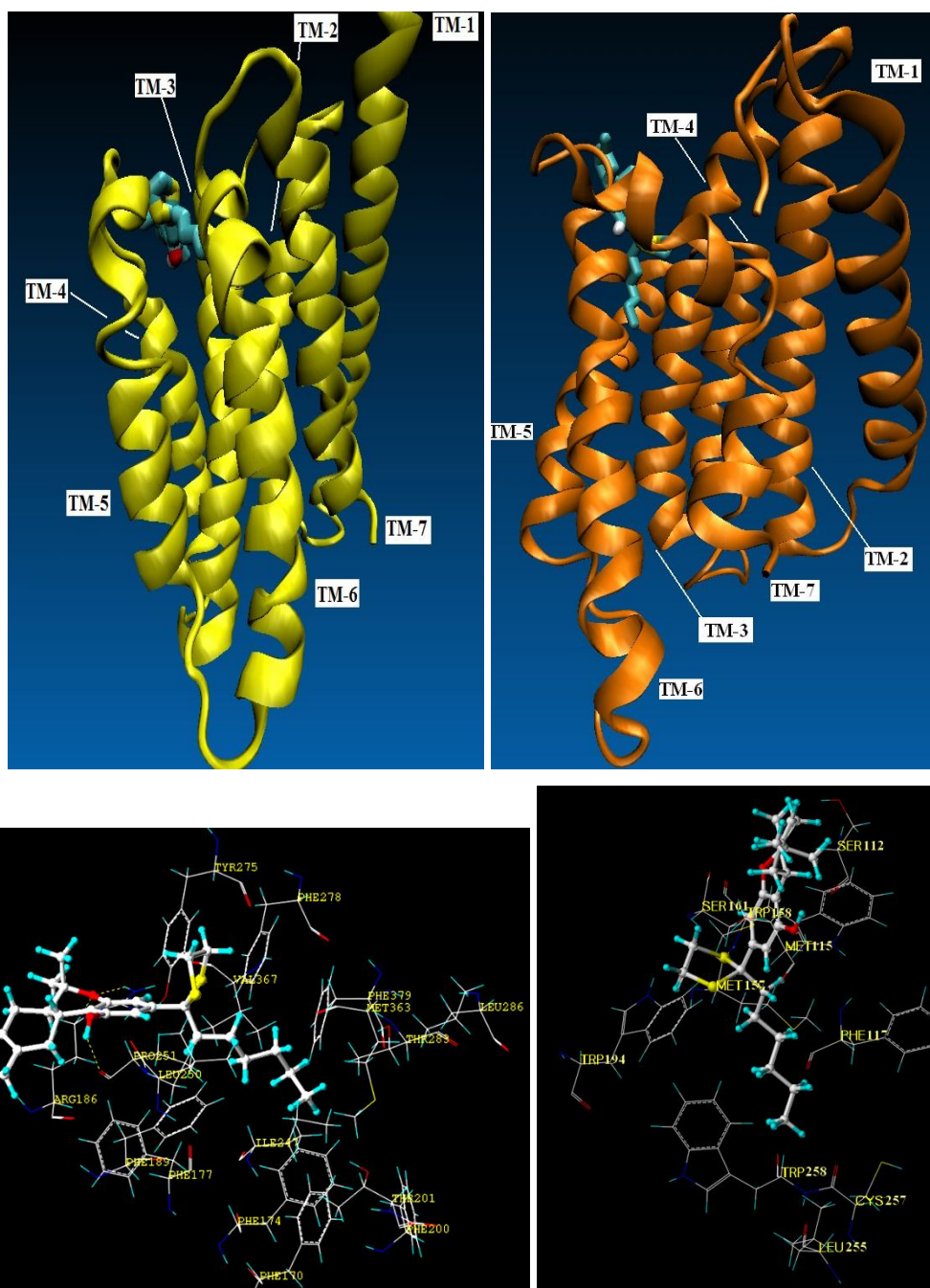
Figure 5.19 (top) Ligand location at the active site of the receptors CB1 (left) and CB2 (right); **(bottom)** AMG3 stabilizes its binding mainly through van der Waals interactions with the non-polar surfaces of the active site residues of CB1 receptor (left), (i.e., Lue193, Phe200, Thr201, Ile247, Pro251, Thr283, Trp356, Leu360, Val387) and CB2 receptor (right), (i.e., Phe117, Trp194, Leu255, Cys257, Trp258).

(i)

Conformer	Mean	Best	Worst	Std. Dev.
A	-9.93	-11.40	-9.26	0.60
C	-10.15	-11.43	-9.50	0.56
F	-9.50	-10.86	-8.90	0.46
J	-9.77	-11.14	-9.18	0.49
K	-9.55	-10.67	-9.01	0.44
L	-9.55	-11.15	-9.07	0.49

(ii)

Conformer	Mean	Best	Worst	Std. Dev.
A	-9.82	-11.50	-9.31	0.48
C	-10.26	-12.52	-10.26	0.69
F	-8.38	-9.22	-8.15	0.25
J	-10.29	-11.80	-9.73	0.55
K	-9.35	-11.15	-9.35	0.66
L	-8.90	-10.66	-8.26	0.66

Table 5.12 The mean, the highest and the lowest values of the best thirty binding scores for each complex of the (i) CB1 and (ii) CB2 receptors with AMG3 conformers, as well as the standard deviations between the binding scores.

5.2.2.2 Molecular Specificity for the S1 and S2 Binding Pockets at CB1 Receptor

Since AMG3 has conformational flexibility that can accommodate S1 and S2 binding pockets, it should be of great interest to study the conformational preferences of these two cavities located at the binding site of the receptor. Design of molecules with specific preference to either site may be of biological significance. S2 is deeper than S1 pocket and accommodates preferably the *all trans* conformation of the alkyl side chain segment of ligands. Unsaturation of the alkyl side chain of compounds leads their orientations towards S2 pocket. For this reason an analogue of AMG3 with four unsaturated bonds at the alkyl

chain was designed in order to specifically be directed to S2 cavity of the CB1 receptor (Figure 5.21i).
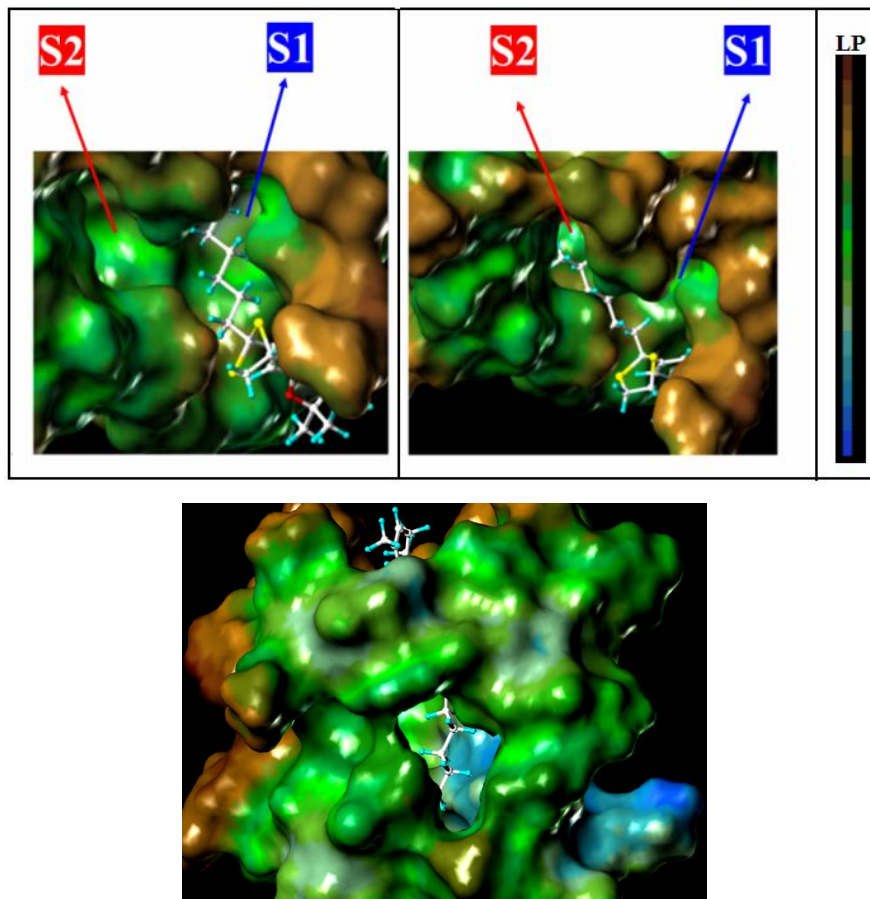
Figure 5.20 (top) Two cavities S1 and S2 observed at the active site of the CB1 receptor: S1 and S2 pockets constitute two cavities that have ~7 Å and ~10 Å depths, respectively and they accommodate the alkyl chain segment of CBs. **(bottom)** AMG3 ligand location at the CB2 receptor. MOLCAD lipophilic potential surface was calculated for the receptor with the Connolly method. Brown color denotes the most lipophilic areas and blue color denotes the most polar areas.

These predictions have been tested with the docking trials of novel analogues possessing unsaturation of alkyl side chain of AMG3. Imposing double bond unsaturation to the C2'-C3' single bond of alkyl chain of AMG3, is not enough for forcing the side chain towards

lateral orientation (top-left on the Figure 5.21ii). However further unsaturations e.g., addition of double bonds to C4'-C5' (top-right on the Figure 5.21ii) and C6'-C7' bonds (mid-left on the Figure 5.21ii) leads the orientation of unsaturated alkyl chain towards S2 binding pocket. The four unsaturated bonded analogue has optimum alkyl side chain length for the S2 ligand binding pocket (mid-right on the Figure 21ii). Further extension is detrimental (bottom on the Figure 5.21ii). Figure 5.21iii depicts the total FlexX binding scores versus number of double bonds at the alkyl side chain of AMG3 analogues mentioned above. Different unsaturation patterns have been also studied at the alkyl side chain of AMG3 (e.g., C2'-C3' and C5'-C6'), in order to cover all possible probabilities without accounting for the synthetic difficulty. The docking results showed that imposing double bond unsaturations to the C2'-C3' and C5'-C6' single bonds of alkyl chain of AMG3, is not enough to orient to the direction of S2 site. These observations may help open new avenues to synthetic chemists for synthesizing novel compounds.

(i)

(ii)

(iii)
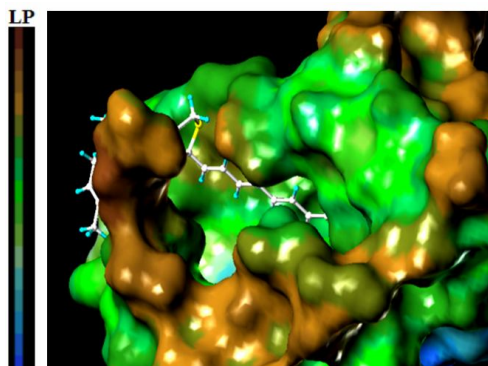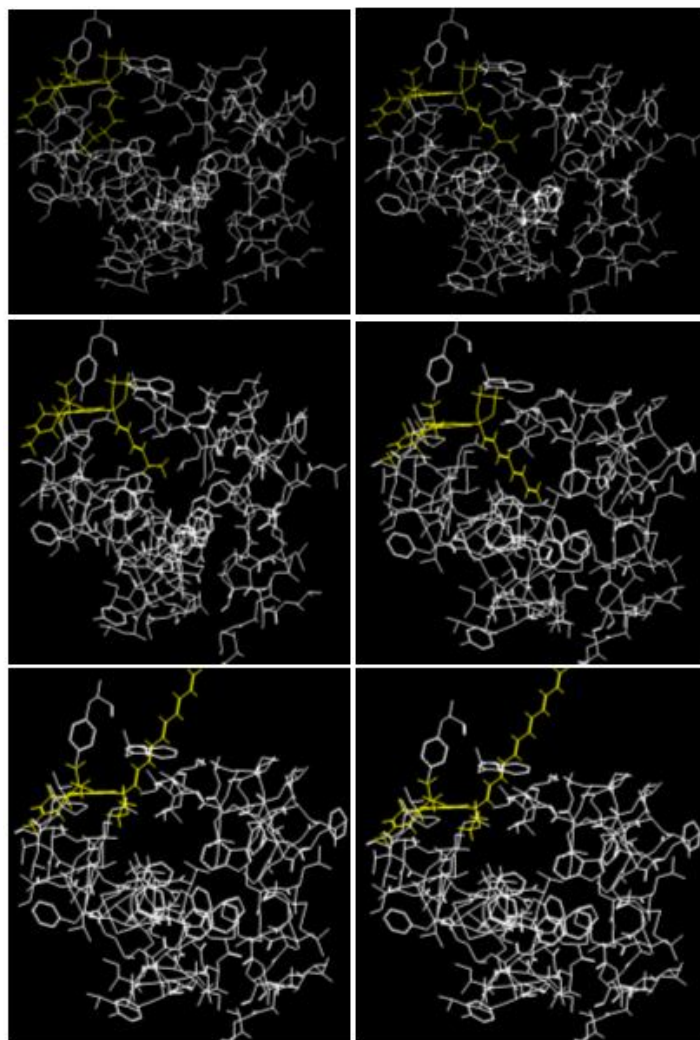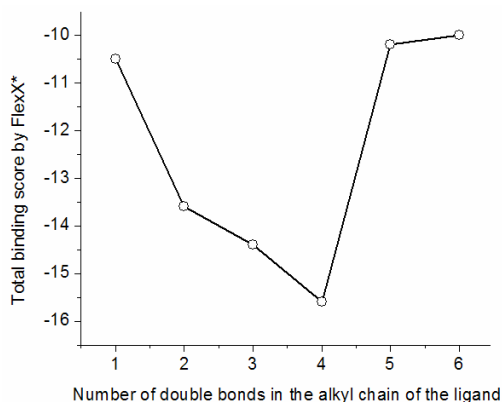
Figure 5.21 (i) Unsaturation of alkyl chain leads the orientation of chain towards S2. For this reason a structure was designed possessing four unsaturated bonds which were directed specifically to S2 cavity. **(ii)** Docking of rationally designed AMG3 analogues possessing double bonds at the alkyl side chain of the CB1 receptor. The degree of unsaturation is critical for the design of analogues to orient towards S1 or S2 binding cavity. **(iii)** Total FlexX binding scores versus number of double bonds at the side chain. The optimal number of double bonds at the alkyl side chain is four at S2 cavity.

5.2.2.3 Second Generation of 3D QSAR Models Based on *in Silico* Docking Results

After acquiring the highest percentage of conformer of AMG3 (conformer **C**) at the active site of the CB1 and CB2 receptors, the effect of used template conformation to the derived QSAR models was evaluated. For comparison, 3D QSAR/CoMSIA models with smoother potential functions have been used for the re-generation of CB1 and CB2 models.

Same common atoms (C_1, C_2, C_3, C_4, C_{4a}, C_{6a}, C_7, C_{10}, C_{10a}, C_{10b} and the oxygen atoms in the template conformers of AMG3) with the initial models were selected for superimposition processes. Figure 5.22 shows the superimpositions of CB analogues used as a training set to construct 3D QSAR/CoMSIA models based on the conformer **C** of template ligand

(for comparison, superimpositions of CB analogues based on conformer **A** of template ligand is also presented).

In order to build the CB1 and CB2 3D QSAR/CoMSIA models, a set of 30 Δ^8-THC and CBD analogues (Table 5.1) were analyzed. The pK_i values were used in the 3D QSAR correlations, and cross-validated PLS analyses were applied. Steric and electrostatic field columns of CoMSIA were generated. The same CoMSIA settings, PLS analyses and validations have been applied as initially generated models. A very high correlation was observed for both models as it is demonstrated by the high values of r^2 (Table 5.13). Additionally, the credibility of the models is proved by the high values of r_{cv}^2 (Table 5.13).
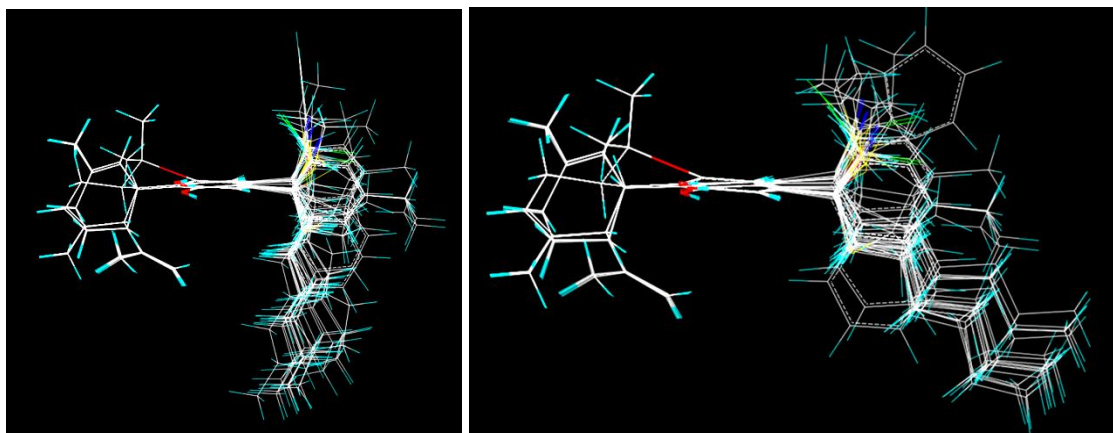
Figure 5.22 Structural alignments of the compounds in the training set for constructing 3D QSAR/CoMSIA models based on conformers **A** (on the left) and **C** (on the right) of the template ligand **12**.

Table 5.14 summarizes the experimental (observed) and CoMSIA-calculated pK_i results for the binding affinities at the CB1 and CB2 receptors. Figure 5.23 shows the relationship between the 3D QSAR/CoMSIA predicted and experimental pK_i values of the non-cross-validated analyses for the constructed models based on conformers **A** (left, top) and **C** (right, top) of **12** for CB1 receptor and corresponding conformers (**A**, left, bottom; and **C**, right, bottom) of **12** for CB2 receptors. The linearity of the plot concerning conformer **C** was better than the linearity of the plot concerning conformer **A**. Both plots showed good correlations for the constructed models.

	CB1 model-initial model (template ligand **12**-conformer **A**)	CB1 model (template ligand **12**-conformer **C**)	CB2 model-initial model (template ligand **12**-conformer **A**)	CB2 model (template ligand **12**-conformer **C**)
Number of compounds in the training set	30	30	29	29
r_{cv}^2	0.746	0.764	0.625	0.645
r^2	0.944	0.953	0.912	0.940
Standard error of estimate	0.296	0.272	0.324	0.247
F	65.031	77.600	47.855	57.491
Relative contributions of steric:electrosatic fields	0.890:0.110	0.890:0.110	0.918:0.082	0.885:0.115
Number of optimal components	6	6	5	5

Table 5.13 PLS analyses for the CB1 and CB2 receptors using the CoMSIA models based on compound **12** as template (using two different conformers **A** and **C**).

Figure 5.24 shows the steric and electrostatic CoMSIA contour maps of **12** and its corresponding CBD analogue **13** for the CB1 model using conformer **C** of AMG3 as template. For easy comparison, CoMSIA contour maps of CB1 model based on conformer **A** is also presented. Figure 5.25 presents the CoMSIA contour maps of **12** and its corresponding analogue **13** for the CB2 model using conformer **C** as a template ligand conformation. For easy comparison of contour plots, corresponding CoMSIA contour maps were also presented based on template conformer **A** of **12**.

Compound	CB1 CoMSIA model		CB2 CoMSIA model	
	pK_i(observed)	pK_i(predicted)	pK_i(observed)	pK_i(predicted)
1	7.02	7.38	7.14	7.42
2	6.20	5.93	6.43	6.29
3	6.92	7.16	7.29	7.25
4	7.24	7.11	6.97	7.17
5	7.93	7.78	8.03	7.98
6	6.12	6.15	6.65	6.77
7	7.55	7.66	7.60	7.82
8	6.59	6.41	6.98	7.01
9	8.08	7.93	8.41	8.02
10	6.50	6.33	6.96	6.88
11	6.77	6.83	6.99	6.93
12	9.49	9.45	9.28	9.21
13	6.87	6.86	7.30	6.99
14	9.28	8.99	9.66	9.01
15	7.24	7.03	6.59	6.53
16	8.74	8.80	8.44	8.79
17	7.49	7.63	7.71	7.70
18	9.35	9.61	8.72	9.25
19	7.32	7.25	7.41	7.57
20	5.90	5.88	6.64	6.29
21	7.66	7.95	-	-
22	9.08	9.20	9.31	9.19
23	9.36	8.79	9.07	9.03
24	7.23	7.03	7.00	7.35
25	8.90	9.06	9.54	9.27
26	6.18	6.85	7.48	7.42
27	9.15	9.05	8.99	9.24
28	6.72	6.81	7.20	7.36
29	7.66	7.89	7.08	7.03
30	8.66	8.39	8.48	8.32

Table 5.14 Summary of experimental (observed) and second generation of CoMSIA model predicted pK_i results of training set for the binding affinity at the CB1 and CB2 receptors.
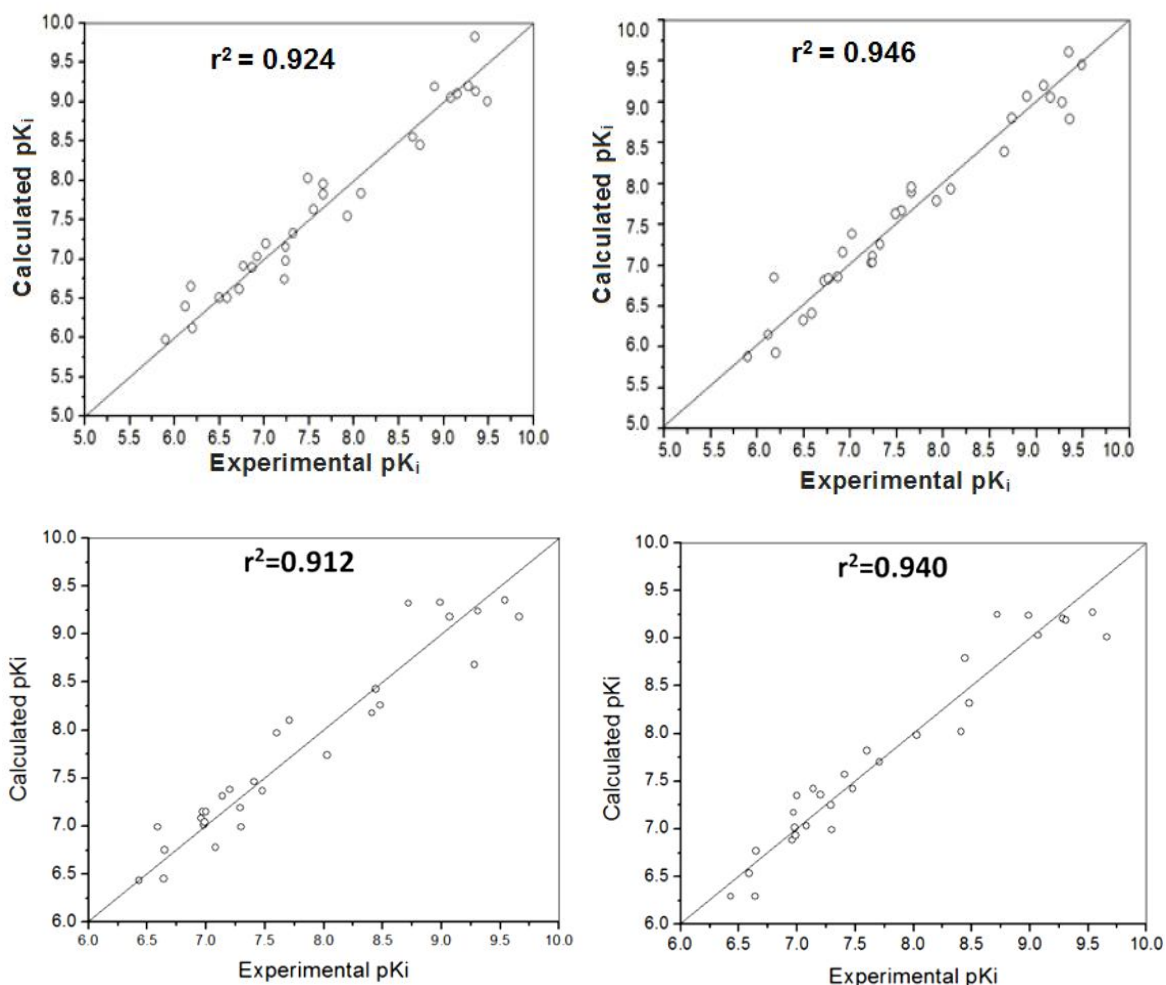
Figure 5.23 (top) Plots of corresponding 3D QSAR/CoMSIA predicted and experimental values of binding affinity (given as pK_i) of CB analogues in the training set at the CB1 receptor for the constructed models based on conformers **A** (on the left) and **C** (on the right) of **12**, respectively for CB1 receptor. **(bottom)** Plots of corresponding 3D QSAR/CoMSIA predicted and experimental values of binding affinity (given as pK_i) of CB analogues in the training set at the CB2 receptor for the constructed models based on conformers **A** (on the left) and **C** (on the right) of **12**, respectively for CB2 receptor.
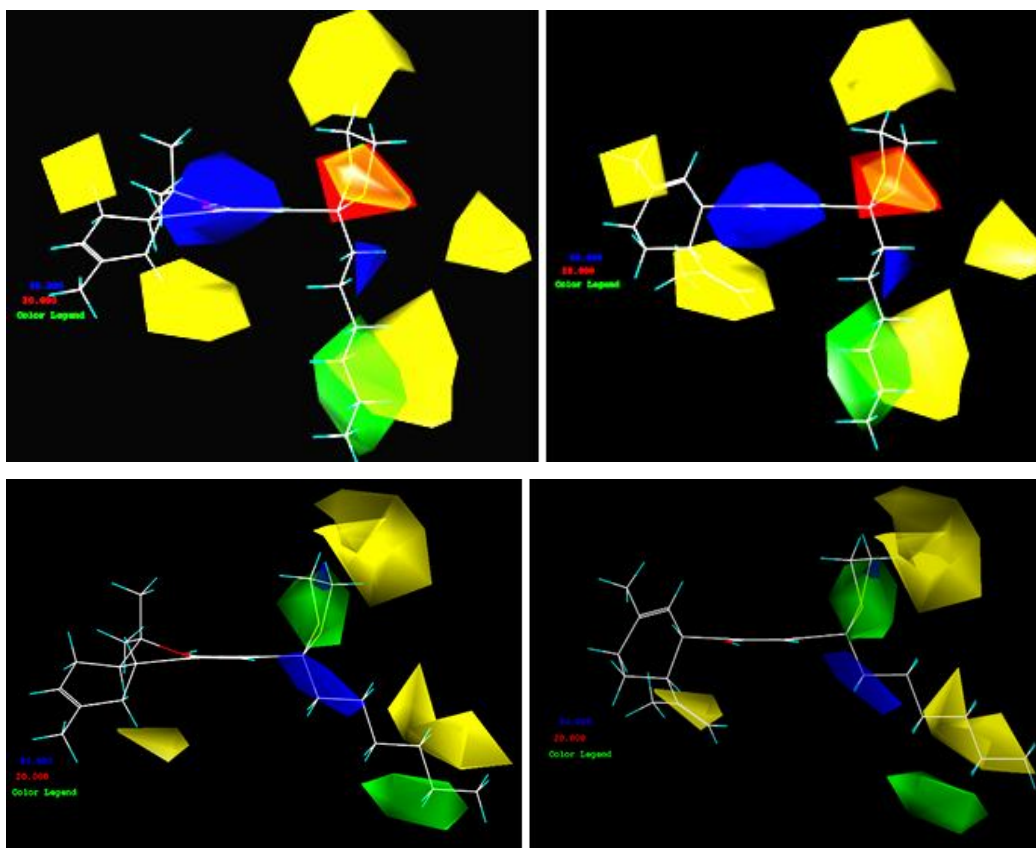
Figure 5.24 (top) CoMSIA contour maps of **12** (on the left) and its corresponding CBD analogue **13** (on the right) for CB1 model. Conformer **A** is used as template. **(bottom)** CoMSIA contour maps of **12** (on the left) and its corresponding CBD analogue **13** (on the right) for CB1 model. Conformer **C** is used as template.

Steric and electrostatic contour maps using template conformer **C** of **12** by 80:20 contribution field ratio for the favored and disfavored fields were not enough to visualize these contour maps at the cyclic ring segment of the compound. However, increasing the disfavored level ratio (e.g., 70:30) leads to visible steroelectronic contour maps and it shows similar properties as contour maps of template conformer **A** of **12** at the cyclic ring segment of the compound. For example, sterically unfavorable areas are located on the methyl or propenyl groups of CBD analogues, these unfavorable regions are located at the vicinity of the tricyclic segment of Δ^8-THC analogues (Figure 5.26). Thus, contour maps confirm higher affinities of Δ^8-THC analogues than their corresponding CBD analogues.
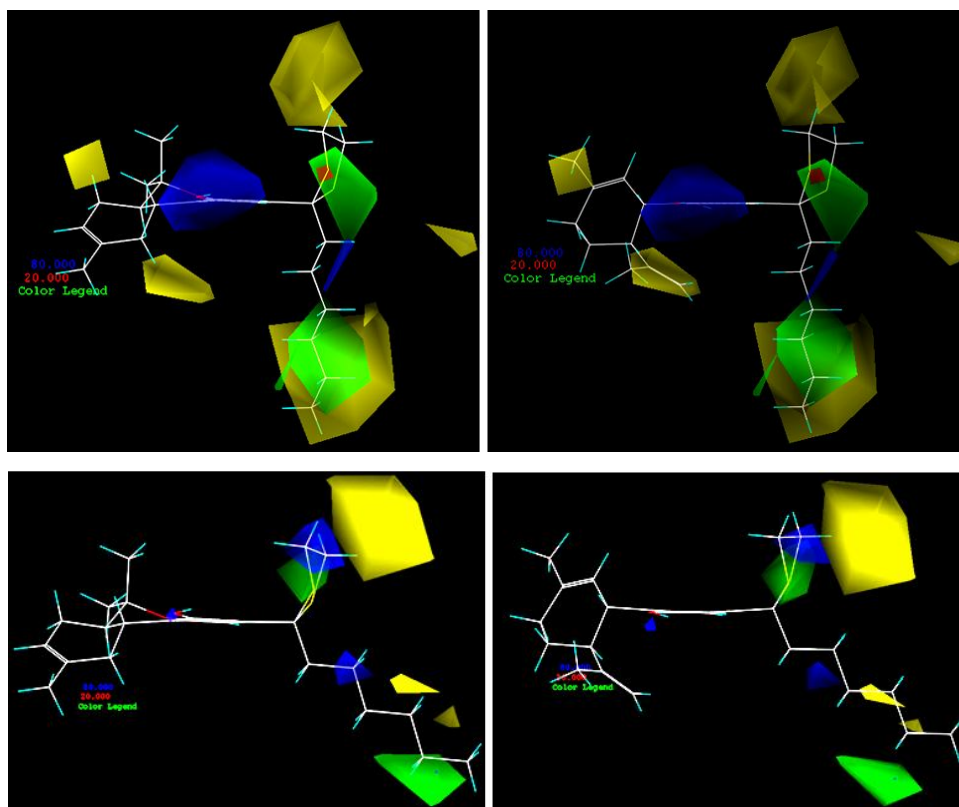
Figure 5.25 (top) CoMSIA contour maps of **12** (on the left) and its corresponding CBD analogue **13** (on the right) for CB2 model. Conformer **A** is used as template. **(bottom)** CoMSIA contour maps of **12** (on the left) and its corresponding CBD analogue **13** (on the right) for CB2 model. Conformer **C** is used as template.
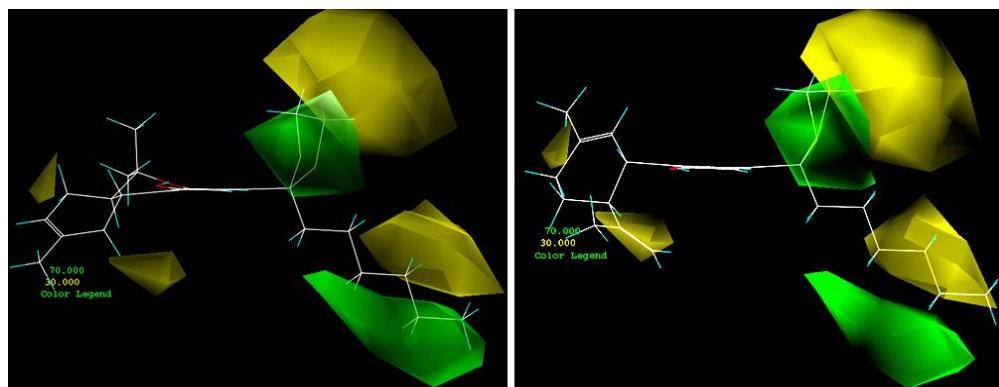
Figure 5.26 (top) CoMSIA steric contour maps of **12** (on the left) and its corresponding CBD analogue **13** (on the right) for CB2 model using 70:30 favored/disfavored field levels. Conformer **C** is used as template.

Three general conclusions could be drawn from the characteristics of derived 3D contour maps of CoMSIA models using both conformations of template ligand **12**:

1. Steric effects determine the binding affinity. The relative contributions of steric fields are larger than the electrostatic fields.

2. The orientation of the C3-alkyl chain plays a crucial role in determining the biological activity. The green colored contours along the left side of the end of the alkyl chain (corresponding to shown snapshot contour plots, Figures 5.24 and 5.25) show that bulky groups enhance the binding affinity, whereas bulky groups in the right sides of the C3-alkyl chain of analogues lead to decreased binding affinity.

3. Because of the structural differences of Δ^8-THC and CBD derivatives at the cyclic ring segment, these groups have different pharmacophoric requirements for their receptors in these regions. While sterically unfavorable areas are located on the methyl or propenyl groups of CBD analogues, these unfavorable regions are located at the vicinity of the tricyclic segment of Δ^8-THC analogues (Figures 5.24 and 5.25). This explains why usually Δ^8-THC analogues have higher binding affinities than their corresponding CBD analogues.

The conformers **A** and **C** of **12** used as a template compound in CoMSIA analyses show similarities and differences in contour maps. Their similarities are reflected in the same regions that contour levels of identical color cover. However, close observation reveals significant differences in their shape and extent of covering of the contour regions. The conformational differences of conformers **A** and **C** are localized in the alkyl chain. Our results confirm the earlier literature reports that the lipophilic alkyl chain plays crucial role in determining cannabimimetic activity for the CB receptors. Thus, the differences of contour maps at alkyl chain are important for the interpretation of pharmacophore groups that affect the binding affinity. When conformer **A** is used as a template, both THC and CBD analogues have green colored contour (depicts sterically favorable groups) at the tail of

alkyl chain (Figure 5.24). However, if conformer **C** is used as a template compound, then at the tail of alkyl chain only THC fits green colored sterically favorable contour (left on Figure 5.25). CBD analogues do not fit green colored contours but they fit yellow colored contours (depicts sterically unfavorable groups) (right on Figure 5.25). These important observations are obtained only by the model that was constructed with conformer **C** of **12**. The contour plots at the tail of alkyl chain which derived by the model that is constructed with conformer **C** of **12** demonstrates the better binding affinity of THC analogues than the corresponding CBD analogues.

In addition, to validate the higher predictive ability of conformer **C** of the template ligand **12**, 10 other Δ^8-THC analogues have been added to the training set and CoMSIA models have been reconstructed (binding affinities have been taken from reported values in the literature[6,67]). The same CoMSIA settings and PLS analyses have been performed for the re-constructed CoMSIA models. The same atoms in the template conformers of **12** have been selected for the structural superimposition processes. Results did not significantly modify the initially obtained models. Re-constructed 3D QSAR/CoMSIA models for the binding affinities at the CB1 and CB2 receptors have a very good cross-validated correlation (0.745 and 0.632, respectively for CB1 and CB2 receptors). Re-constructed models validate the initially obtained results: The model based on conformer **C** of **12** shows better statistical results than the model based on conformer **A** of **12**.

5.2.2.4 MD Simulations of AMG3 at the Active Site of Membrane-associated CB1 and CB2 Receptors

MD simulations have been performed to the systems including AMG3 at the binding site of the CB1 and CB2 receptors merged with membrane bilayer in order to analyze the effect of critical amino acid residues at the active site of CB receptors to the conformational properties of ligands in a more realistic environment. Figure 5.27 shows a representative picture of used systems. For these simulations, docked poses of complexes that have high population were used as initial ligand-receptor complex coordinates. Torsional angle val-

119

ues of the alkyl side chain of AMG3 throughout the simulations were screened with trajectory analysis. Results showed that adopted conformations of AMG3 at CB1 and CB2 receptors have different conformational properties. Torsional angles at the alkyl chain of AMG3 adopt *trans* for τ_1, τ_3, τ_5 and τ_6 at the binding site of CB2 receptor with some fluctuations near this dihedral angle. Trajectory analysis of torsional angles τ_2 and τ_4 show a propensity to be *gauche-;* τ_4 also adopts *trans* dihedral angles at the active site of the CB2 receptor (Figure 5.28). Dihedral angle screening throughout the MD simulations results showed that τ_2, τ_4 and τ_6 mainly form a *trans* conformations at the binding site of the CB1 receptor. Dihedral angles τ_3 and τ_5 are very flexible and adopt *gauche±* and *trans* conformations at the binding site of the CB1 receptor, however, *gauche+/trans* and *trans/gauche-* torsional angles are mainly observed for τ_3 and τ_5, respectively (Figure 5.28).

Therefore, five additional conformations (**M**, **N**, **O**, **P** and **R**) have been obtained from these simulations (Figure 5.29). Torsional angle screening results of MD analysis showed that conformers **C**, **M**, **N** and **O** of AMG3 favor at the active site of the CB1, and conformers **P** and **R** favor at the binding site of the CB2 receptor.

Although the CB1 and CB2 receptors exhibit a very high sequence homology which rises to 68% in the TM regions, there are certain behavior differences of AMG3 conformers at the binding sites of receptors. One of the main differences between the MD simulations of ligand at the CB1 and CB2 receptors is the different behavior of the first dihedral angle τ_1 of the alkyl side chain of AMG3. In the CB1 receptor, there is a high propensity of τ_1 to establish a *gauche+* conformation, however in the CB2 receptor; it prefers to have a *trans* conformation. It is well-known that, different conformational re-arrangements of third and sixth TMs of GPCR determine the activation of CBs. In CB2 receptor, alkyl side chain of AMG3 conformers align parallel in the ligand recognition part of TM3, while in the CB1 receptor they align perpendicular. This observation may help to understand the selectivity of CB ligands for the CB1 and CB2 receptors.

5.2.2.5 Third Generation of 3D QSAR Models Based on Conformational Analysis Results of AMG3 by MD Simulations at the Active Site of the Receptor

In order to test the effect of favored conformations of template compound (AMG3) at the binding site of the CB receptors to the QSAR models, initially all possible conformations have been tested as template conformer and 3D QSAR/CoMSIA models were re-obtained. Same common atoms in the template conformers of AMG3 with the initial models were selected for superimposition processes. Compounds in Table 5.1 and their measured activities were used for the third generation of CB1 and CB2 models. Initial statistical tests by obtained models showed that optimal statistical results derived using template conformer **O** for CB1 and conformer **P** for CB2 models. Therefore, further analysis is performed using these template conformers. Figure 5.30 shows the superimpositions of CB analogues used as a training set to construct 3D QSAR/CoMSIA models based on the conformers **O** and **P** of template ligand. Table 5.15 shows the derived statistical results for CB1 and CB2 models (in order to easy comparison, results of first and second generation of models were also included in the Table). It is clearly shown that the using template conformer obtained from MD simulations of ligand at the binding site of the receptor improves statistical results, significantly.
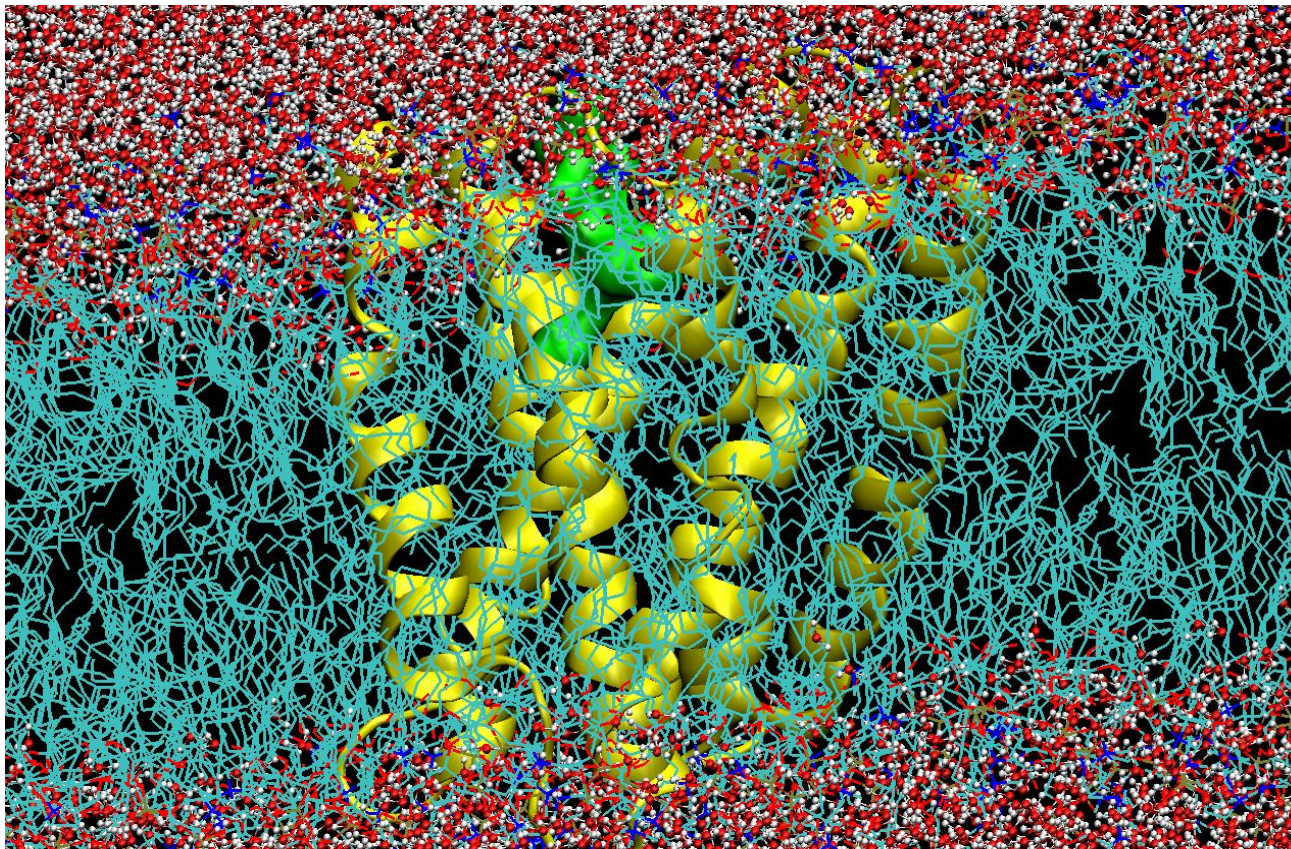
Figure 5.27 Representative picture of used system for MD simulations (ligand at the binding site of the receptor which is merged with lipid bilayer).

(i)
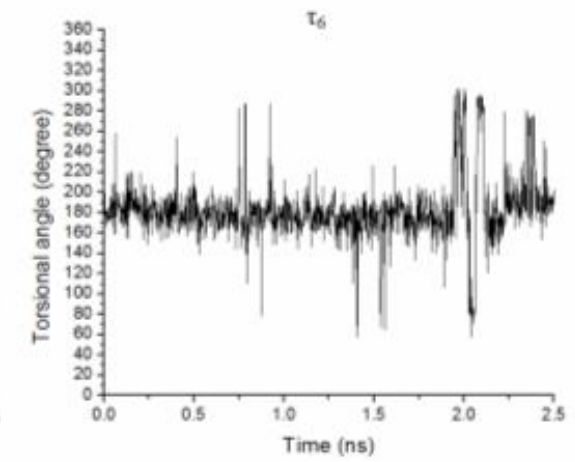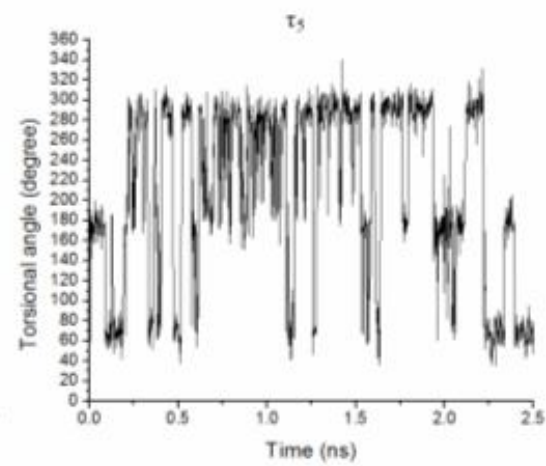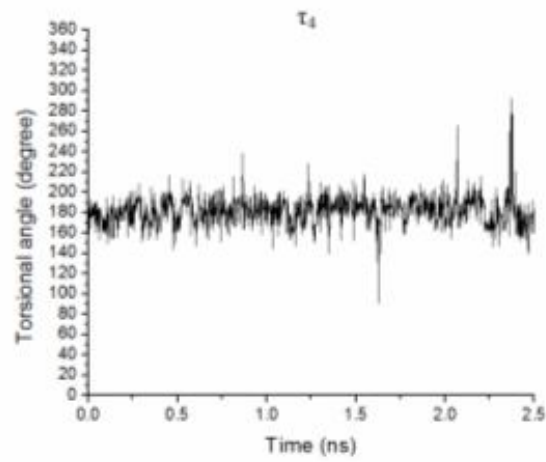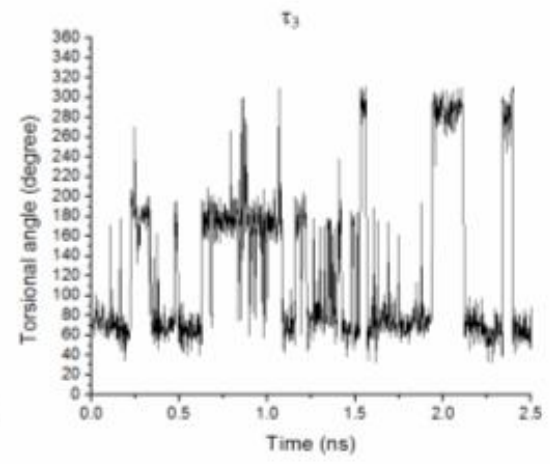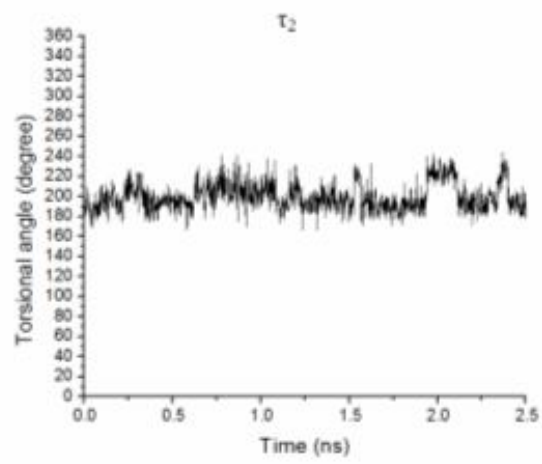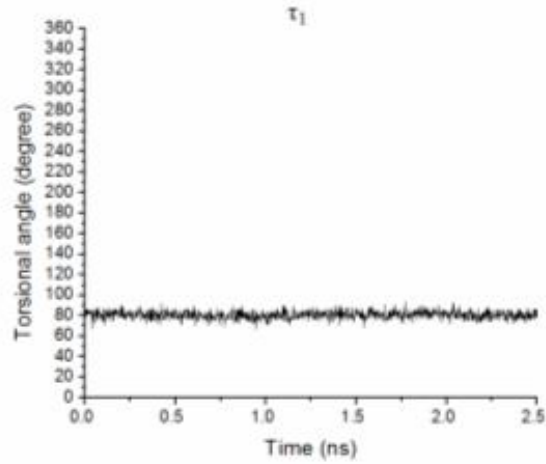
(ii)
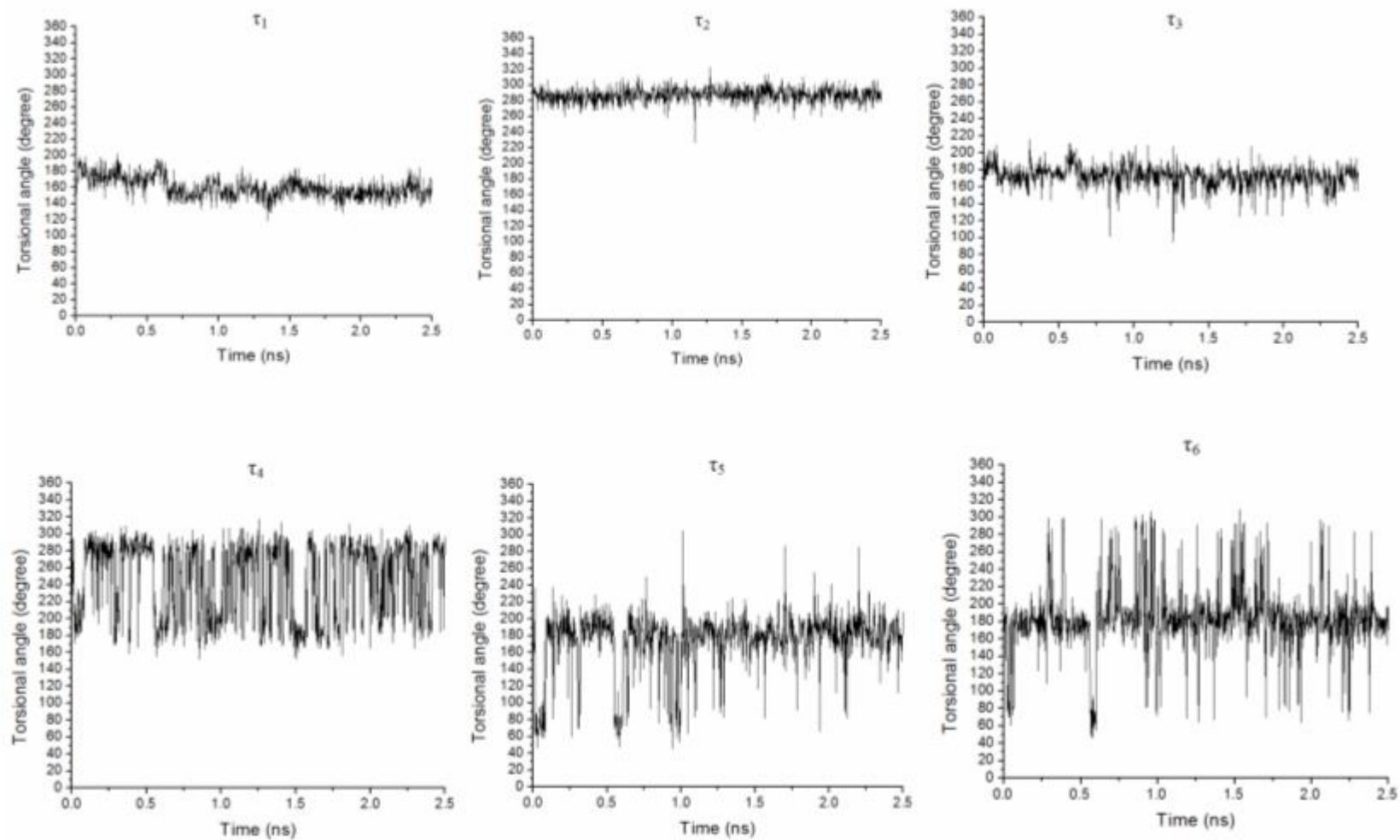
Figure 28. Torsional angle screening of alkyl side chain of AMG3 throughout the MD simulations at the binding site of **(i)** CB1 and **(ii)** CB2 receptors.
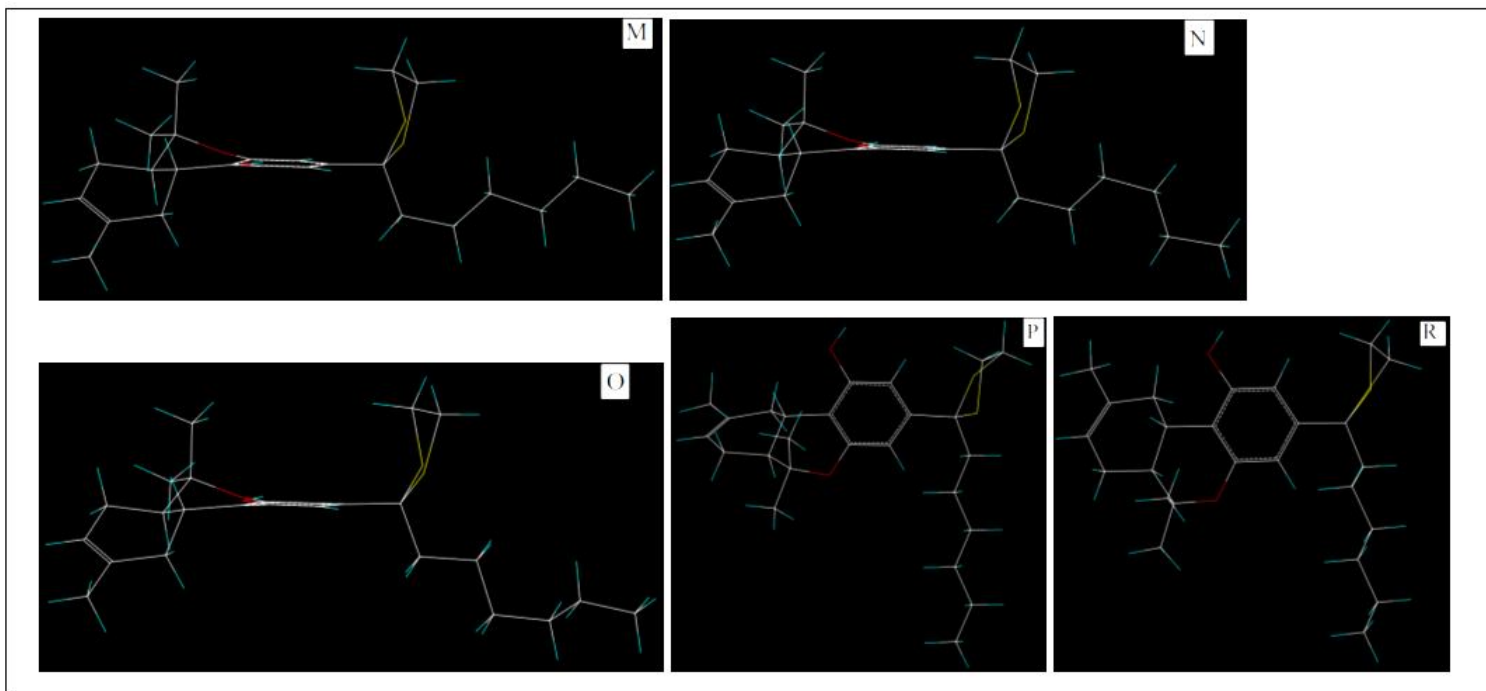
Figure 5.29 MD simulations of AMG3 at the active site of the receptors produced five additional conformations (**M**, **N**, **O**, **P** and **R**). At the active site of the CB1 receptor, conformers **C**, **M**, **N** and **O**; at the active site of the CB2 receptor, conformers **P** and **R** are found as favored conformations of AMG3. At the alkyl side chain these conformers have following dihedral angles: **M** (*gauche+/trans/gauche+/trans/trans/trans*); **N** (*gauche+/trans/gauche+/trans/gauche-/trans*); **O** (*gauche+/trans/trans/trans/gauche-/trans*); **P** (*trans/gauche-/trans/trans/trans/trans*) and **R** (*trans/gauche-/trans/gauche-/trans/trans*).

Experimental and corresponding predicted affinities of CB analogues were presented at the Table 5.16 and plotted at Figure 5.31.
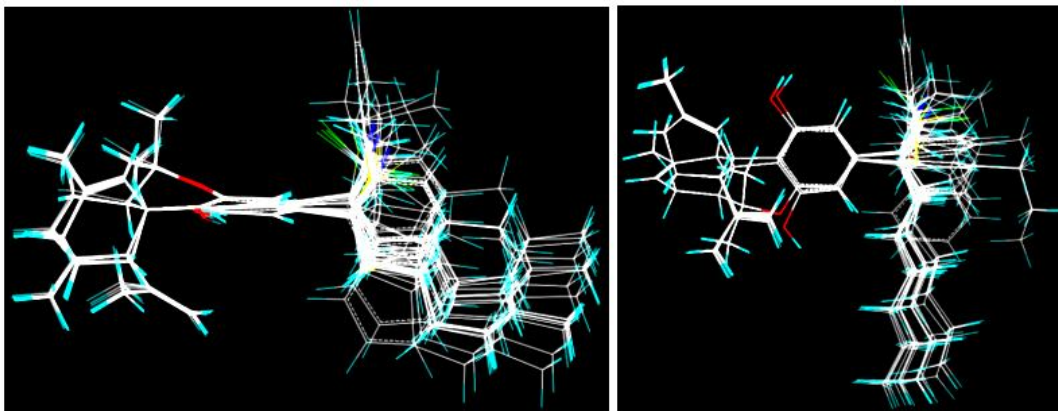
Figure 5.30 Structural alignments of the compounds in the training set for constructing 3D QSAR/CoMSIA models based on template ligand **12**. Superimposition was performed based on conformer **O** for CB1 model (on the left); and conformer **P** for CB2 model (on the right).

Figure 5.32 shows the steric and electrostatic contour maps of **12** (on the left, top) and its corresponding CBD analogue **13** (on the right, top) for the CB1 receptor using as template ligand the conformer **O**; and corresponding stereoelectronic contour maps (compounds **12** (on the left, bottom) and **13** (on the right, bottom) for the CB2 receptor using as template ligand the conformer **P**. Similar contour plots with initially generated models were obtained from the third generation of the models, however, the use of different template conformers for CB1 and CB2 models showed some fine and critical differences. For example, the orientation of alkyl chain of AMG3 is more restricted in the CB2 models, since left and right sides of the tail of the alkyl chain include yellow colored sterically unfavorable contours, and between these contours there is a green colored sterically favorable contour (Figure 5.32, bottom).

	CB1 model-first generation (template ligand **12**-conformer **A**)	CB1 model second generation (template ligand **12**-conformer **C**)	CB1 model third generation (template ligand **12**-conformer **O**)	CB2 model-first generation (template ligand **12**-conformer **A**)	CB2 model second generation (template ligand **12**-conformer **C**)	CB2 model third generation (template ligand **12**-conformer **P**)
Number of compounds in the training set	30	30	30	29	29	29
r_{cv}^2	0.746	0.764	0.771	0.625	0.645	0.710
r^2	0.944	0.953	0.962	0.912	0.940	0.952
Standard error of estimate	0.296	0.272	0.244	0.324	0.247	0.246
F	65.031	77.600	97.623	47.855	57.491	72.373
Relative contributions of steric:electrosatic fields	0.890:0.110	0.890:0.110	0.852:0148	0.918:0.082	0.885:0.115	0.853:0.147
Number of optimal components	6	6	6	5	5	5

Table 5.15 Derived statistical results of third generation (based on template conformation derived from MD simulations of template molecule at the membrane-associated CB1 and CB2 receptors) of CB1 and CB2 models. (In order to easy comparison, results of first and second generation of models were also included in the Table. First generation of models was based on favored template conformation *in solution*, and second generation of the models was based on docking results).

Compound	CB1 CoMSIA model		CB2 CoMSIA model	
	pK$_i$(observed)	pK$_i$(predicted)	pK$_i$(observed)	pK$_i$(predicted)
1	7.02	7.11	7.14	7.25
2	6.20	6.00	6.43	6.48
3	6.92	6.98	7.29	7.23
4	7.24	6.96	6.97	7.08
5	7.93	7.73	8.03	7.91
6	6.12	6.55	6.65	6.43
7	7.55	7.64	7.60	7.84
8	6.59	6.67	6.98	6.99
9	8.08	7.95	8.41	8.03
10	6.50	6.68	6.96	6.97
11	6.77	6.82	6.99	7.25
12	9.49	9.23	9.28	9.03
13	6.87	6.58	7.30	7.23
14	9.28	9.56	9.66	9.56
15	7.24	7.18	6.59	6.37
16	8.74	8.81	8.44	8.51
17	7.49	7.85	7.71	7.79
18	9.35	9.37	8.72	9.28
19	7.32	7.15	7.41	7.46
20	5.90	5.82	6.64	6.48
21	7.66	7.94	-	-
22	9.08	9.14	9.31	9.35
23	9.36	9.33	9.07	9.10
24	7.23	6.98	7.00	7.05
25	8.90	8.87	9.54	9.13
26	6.18	6.30	7.48	7.33
27	9.15	8.90	8.99	9.10
28	6.72	6.61	7.20	7.24
29	7.66	7.77	7.08	7.38
30	8.66	8.75	8.48	8.45

Table 5.16 Summary of experimental (observed) and CoMSIA-predicted pK$_i$ results of training set by third generation of the models for the binding affinity at the CB1 and CB2 receptors.

Obtained contour plots also merged with the receptor coordinate files (Figure 5.33, top). Results confirmed the accuracy of obtained contour plots by third generation of QSAR models. For example, sterically unfavorable contours of CB1 model fit with the side chains of the amino acid residues at the binding site of the CB1 receptor (e.g., Leu190, Arg186, Trp279, Tyr275, Tyr355 and Val364); (Figure 5.33, bottom).
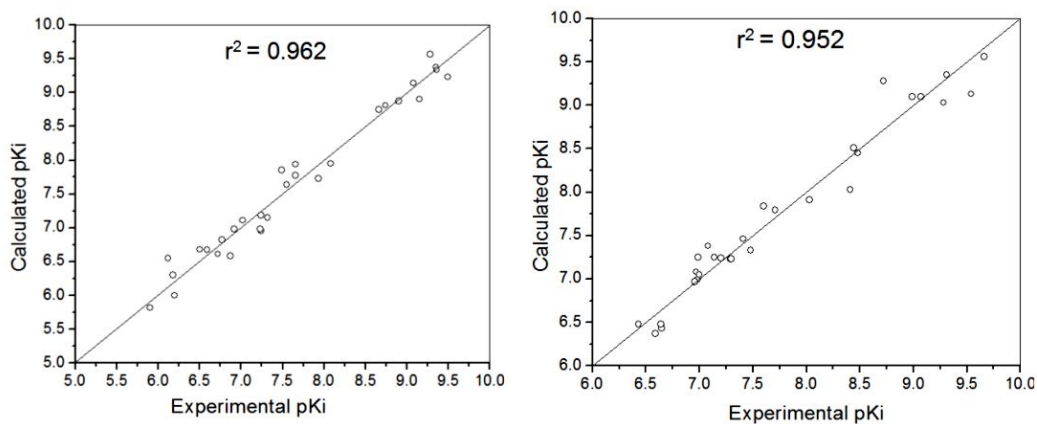
Figure 5.31 Plots of experimental and predicted values of binding affinities (given as pK$_i$) of CB analogues in the training set obtained by third generation of 3D QSAR (left) CB1 model and (right) CB2 model.
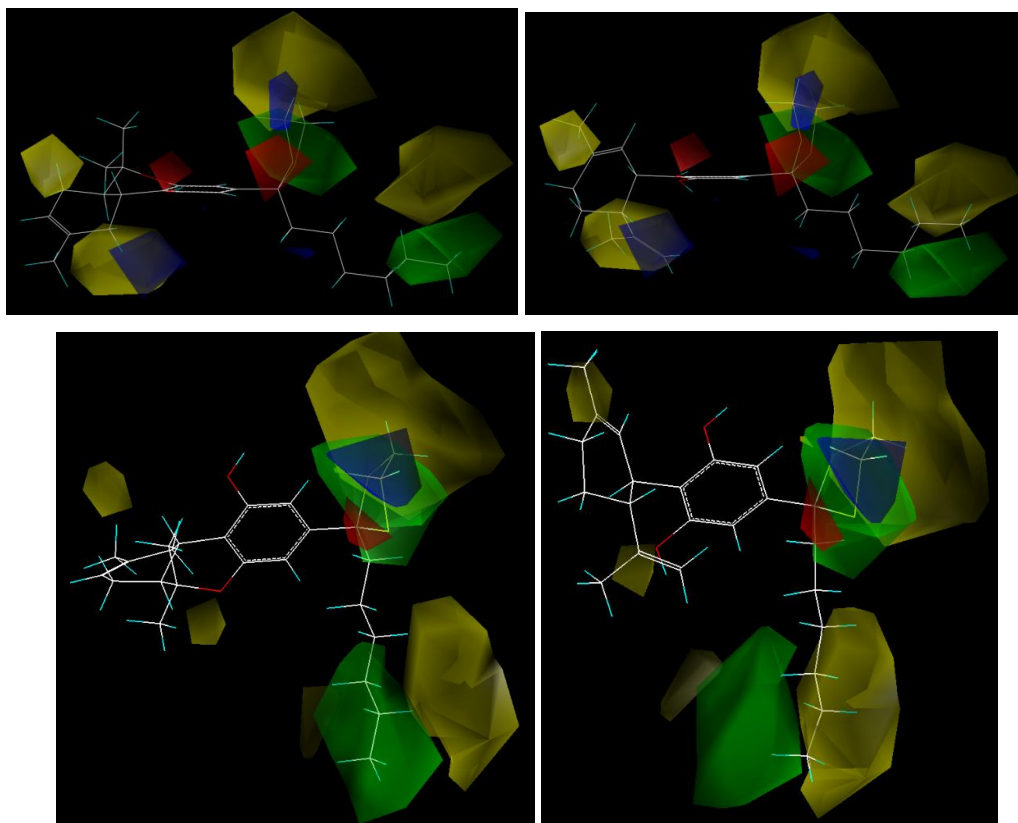
Figure 5.32 Steric and electrostatic contour maps of **12** (on the left, top) and its corresponding CBD analogue **13** (on the right, top) for the CB1 receptor using the template ligand as conformer **O**; and corresponding stereoelectronic contour maps (compounds **12** (on the left, bottom) and **13** (on the right, bottom) for the CB2 receptor using the template ligand as conformer **P**.
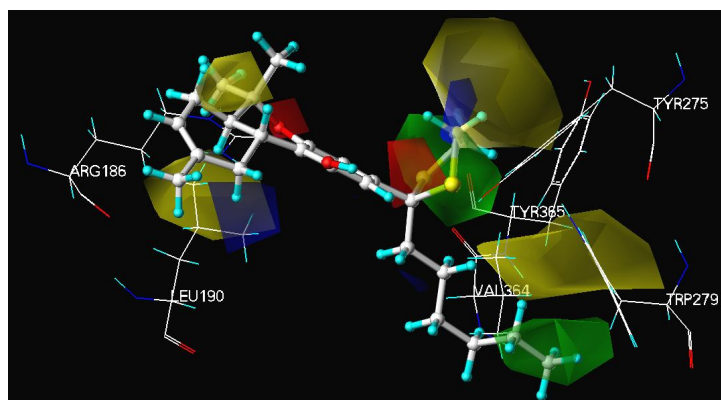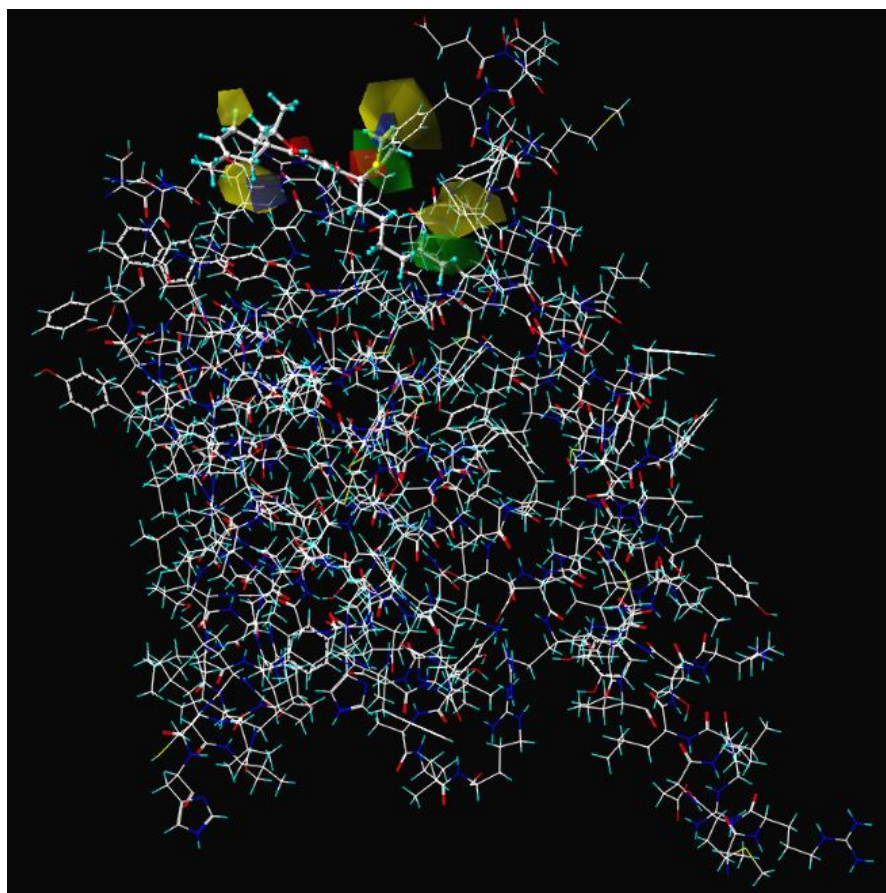
Figure 5.33 (top) Contour plots of CB1 model merged with CB1 receptor. **(bottom)** Close-look to the contour maps at the binding site of the CB1 receptor.

5.2.3 *De Novo* Drug Design Studies of CB Analogues

The optimal derived PLS analyses of CB analogues, which are produced from third generation of QSAR models, were used to generate each site points for CB1 and CB2 models. These site points were used in the *de novo* drug discovery program LeapFrog,
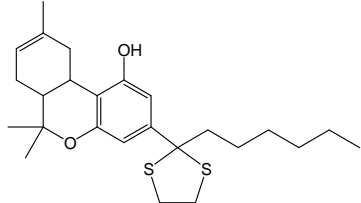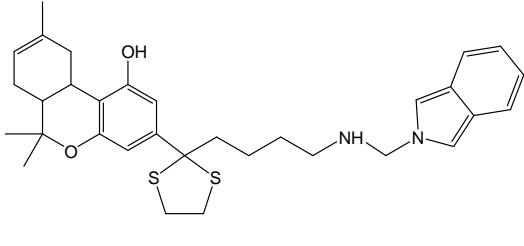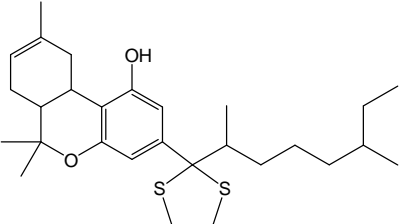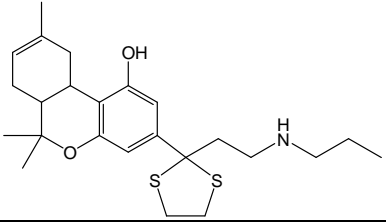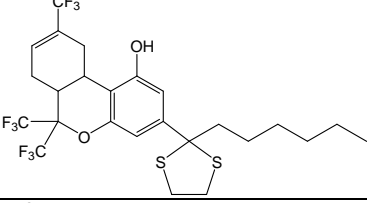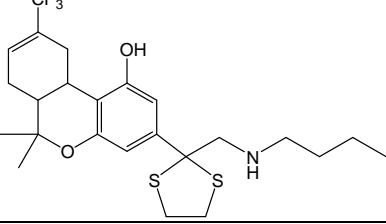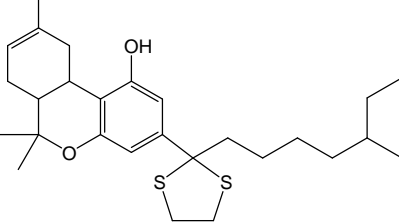
for the predictions of novel hits by repeatedly making structural changes and then either keeping or discarding the results. Since third generation of CB1 and CB2 models were derived based on different conformations of template conformers, a selectivity of new generated ligand structures by LeapFrog can be expected. The basic information about LeapFrog was explained in the "Computational Details" chapter of the thesis. The new ligand structures were evaluated on their binding energies, and structures that have better binding energy than reference compound (template compound, AMG3) were collected. Tables 5.17 and 5.18 show these structures for CB1 and CB2 receptors, respectively. Predicted binding affinities based on derived QSAR models were also included in the tables. Derived molecules that have highest predicted binding affinity for CB1 and CB2 receptors (**D1** and **D12**, Tables 5.17 and 5.18) were docked at the binding site of the CB1 and CB2 receptors, respectively (Figures 5.33 and 5.34). Their predicted high binding affinities confirmed by better docking scores than AMG3 (i.e., **D1** and **D12** have binding scores of -19.11 kJ/mol and -19.23 kJ/mol, respectively). The **D1** stabilizes its interactions with the binding site forming H-bonds with the amino acid residues (e.g., Arg186, Thr197 and Pro251) of CB1 as well as van der Waals interactions with the non-polar surfaces of the active site residues of CB1 receptor (e.g., Thr197, Phe200, Thr201, Ile247, Leu250, Pro251, Tyr275, Trp279, Thr283, Leu360). The **D12** stabilizes its interactions with the binding site forming H-bonds with the amino acid residues (e.g., Leu160, Leu163, Ser165, Tyr166, Leu167, Pro168) as well as van der Waals interactions with the non-polar surfaces of the binding site residues of CB2 receptor (e.g., Lys109, Leu160, Leu163, Pro168, Pro187, Tyr190, Trp194, Trp258).

5.2.4 Homology Modeling Studies of CB Receptors

If the X-ray structure of a ligand-bound receptor is not available, homology models of the protein can be used to obtain the ligand binding cavities. The steroelectronic properties of these cavities are directly related to the performed molecular model coordinates. Thus, the use of different template structures for homology may result in variation of ligand binding modes. In order to validate the obtained results using CB1 and CB2 receptor models based on bovine rhodopsin (1F88, pdb code), alternative CB1 and CB2 comparative models have been attempted employing template structure of β2-adrenergic receptor. The initial structure was taken from the cholesterol bound

form of human β2-adrenergic receptor (pdb code, 3D4S).[127] The water molecules and the cholesterol were removed from the system. Sequence alignment has been obtained with Biopolymer module of Sybyl.[27] CB1 and CB2 receptors show 28% and 24% sequence identity, respectively (Figure 5.35). Initial geometry optimization calculations have been carried out with Powell algorithm using Tripos force field.[27] Subsequently, these receptors have been subjected to 2 ns MD simulations using Gromacs.

Before the simulations, geometry optimization of receptors has been performed without constrains using steepest descent integrator for 10000 steps with the minimization tolerance of 100 kJ/(mol.nm). Cluster analysis of obtained coordinate file of trajectories has been performed with *g_cluster* module of Gromacs and each simulation yielded 8 clusters. Sequence alignment of representative of these clusters with the rhodopsin based receptor models has been performed with Accelrys DS 2.0 program[128] and models that have smallest RMSD values (using C^{α} atoms as the reference points) for CB1 and CB2 were used for further analysis. Obviously there are some fine differences between models based on rhodopsin and β2-adrenergic receptor for each receptor; however, the motifs of the seven member TM helices have been kept. It should be noted that the active site residues have smaller RMSD values (< 3Å) than the other amino acid residues. For clarity, superimposition of rhodopsin and β2-adrenergic based receptor models of CB1 has been shown at Figure 5.36. In Figure 5.37, the ligand binding pockets of CB1 have been shown with novel obtained model based on β2-adrenergic receptor; figure clearly confirms the two obtained binding pockets from the previous model based on rhodopsin.

Compound		CB1 model Predicted pK$_i$
Ref.		9.23
D1		9.73
D2		9.43
D3		9.39
D4		9.35
D5		9.32
D6		9.32
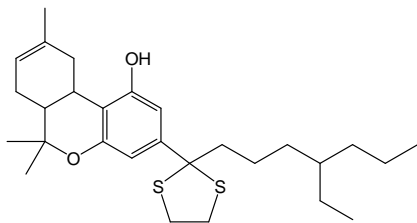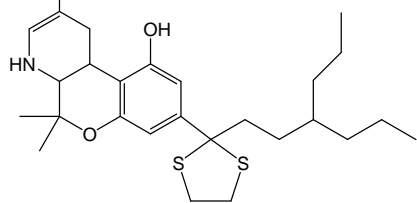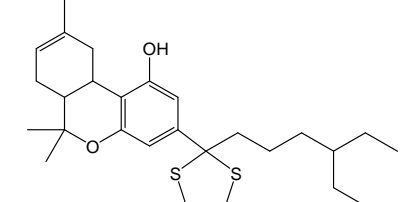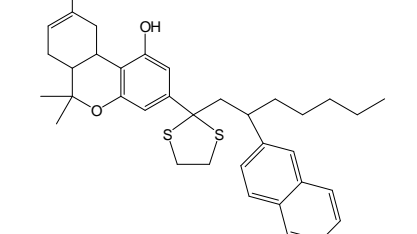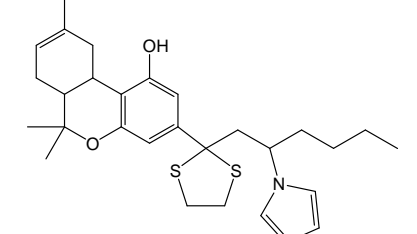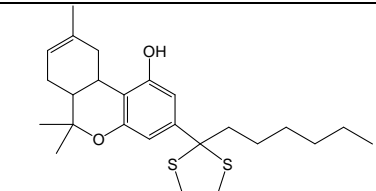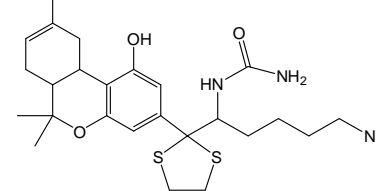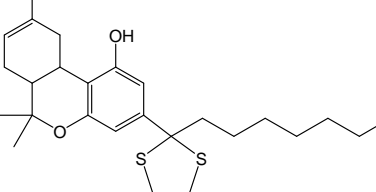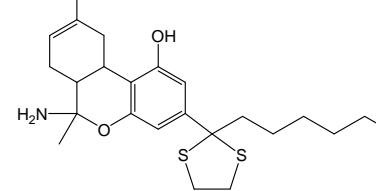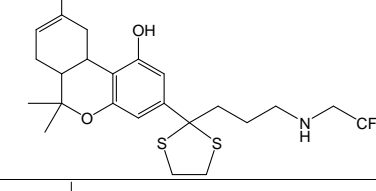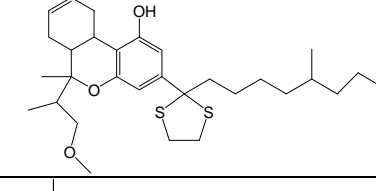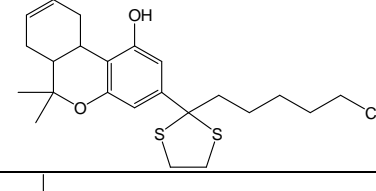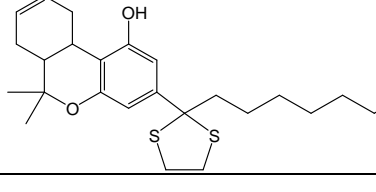
D7		9.31
D8		9.28
D9		9.28
D10		9.27
D11		9.24

Table 5.17 The proposed novel CB analogues by *de novo* drug design program Leap-Frog based on CB1 model and their predicted pK_i values for CB1 receptor.
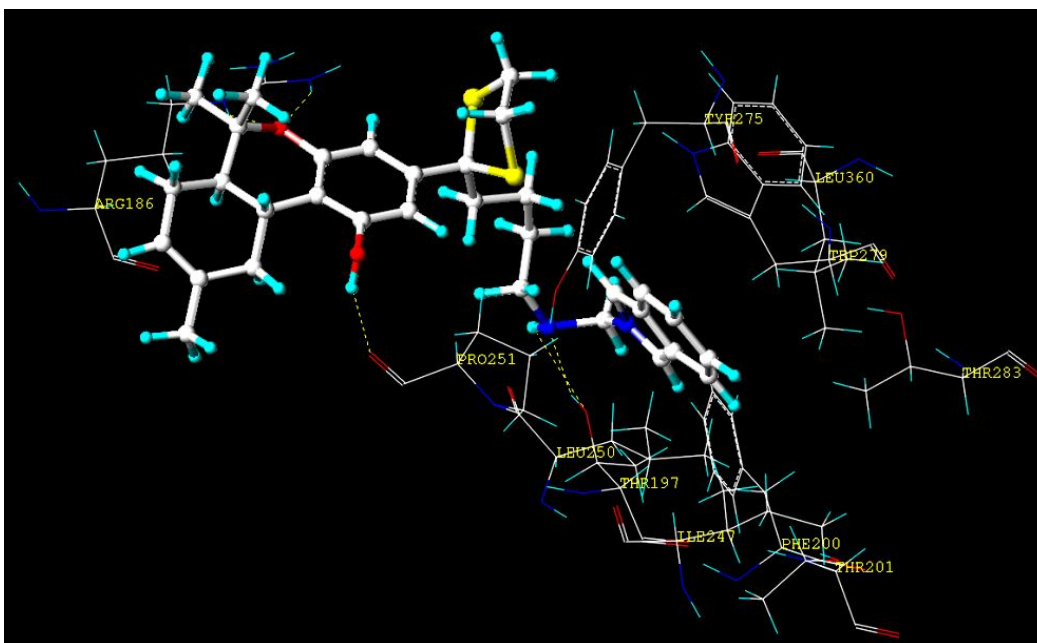
	Compound	CB2 model Predicted pKi
Ref.		9.03
D12		9.52
D13		9.42
D14		9.40
D15		9.39
D16		9.36
D17		9.34
D18		9.33
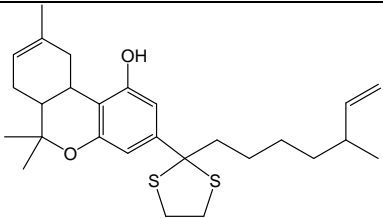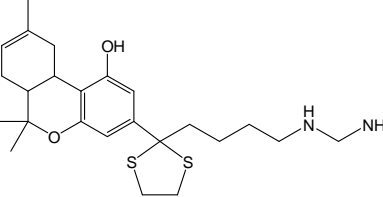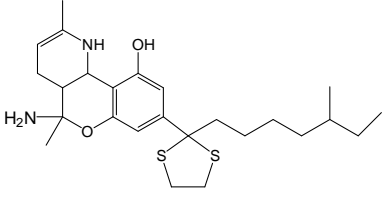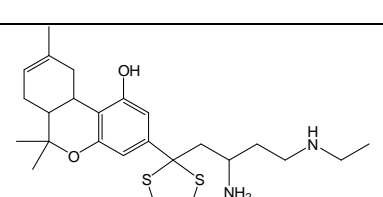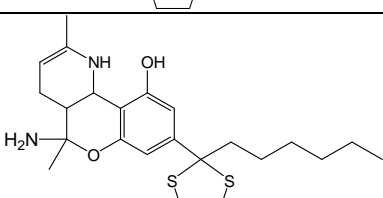
D19		9.26
D20		9.26
D21		9.25
D22		9.23
D23		9.15

Table 5.18 The proposed novel CB analogues by *de novo* drug design program Leap-Frog based on CB2 model and their predicted pK_i values for CB2 receptor.

Figure 5.33 Binding interactions of **D1** at the binding site of the CB1 receptor.
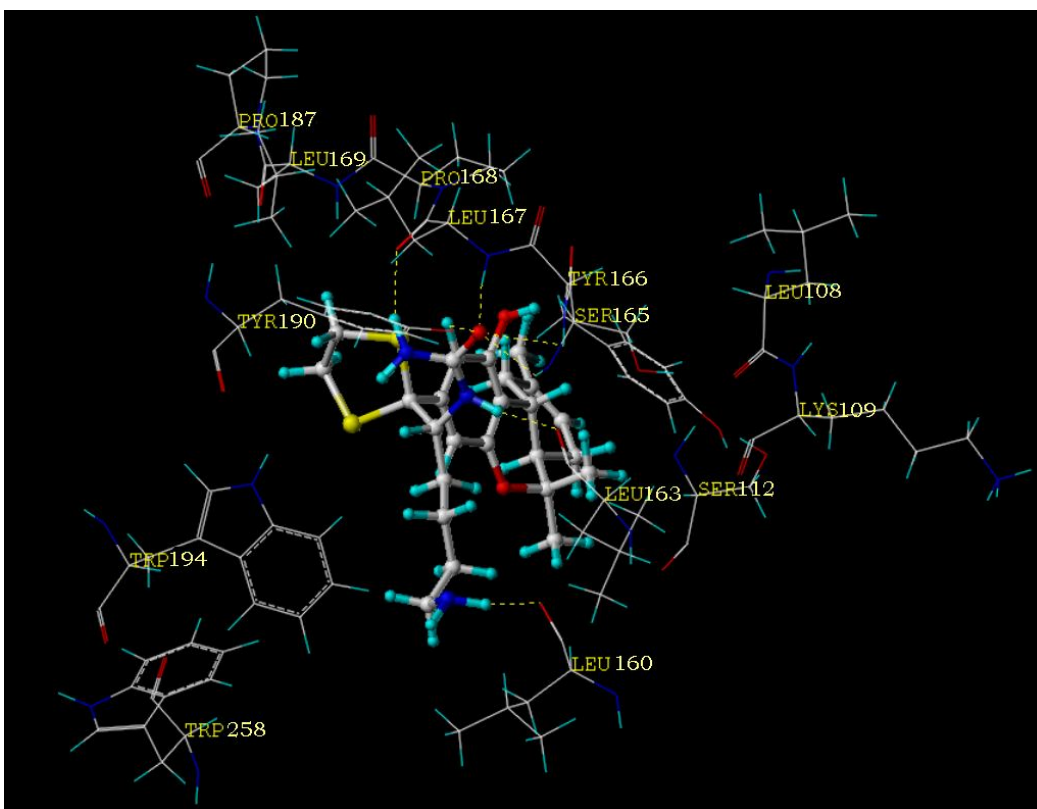
Figure 5.34 Binding interactions of **D12** at the binding site of the CB2 receptor.
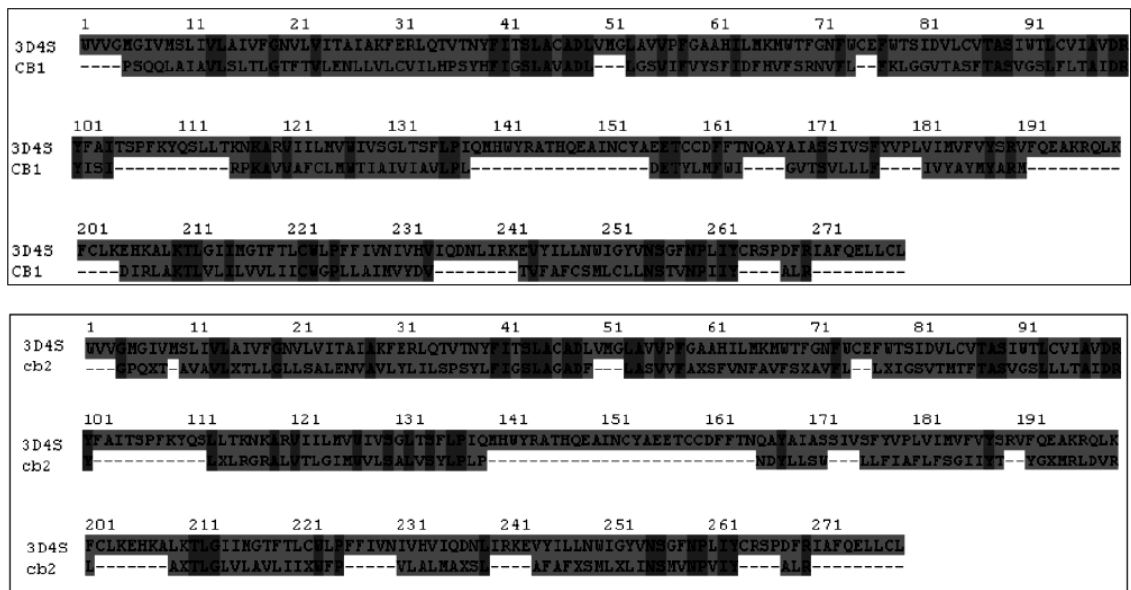
Figure 5.35 Sequence alignment of CB receptors based on β2-adrenergic receptor.
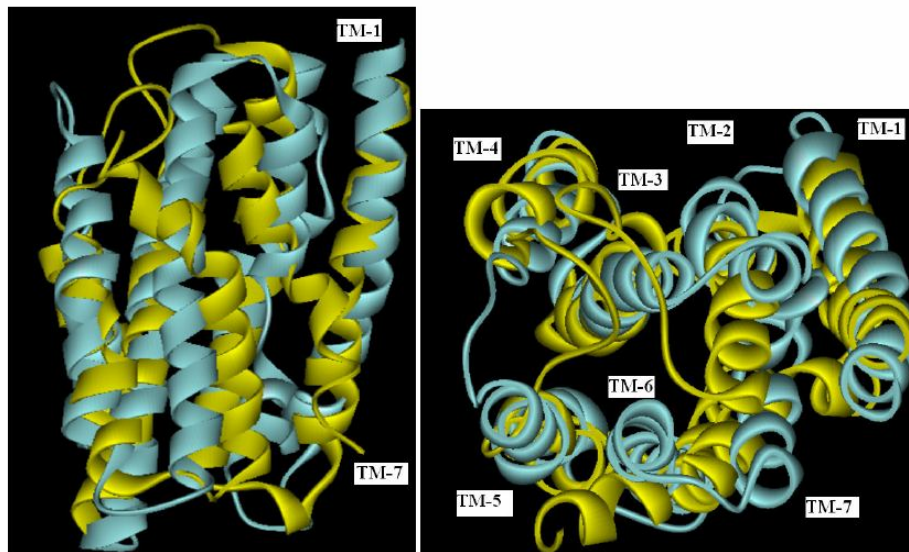
Figure 5.36 Superimposition of rhodopsin (yellow colored) and β2-adrenergic (cyan colored) based CB1 receptor models from both side and top views.
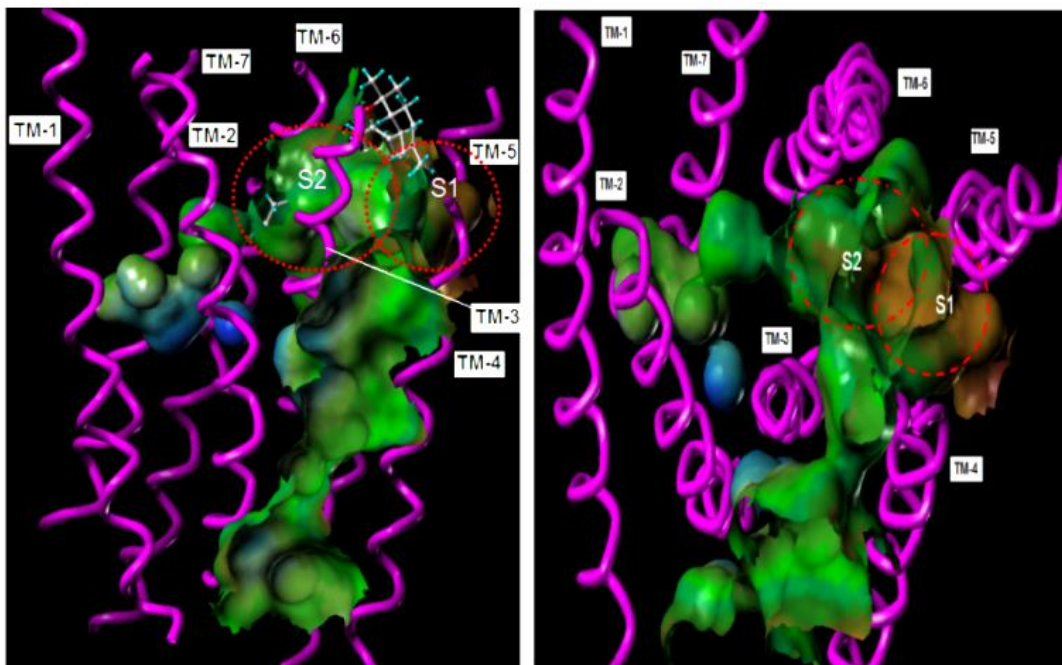
Figure 5.37 Two binding pockets (S1 and S2) were identified at the rhodopsin based CB1 model (left). CB1 receptor model obtained by using template of human β2-adrenergic receptor confirmed these two ligand binding pockets between the TM3-TM6 (right).

Chapter 6. Computational Design of Novel Fullerene Analogues as Potential HIV-1 PR Inhibitors: Analysis of the Binding Interactions between Fullerene Inhibitors and HIV-1 PR Residues Using 3D QSAR, Molecular Docking and MD Simulations
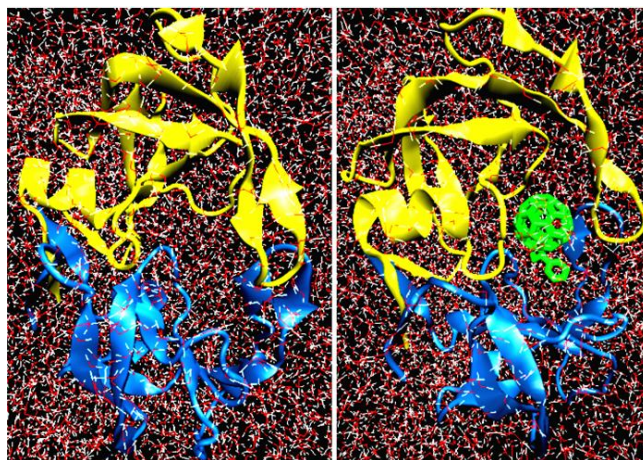
6. 1 Introduction

Inhibition of HIV-1 PR leads to the production of immature virus particles and pre-vention of further rounds of infection.[129,130] HIV-1 PR is an important target enzyme for anti-acquired immunodeficiency syndrome (AIDS) drug design as its inhibition leads to the production of non-infectious viral particles.[129] In the last few years, many potent and selective HIV-1 PR inhibitors have been developed and approved as drugs for the inactivation of this enzyme by the Food and Drug Administration, while sev-eral others are under clinical investigation.[129,131] Since the nature of most of these drugs is peptide-like, their oral bioavailability and half-life are limited.[132] Experimen-tal findings indicate the rapid emergence of drug resistance to most of the HIV-1 PR inhibitors, because site specific mutations in the enzyme occur at one or more resi-dues.[133] These mutations are conservative and involve a similar set of amino acid resi-dues in response to exposure to different inhibitors, thus giving rise to cross resis-tance.[133] Therefore, several research groups tried to develop non-peptidic HIV-1 PR inhibitors, which may block the mutations responsible for this resistance.[11,134]

In the last decade, fullerene and its derivatives have been extensively investigated for biomedical applications. Inhibition of HIV-1 PR by fullerene analogues demonstrated by Friedman *et al.*[11] and complexation of HIV-1 PR with fullerene compounds has been supported by molecular modeling studies.[13] These studies showed that the fullerene can be perfectly accommodated inside the binding pocket of HIV-1 PR (Figure 6.1). The active site of the HIV-1 PR is approximately an open-ended cylin-drical hydrophobic cavity with 10 Å diameter composed of catalytic aspartic acid residues Asp25 and Asp25'.[11,13] The complementary spatial relationship between [60]fullerene and the active site of the HIV-1 PR enzyme has led to the suggestion that fullerene-based derivatives could have potential use as effective HIV-1 PR inhibi-tors.[11,13] The binding interactions of [60]fullerene derivatives in the active site of the HIV-1 PR have been examined through a combination of several molecular modeling techniques.[13,134] Kinetic analysis of the HIV-1 PR enzyme in the presence of various water-soluble fullerene derivatives suggests a competitive mode of action.[13] This is attributed to the ability of fullerene derivatives to form bonds with the catalytic site and the van der Waals interactions with the non-polar HIV-1 PR surface thereby, im-

proving the binding.[134] Since the binding affinity values of "first generation" fullerene inhibitors were not significant (EC_{50} ~10^{-6} M), further structural investigation is required in order to propose new HIV-1 PR/fullerene complexes with better binding affinity. For this aim, a set of synthetically reported fullerene derivatives (Table 6.1)[135,136] have been used to construct models with the 3D QSAR methodologies: CoMFA[9] and CoMSIA[10].
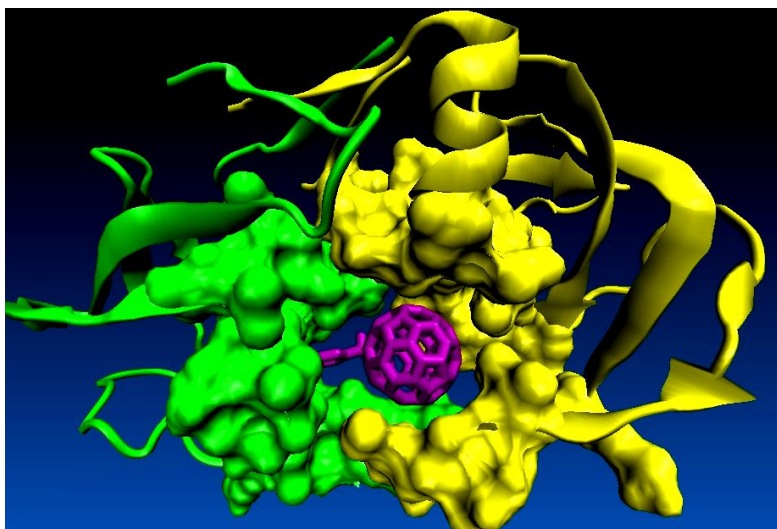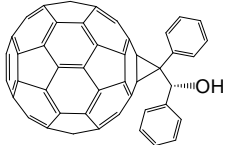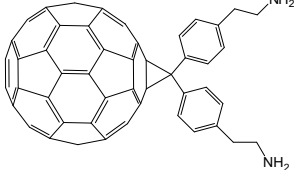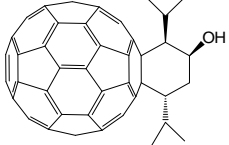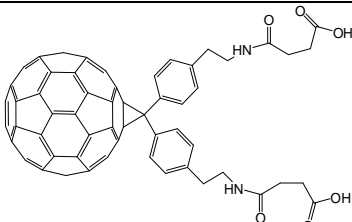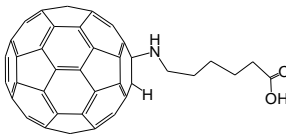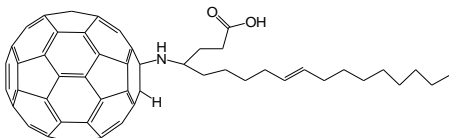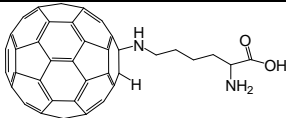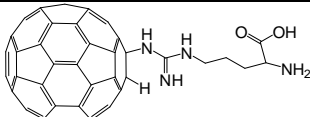

Figure 6.1 Perfect fit of a fullerene derivative at the active site of the HIV-1 PR (active site residues have been shown as molecular surface for clarity).

6.2 Results and Discussion

In order to improve the structure alignment and derive statistically reliable models, initially, the most potent fullerene analogue in the data base (compound **1**, Table 6.1) has been docked in the binding cavity of 3D model of HIV-1 PR (hitherto X-ray structure of a fullerene ligand-bound HIV-1 PR is not available, thus, HIV-1 PR structure has been received from the X-ray structure of haloperidol-bound HIV-1 PR (pdb code, 1AID)[35]. The derived best docked complex structure has been subjected to MD simulations, in order to stabilize the localization of fullerene at the HIV-1 PR. 3D-QSAR/CoMFA and CoMSIA methods were applied to the data set, which was divided into training and test sets. To our knowledge, this study is the first 3D QSAR application to the fullerene based compounds. Both CoMFA and CoMSIA studies gave similar results indicating that the steric effects are essential for the activity; which is reasonable because of high non-polar property of compounds at the data set.

The *de novo* design Leapfrog routine of Sybyl was used as an aid for the discovery of new molecules based on the 3D QSAR results.

Comp. No.	Compound	Exp. Binding Energy (kJ/mol)	Calculated Binding Score (kJ/mol)	Exp. Binding Affinity (μM)
1		-40.1	-35.2	0.1
2		-30.4	-29.1	5.0
3		-39.2	-29.2	0.15
4		-29.5	-31.5	7.3
5		-36.2	-35.6	0.49
6		-16.6	-24.8	1300
7		-23.7	-32.3	75
8		-20.9	-28.1	230

9		-22.1	-20.1	140
10		-32.2	-27.4	2.50
11		-34.7	-36.0	0.9
12		-31.8	-33.0	7.3
13		-32.5	-33.5	2.2
14		-29.9	-31.4	6.3
15		-29.5	-32.9	2.9
16		-26.8	-27.1	21.7
17		-22.2	-26.8	137
18		-27.3	-29.3	17.6

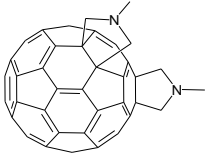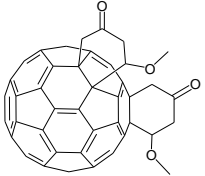|				
19		-23.8	-24.6	72.7
20		-29.4	-18.4	7.70

Table 6.1 Chemical structures, measured affinities and binding energies and calculated binding scores of fullerene derivatives.

After obtaining the optimum position of most potent fullerene analogue **1** inside the HIV-1 PR, 3D QSAR/CoMFA and CoMSIA methods have been performed. The logarithmic $1/EC_{50}$ values (pEC_{50}) were used in the 3D QSAR correlations, as they are related to changes in the free energy of binding. Several variations in the alignment procedures are considered by superimposing similar pharmacophoric features. Highlighted carbon atoms (32 central carbon atoms of fullerene) for the template ligand **1** are selected for the structural superimposition processes (Figure 6.2i). The alignment of the molecules was based on atom-by-atom superimposition of selected atoms, which are common in all compounds. Figure 6.2ii illustrates the superimposition of the molecules in the training set. Compounds **5**, **16** and **17** have been used as a test set. They possess higher, similar, and lower inhibition effects than average pEC_{50} values of the studied compounds which represents the whole data set. The cross validated PLS method was then subjected to the training set. Table 6.2 summarizes the statistical results.
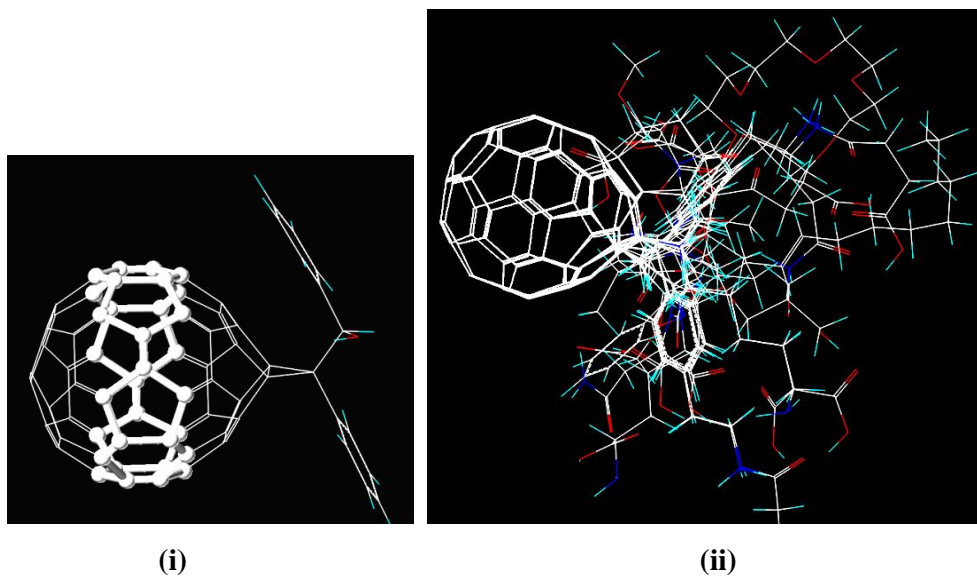
(i)   **(ii)**

Figure 6.2 (i) Selected atoms of the template compound **1** for structural superimposi-
tions of the compounds in training set. **(ii)** Structural alignments of the compounds in
the training set for constructing 3D QSAR/CoMFA and CoMSIA models at HIV-1
PR receptor.

	CoMFA	CoMSIA
r^2_{cv}	0.549	0.555
r^2	0.994	0.997
Standard error of estimate	0.097	0.051
F	509.545	1216.442
Relative contributions of steric/electrosatic fields	0.776:0.224	0.872:0.128
Number of optimal components	4	5

Table 6.2 Cross-validated analyses using CoMFA and CoMSIA methodologies,
based on template compound **1**.

3D QSAR/CoMFA and CoMSIA studies gave r_{cv}^2 values of 0.549 and 0.555, respectively. The non-cross validated PLS analysis yielded r^2 values of 0.994 and 0.997, respectively. Figure 6.3 shows the relationship between the experimental and predicted pEC_{50} values of cross-validated PLS analysis for CoMFA and CoMSIA. The significance of the proposed models is verified by the good predictions of the activity of compounds belonging to the test set. Test set compounds **16** and **17** both in CoMFA and CoMSIA gave good prediction results (Table 6.3). Compound **5** is underestimated by CoMFA about 1.0 unit but CoMSIA result gave satisfactory estimation of its activity (error on prediction is less than a unit).

Compound	pEC_{50} (observed)	CoMFA pEC_{50}(predicted)	CoMSIA pEC_{50} (predicted)
1	7.00	7.05	6.95
2	5.30	5.25	5.31
3	6.80	6.65	6.84
4	5.13	5.22	5.11
5	6.31	*5.22*	*5.76*
6	2.89	2.84	2.97
7	4.12	4.03	4.08
8	3.64	3.69	3.56
9	3.85	3.92	3.80
10	5.60	5.52	5.62
11	6.05	5.96	5.94
12	5.53	5.59	5.61
13	5.66	5.83	5.66
14	5.20	5.14	5.16
15	5.14	5.23	5.23
16	4.66	*4.68*	*4.67*
17	3.86	*3.46*	*3.67*
18	4.75	4.64	4.73
19	4.14	4.17	4.15
20	5.11	5.17	5.18

Table 6.3 Summary of experimental (observed) versus CoMFA and CoMSIA-predicted pEC_{50} results for the binding affinity of fullerene derivatives at the HIV-1 PR.

The CoMFA and CoMSIA contour maps were used in order to visualize the steroelectronic requirements of the binding cavity of HIV-1 PR. Figures 6.4 and 6.5 show the steric-electrostatic contour maps of the CoMFA and CoMSIA models (for the compounds that show the best and worst inhibition effects within the data set for the HIV-1 PR receptor, compounds **1** and **6** in Table 6.1, respectively).
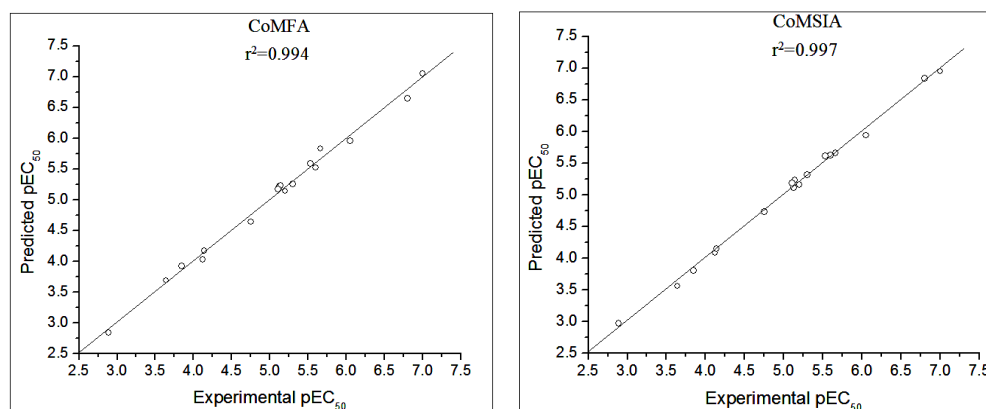
Figure 6.3 Plots of observed and 3D QSAR/CoMFA (left) and CoMSIA-predicted (right) binding affinities of fullerene analogues in the training set at the HIV-1 PR.

Template compound **1** has better inhibition effect than **6**, around 10^4-fold. This can be explained by different topographical requirements of **1** and **6**; both aromatic rings of **1**, fit very well with the green colored contour map which shows sterically favorable places, there are no bulky groups of **6** in this field (Figure 6.4). In addition, inactive compound **6** has a flexible chain and fits with the yellow colored contour which shows sterically unfavorable areas at the CoMSIA model (Figure 6.4). Moreover, electrostatic contour maps of **1** and **6** clearly show the differences both in CoMFA and CoMSIA models. For example, the -OH group of **1** fits on red colored contour (negative charged favored) while, -C=O group of **6** fits on blue colored contour (positive charged favored), (Figure 6.4). Relative contributions of steric and electrostatic fields obtained by CoMFA and CoMSIA are 0.78:0.22 and 0.87:0.13, respectively.
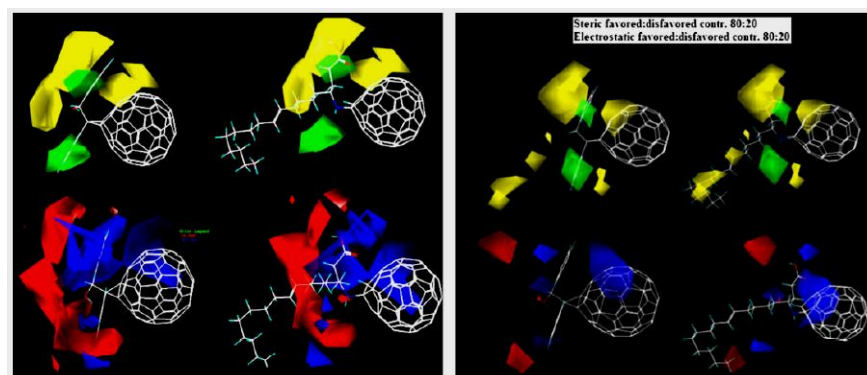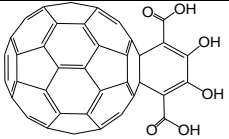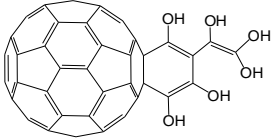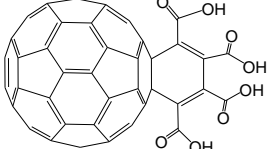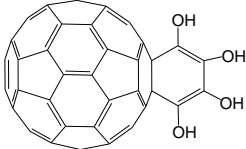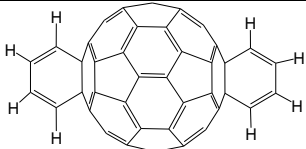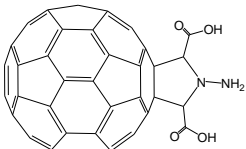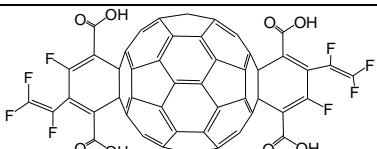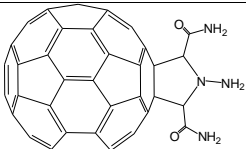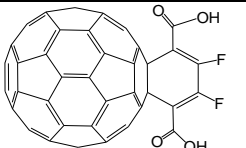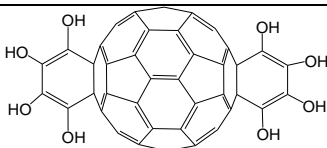

Figure 6.4 (Left) CoMFA contour maps of template compound **1** (shown in top-left and bottom-left which has highest binding affinity in the training set) and compound **6** (shown in top-right and bottom-right which has lowest binding affinity in the training set). (Right) Corresponding CoMSIA contour maps for compounds **1** and **6**.

Hitherto, a lot of fullerene derivatives have been synthesized by several experimental groups but a limited number of them have been subjected to bioactivity test against HIV-1 PR. Obviously, when the number of compounds increases in the data set, the stability of the constructed model is expected to be increased. However, in this particular case, it is difficult to increase the number of compounds in the data set because the experimentally examined fullerene compounds for HIV-1 PR inhibition are limited and measured ones do not show very diverse binding affinities. For this reason, an attempt has been performed in QSAR studies: The binding energy results of biologically evaluated fullerenes in HIV-1 PR cavity was examined with docking simulations and a correlation between experimental versus theoretical values has been found. Thus, it is proposed to enlarge the data set including the structures from computationally designed compounds and their activities from the docking simulations to construct QSAR models. This idea may be used by other research groups if the experimental data are limited and/or not diverse, the coordinate file of the receptor (using X-ray data or homology model) is available and the active site of the receptor is well defined.

The large variations in binding affinities of the designed fullerene inhibitors with HIV-1 PR and the relations between biological activity and the flap motion of the enzyme, as well as, the connection between the biological activity and the conformational changes in the catalytic site of the HIV-1 PR, were also analyzed. The following three steps in our computational design strategy have been studied: (i) Initially, several monoadducts and bisadducts of [60]fullerene have been designed with some modifications of reported structures in the literature (Table 6.1) in order to explore more conformational space. Most of the designed fullerene structures include 1,3-cyclohexadiene derivatives. The experimental methodology to synthesize this kind of fullerene derivatives was reported by An *et al.*[137] (ii) In order to diminish the stability problems, CoMSIA models with smoother potential functions have been used in further 3D QSAR investigations of the fullerenes. Novel monoadducts and bisadducts of [60]fullerene have been designed with the aid of 3D QSAR/CoMSIA models and their binding affinities at the HIV-1 PR have been tested employing molecular docking. In order to use proper input coordinates of HIV-1 PR/fullerene derivative complex in the docking simulations, MD simulations were employed by Gromacs program.[36] In order to understand the effect of fullerene ligand to the conformational changes of amino

acid residues of active site of HIV-1 PR, both ligand-free and ligand-bound systems were used in the simulations (Figure 6.5). The flexibility of flap region of HIV-1 PR has been discussed by both experimental and computational studies.[138,139] The free protease adopts closed, semi-open or fully open forms in a dynamic equilibrium. However, the closed form is favored when the inhibitor is bound in the cavity of the reaction site.[138] Friedman et al.[11,13] proposed that there is a direct correlation between inhibitory of a compound and the amount of hydrophobic surface area that it can desolvate. The MD simulations by Zhu et al.[140] showed the exclusion of water near the flap regions in order to accommodate the fullerene inhibitor. The decreasing of the water density in the cavity leads to the enhancement of the hydrophobic interaction between the fullerene derivative and the active site of the enzyme. (iii) The Leapfrog de novo drug design program and 3D QSAR/CoMSIA contour maps have been used in order to generate a series of potent fullerene based HIV-1 PR inhibitors. This approach is fruitful and can be of general use because: (a) it minimizes the compounds to be synthesized; (b) it leads a rational design for the subsequent compounds to be synthesized and (c) it saves time and reduces the cost and human effort.

The inhibition effect of limited monoadducts and bisadducts of [60]fullerene has been biologically evaluated. Table 6.1 shows the biological activities of 20 fullerene derivatives reported in the literature.[13,135,136,143] In order to explore more conformational space and include more diverse binding affinities in the data set, computationally designed fullerene analogues (Table 6.4) have been analyzed using docking studies. The formula $\Delta G = -RT\ln K_i$ was used to convert the FlexX binding energies to estimated binding affinities. Since the experimental binding activities of most of the derivatives, which are of interest in this study only reported as median effective concentration (EC_{50}), these values are assumed to be equal with K_i in the calculations of the free binding energies. A similar working hypothesis has been used by Naik et al.[141] and Conn et al.[142] The reported K_i and EC_{50} values for compound **4** are 5.3 μM and 7.3 μM, respectively. Assuming that K_i and EC_{50} are identical, the resulting error for ΔG of compound **4** is 2.6 %.

Compound No.	Compound	Calculated Binding Score (kJ/mol)	Estimated Binding Affinity (μM)
21		-37.9	0.25
22		-41.9	0.051
23		-50.3	0.002
24		-35.3	0.71
25		-19.1	472
26		-35.6	0.63
27		-33.4	1.53
28		-35.5	0.66
29		-38.4	0.21
30		-27.0	19.9

151

31		-27.4	16.9
32		-31.6	3.15
33		-37.6	0.28
34		-23.2	91
35		-24.3	58.7
36		-12.9	5670
37		-19.3	436
38		-32.9	1.87
39		-26.5	24.3
40		-17.6	862
41		-42.7	0.04
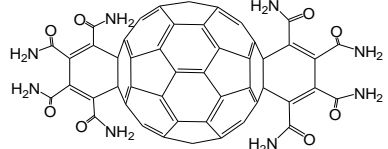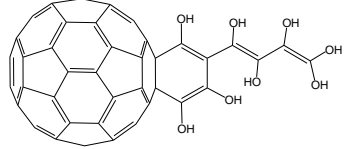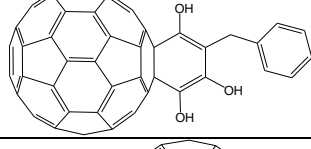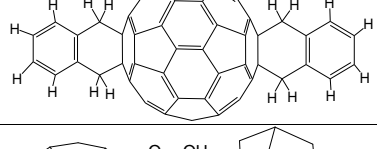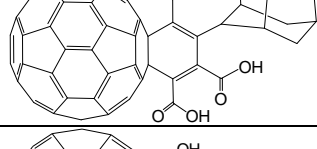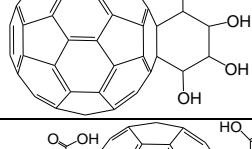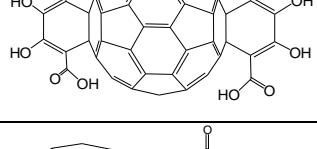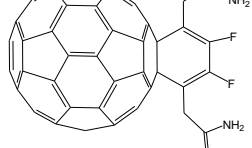
42		-42.6	0.04
43		-38.8	0.18
44		-32.2	2.47
45		-23.7	75
46		-40.2	0.10
47		-35.3	0.71
48		-30.0	5.98
49		-34.9	0.84

Table 6.4 Computationally designed fullerene derivatives, their binding scores and estimated binding affinities.

A deeper understanding of the mechanistic events associated with HIV-1 PR binding is important for the design of new inhibitors with enhanced activity. In the docking calculations the FlexX docking algorithm was used which considers docking to a rigid protein. It is well known that, the conformations of binding pocket residues of a pro-

tein affect the binding score. The coordinates of active site residues of fullerene bound HIV-1 PR may be different than the coordinate of haloperidol derivative bound HIV-1 PR. Hence, the correct binding mode of the studied inhibitors can best be obtained from the MD simulations of HIV-1 PR/fullerene derivative complex. The most potent reported compound in Table 6.1 (compound **1**) was docked in the binding cavity of HIV-1 PR and docking has been employed using the FlexX docking program. The coordinate file of the best FlexX molecular docking pose (which is associated with the biggest -in absolute value- binding energy) is used in the MD simulations of fullerene-bound HIV-1 PR. The average coordinate file of HIV-1 PR from the final 1 ns trajectory files of MD simulations of complex system has been used in the re-docking calculations.
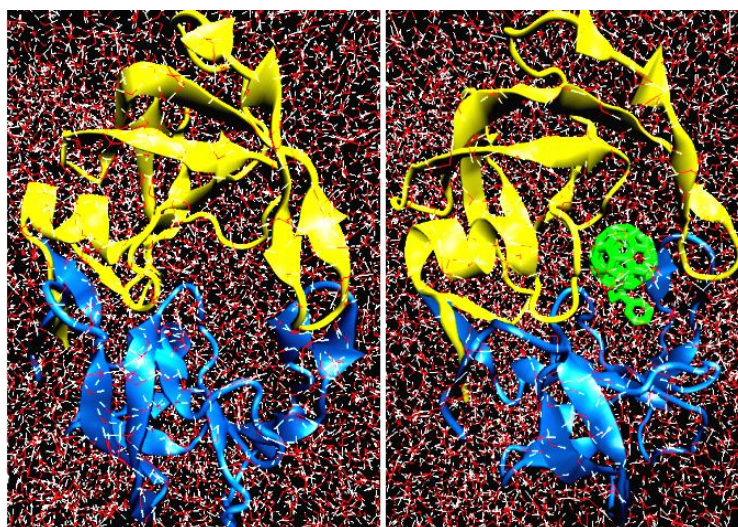
Figure 6.5 Two different systems were used in MD simulations: **(i)** A ligand-free rectangular box with HIV-1 PR and solvent (water) molecules, (left on the figure), and **(ii)** the rectangular box includes fullerene analogue at the binding site of the enzyme and solvent (water) molecules, (right on the figure).

HIV-1 PR is a C2 symmetric homo dimer with a large substrate binding pocket covered by two glycine-rich β-hairpins, or flaps.[139] Consistent structural differences have been found between the free and bound systems of the HIV-1 PR (average structures from simulation have been considered; in Figure 6.6, initial forms have been shown with turquoise color and average structures from MD simulations have been shown with red color). In the fullerene inhibitor bound forms, the flaps are pulled in toward

the dual Asp25 catalytic site (the closed form), while the structure for the free HIV-1 PR adopts a semi-open form with flaps shifted away from the catalytic site, however, still substantially closed to the active site. The relative orientation of the β-hairpin flaps is reversed in the two forms (Figure 6.6). The results are in agreement with the MD simulations of non-fullerene based HIV-1 PR inhibitor.[139]

Changes of the HIV-1 PR binding cavity for the free and fullerene-bound systems were screened by employing the change of distance between C^α atoms of the two catalytic residues, Asp25 and Asp25', as well as, the change of distance between C^α atoms of the residues at the flap region (Gly48, Gly49, Ile50, Gly51 and Gly52 amino acid residues in chain A and their corresponding residues in chain B). In the ligand-free system, only some perturbations around the initial distance have been found for the distance between C^α atoms of Asp25 and Asp25'; Gly48 and Gly48'; Gly51 and Gly51'; Asp25' and Gly49'; Gly52 and Gly52', however, distances between the C^α atoms have been found increased ~4.0 Å between Asp25 and Gly49; ~1.0 Å between Gly49 and Gly49'; and ~0.5 Å between Ile50 and Ile50' through the simulations (Figure 6.7i). In the fullerene-bound system a detectable decrease in distances has been observed during the simulation with some perturbations at certain intervals (Figure 6.7ii). For example, distances between C^α atoms have been decreased ~1.5 Å between Asp25 and Asp25'; ~0.5 Å between Gly49 and Gly49'; ~2.0 Å between Asp25' and Gly49; ~2.0 Å between Gly48 and Gly48'. Distances between Ile50 and Ile50'; Gly51 and Gly51'; Gly52 and Gly52' did not change significantly throughout the simulations (Figure 6.7ii). A small increase in the distance (~1.0 Å) has been observed between the C^α atoms of the amino acids Asp25 and Gly49. Therefore, fullerene inhibitor tends to keep HIV-1 PR cavity in a closed form.

The difference of behavior at the flap region for these systems leads to two different orientations. It has been observed that flap loop of chain A lies perpendicular to that of chain B in both the fullerene bound and ligand free systems. However, in contrast to the ligand free system, chains orient toward the catalytic dual Asp25 residues in fullerene bound system.
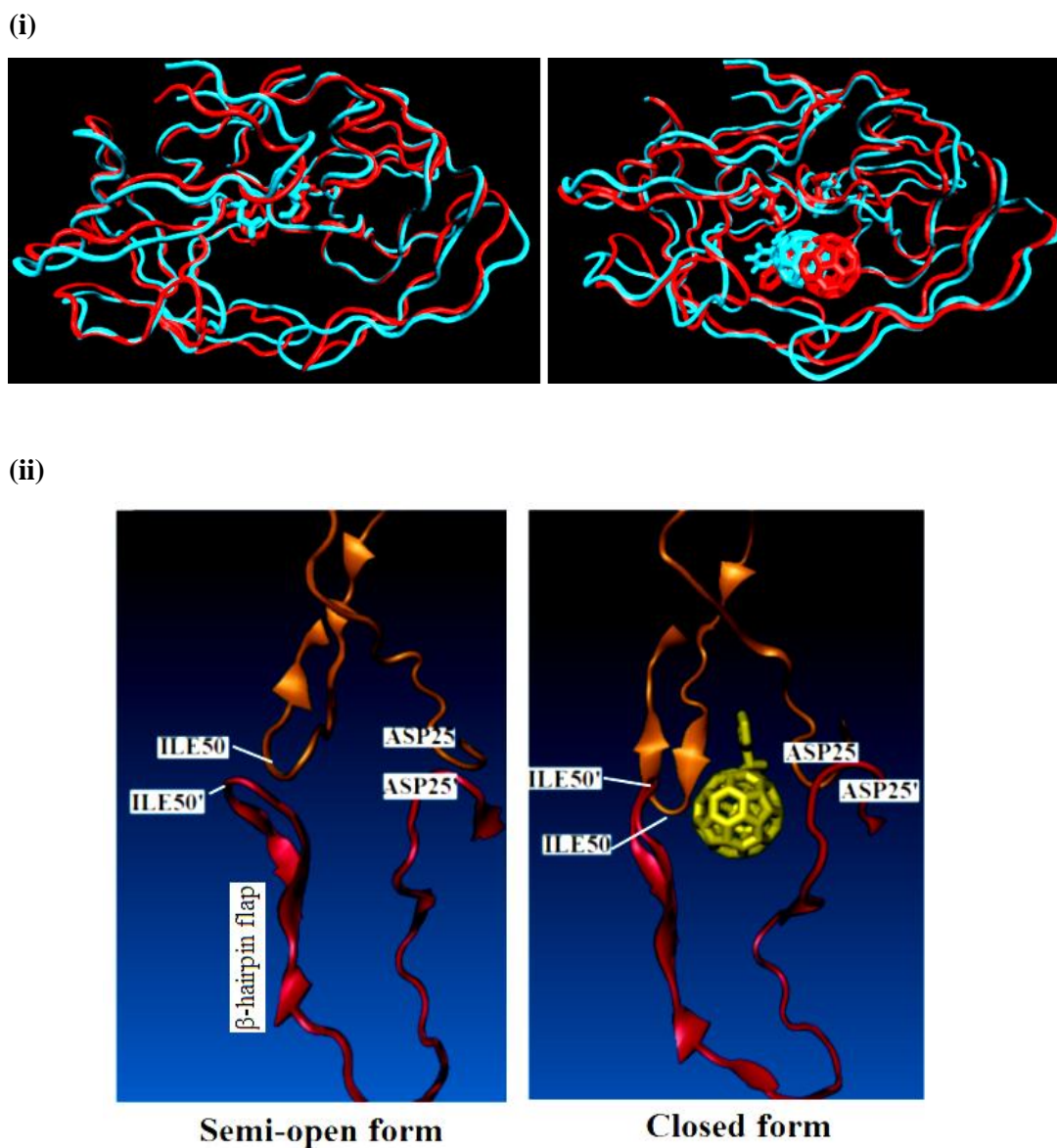
(i)

(ii)

Semi-open form   Closed form

Figure 6.6 (i) In the fullerene inhibitor-bound forms, the flaps are pulled in toward the bottom of active site (the closed form), while the structure for the free HIV-1 PR adopts a semi-open conformation with flaps shifted away from the dual Asp25 catalytic site, however still substantially close to the active site. In figure, initial forms are shown with turquoise color and average structures of simulation are shown with red colors. For clarity, only heavy atoms of the fullerene have been shown in the average structure of the simulation; **(ii)** A side-view of binding cavity of HIV-1 PR (from average structures of MD simulations of free (left) and inhibitor bound (right) systems) illustrates the semi-open and closed forms.

The torsional angles (OD2-C^γ-C^β-C^α) of the catalytic dyad (Fig. 6.8i) were also investigated for the ligand-free and ligand-bound systems. The trajectory analysis of the defined torsional angles of Asp25 and Asp25' for ligand free HIV-1 PR have shown that they keep their value mainly at ~140°/~300° and ~100°/~300°, respectively. (Figure 6.8ii). These values are ~200° for Asp25 and ~280° for Asp25' in fullerene-bound HIV-1 PR (Fig. 6.8iii).

Atom positions between the crystal (1AID) and the average structure from MD have been compared for both ligand-free and ligand-bound systems in order to understand which parts of the receptor are more stable and which are floppy. For this purpose, a script with color scale has been used under VMD program (version 1.8.6)[40] where blue colored places show no change in distance (stable) and red colored places show floppy areas (Figure 6.9). As it is clearly shown, flap regions, catalytic part, and termino-lateral fields of HIV-1 PR at both ligand-free and fullerene-bound systems show more flexible conformations throughout the simulations.

After obtaining the average structure of HIV-1 PR from MD simulations, computed binding scores of the synthetic derived fullerene analogues (Table 6.1) from docking simulations were compared with their corresponding experimental results (Table 6.1). The linearity of the plot (r^2=0.69, N=19, compound **20** used as outlier) shows a good correlation between computed binding scores and experimental binding energies, (Figure 6.10).

Tables 6.1 and 6.4 show the list of fullerene derivatives used in molecular docking studies, their binding energies and their estimated binding affinities from computed binding energies. In order to design novel fullerene derivatives with high inhibition effect of HIV-1 PR, 3D QSAR/CoMSIA models were employed. As it is mentioned above, binding affinity results from experimental measurements of fullerene-based inhibitors at the HIV-1 PR are limited. Since experimental and computed binding energies showed a good correlation, both experimental binding affinities of structures at Tables 6.1 and estimated binding affinities at Table 6.4 have been used to form 3D QSAR models. The used structures have a wide variation of biological activity (~10^6-fold variances in binding affinity). 43 molecules were used as the training set and the

rest 6 derivatives (compound numbers: **16**, **27**, **34**, **39**, **43** and **48**, Tables 6.1 and 6.4) were used as test set at the CoMSIA analysis. The test set includes compounds representing all categories of activity of the training set, e.g., inactive and active compounds comprising all structural features that are important for the activity. Among the analogues at Tables 6.1 and 6.4, compound **23** was selected as a template, because it has the highest binding affinity at the HIV-1 PR in the training set. Several variations in the alignment schemes are considered by superimposing the similar pharmacophoric features. Highlighted carbon atoms (32 central carbon atoms of fullerene) for the template ligand **23** are selected for the structural superimposition processes (Figure 6.11). The alignment of the molecules was based on atom-by-atom superimposition of selected atoms, which are common in all compounds. Figure 6.12 illustrates the superimposition of fullerene analogues used as the training set to construct CoMSIA models. The cross validated PLS method was then applied to the training set. Table 6.5 summarizes the statistical results.

Five different combinations of steroelectronic fields of 3D QSAR/CoMSIA models were obtained from the set of biologically evaluated and computationally designed fullerene derivatives (training set=43, test set=6) in order to predict novel compounds with improved inhibition effect. The best 3D QSAR/CoMSIA model (CoMSIA-4: steric/electrostatic/H-bond donor/H-bond acceptor) yielded a cross validated r^2 value of 0.739 and a non-cross validated r^2 value of 0.993. Using the neutral and ionized states of the fullerene analogues in data set did not affect significantly the constructed QSAR models. Thus, neutral states of fullerene derivatives were used in the further analyses. The selected 3D-QSAR/CoMSIA model (CoMSIA-4) for the estimated binding affinities of fullerene analogues at the HIV-1 PR has a good cross validated correlation. Figure 6.13 shows the relationship between the estimated pEC_{50} values from FlexX binding energies and CoMSIA-predicted results of the non-cross-validated analyses for the HIV-1 PR. Linearity of the plot shows very good correlation for the CoMSIA model. Table 6.6 summarizes the estimated binding affinities from molecular docking results and CoMSIA-predicted pEC_{50} results for the fullerene derivatives at the HIV-1 PR. A good correlation was observed in CoMSIA of the fullerene derivatives as it is demonstrated by the very high values of r^2. Additionally, the credibility of the models is evidenced by the high values of r_{cv}^2. However, the real
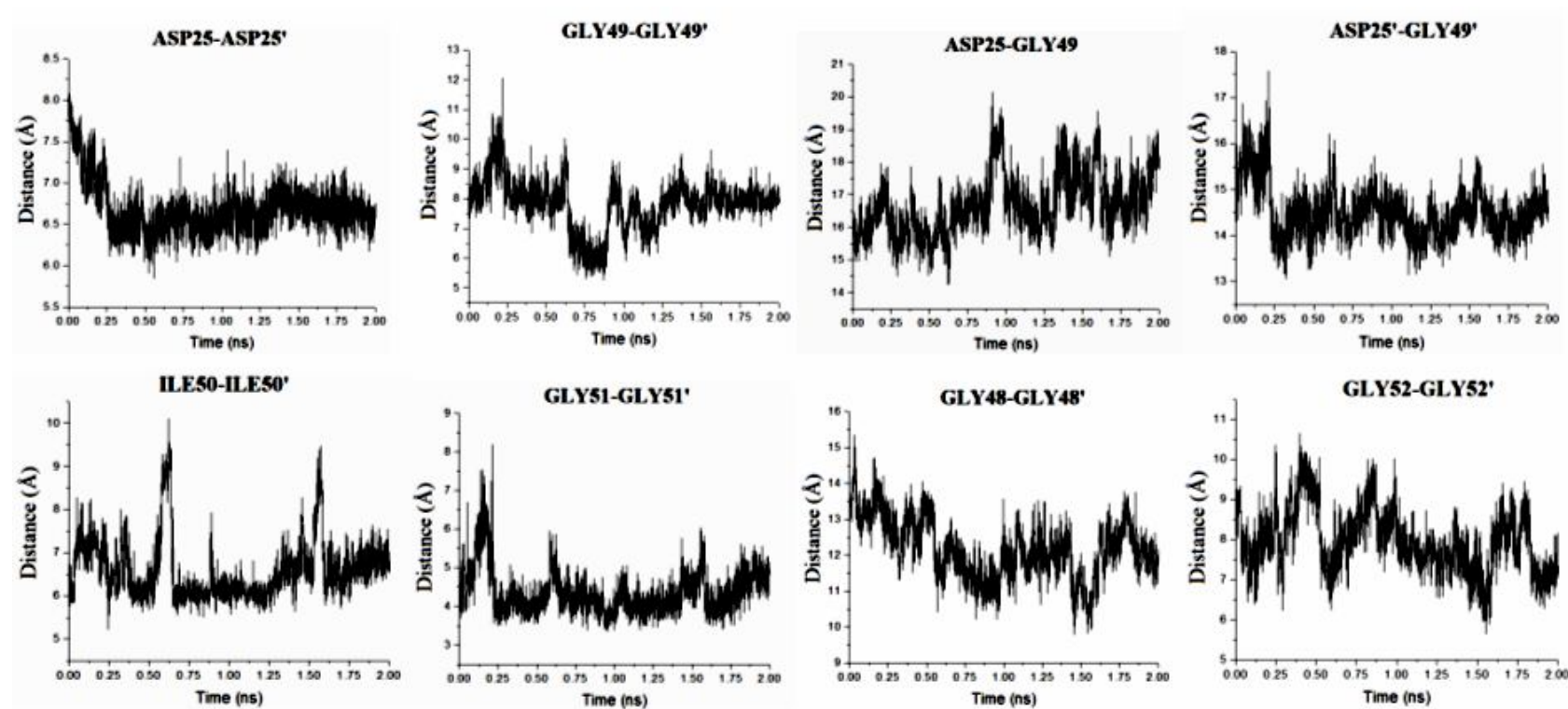
(i)

159
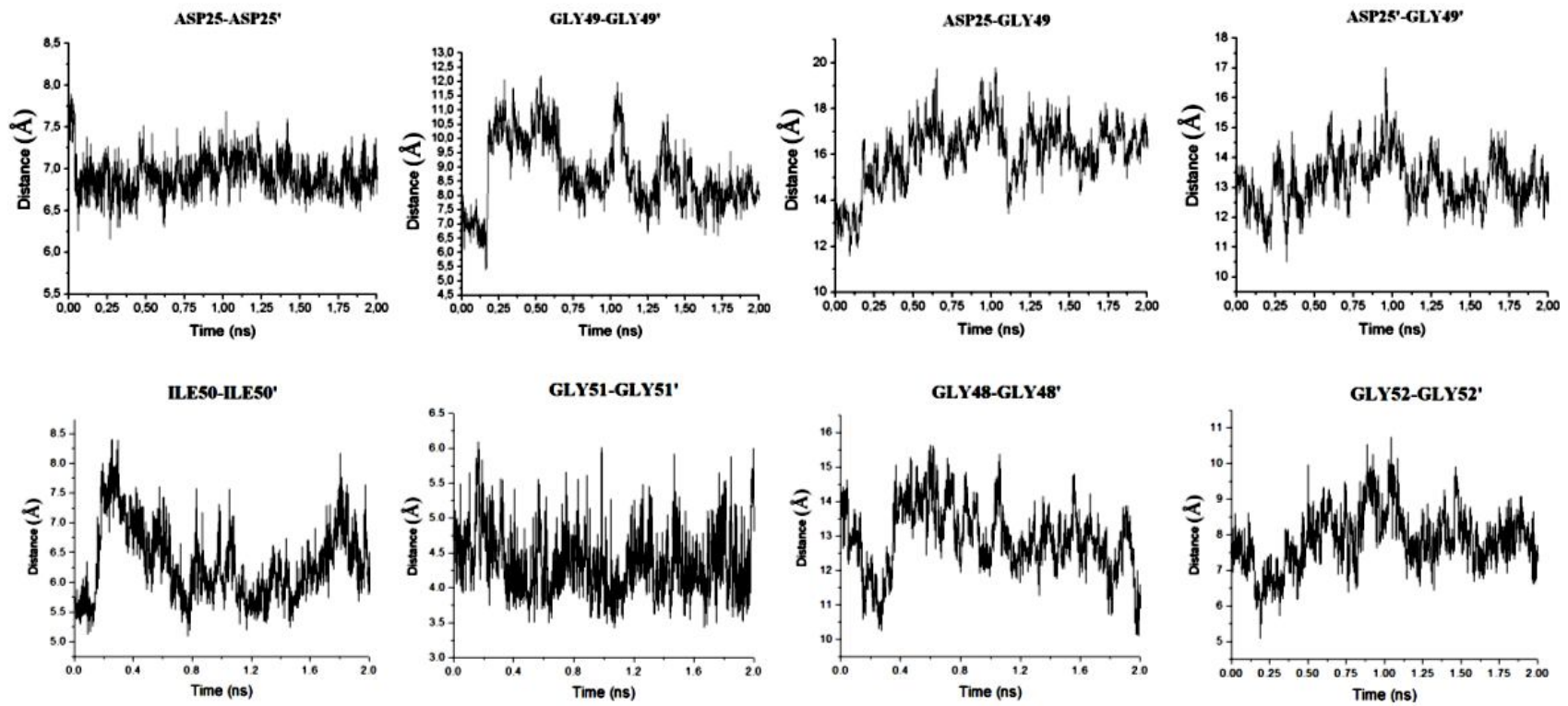
(ii)

Figure 6.7 (i) Change of the HIV-1 PR binding cavity for the ligand-free and; **(ii)** fullerene-bound systems were evaluated in terms of distance between the catalytic residues, Asp25 and Asp25' as well as distance between the residues of the flap region (Gly48, Gly49, Ile50, Gly51 and Gly52) were screened throughout the MD simulations.

160

(i)
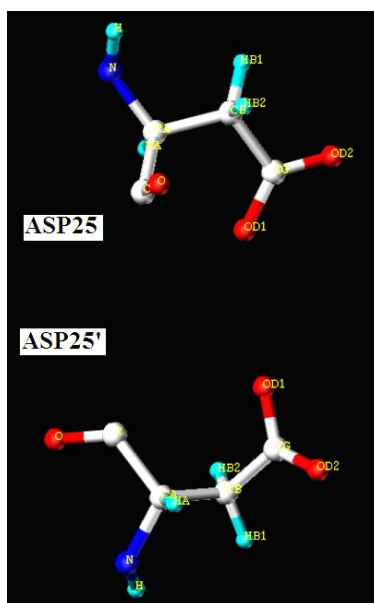
(ii)
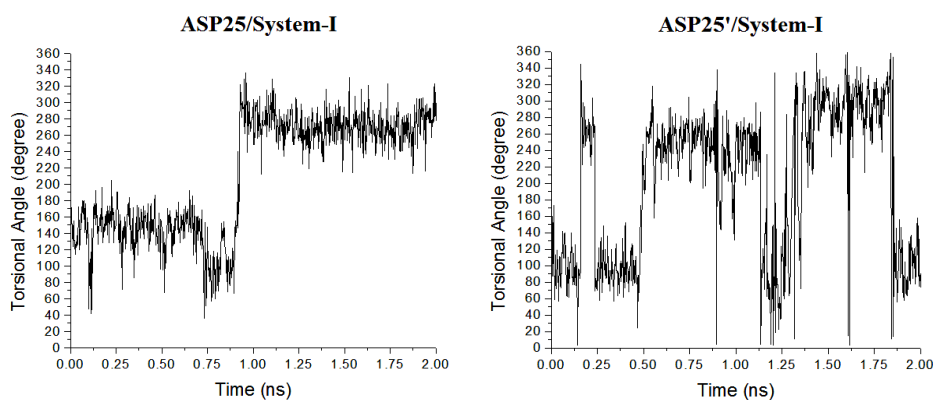
(iii)
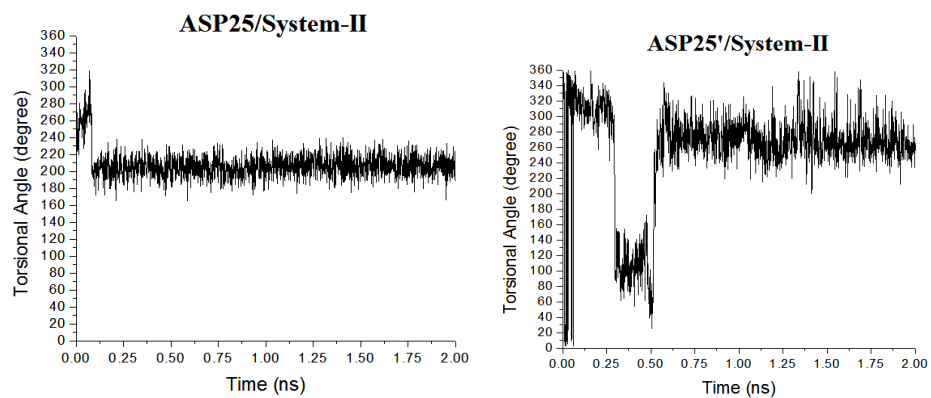
Figure 6.8 (i) The torsional angle of (OD2-C^γ-C^β-C^α) of the catalytic dyad for the **(ii)** ligand-free and the **(iii)** ligand-bound systems throughout the MD simulation.
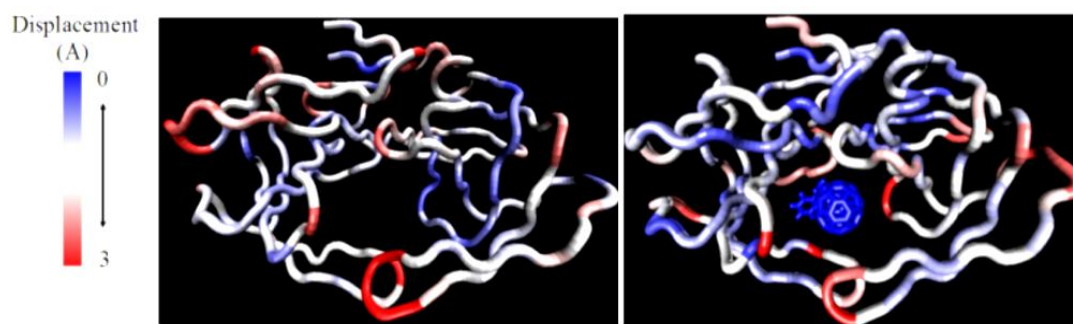
Figure 6.9 Atom positions between the crystal (1AID) and the average structure from MD simulations have been compared for both systems ligand-free and inhibitor-bound systems in order to understand which parts of the enzyme are more stable and which are floppy. Blue colored fields show no change in displacement, red colored fields show high flexible regions.
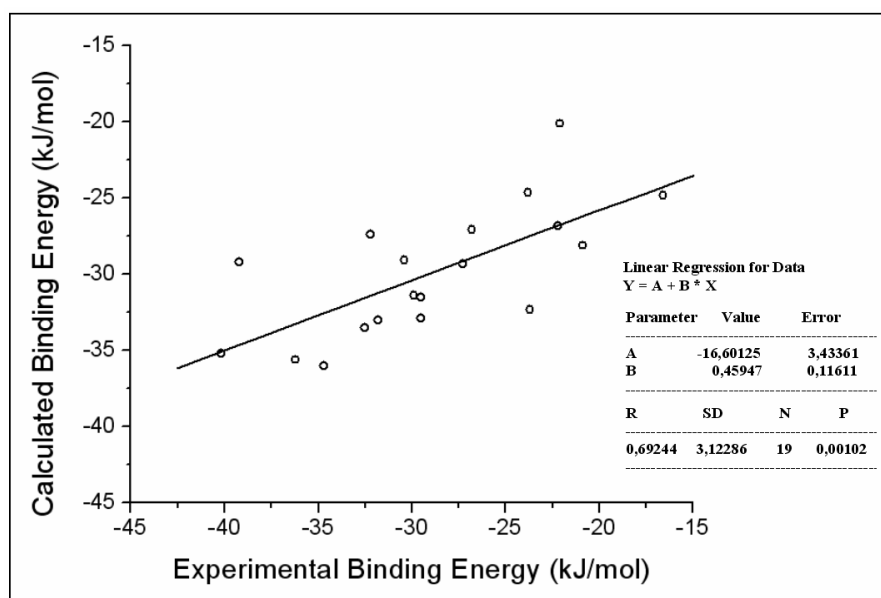
Figure 6.10 Plot of experimental and computed binding energies for the reported fullerene analogues.

significance of the proposed model is verified by the good predictions of the activity of compounds belonging to the test set (Table 6.7). Their pEC_{50} values of these compounds ranges between 4.04 and 6.74 and their biological activities were predicted

from the PLS equations derived from CoMSIA-4 model. All compounds showed predicted values within one logarithmic unit difference from the estimated pEC_{50} values.
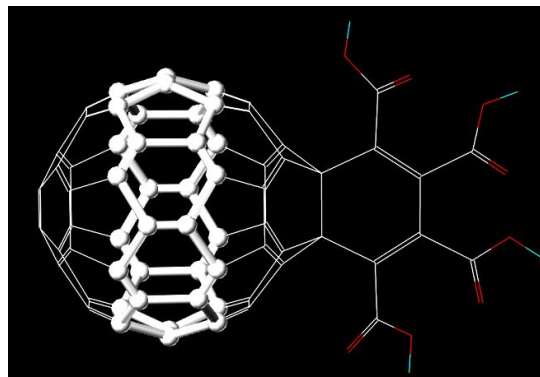
Figure 6.11 Selected atoms of the template compound **23** for structural superimpositions of the compounds in training set.
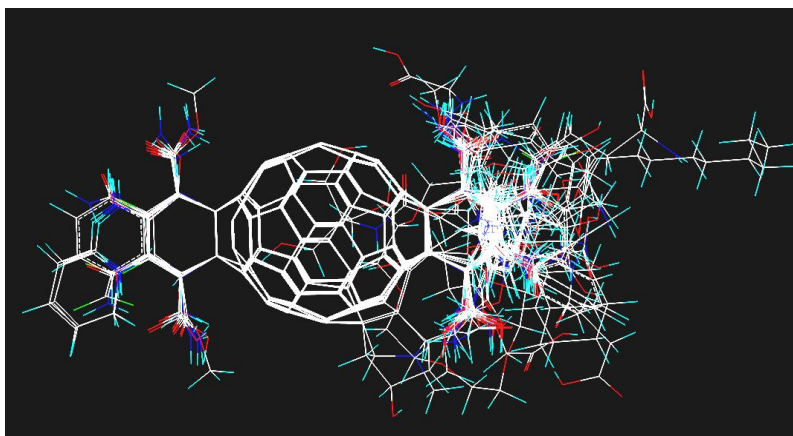
**Figure 6.12** Structural alignments of the compounds in the training set for constructing 3D-QSAR/CoMSIA model at HIV-1 PR.

The contour maps were used to create a matrix in the place of the active site and variations of the used ligands can be generated as long as they fit better into the binding site. Figure 6.14i shows the steric-electrostatic contour maps of the CoMSIA models for the compounds **23** that shows the highest and **36** that shows the lowest inhibition effects within the data set for the HIV-1 PR enzyme. In addition, H-bond donor and H-bond acceptor contour maps are shown in Figure 6.14ii. The individual contributions from the H-bond donor and H-bond acceptor favored and disfavored levels are fixed at 80% and 20%, respectively. The contours for H-bond donor favored fields have been shown in cyan color while its disfavored fields have been

	CoMSIA1 (STR/ES)	CoMSIA2 (STR/ES/ACC)	CoMSIA3 (STR/ES/DON)	CoMSIA4 (STR/ES/ACC/DON)	CoMSIA5 (STR/ES/ACC/ DON/HYD)
r_{cv}^2	0.616	0.733	0.630	0.739	0.670
r^2	0.970	0.991	0.985	0.993	0.993
Components	6	6	6	6	6
F	191.713	632.364	392.176	824.144	861.108
Std. Err.	0.266	0.148	0.188	0.130	0.127
Rel. Contr.					
Steric	0.707	0.512	0.551	0.426	0.243
Electrostatic	0.293	0.143	0.197	0.127	0.086
Hydrophobic	-	-	0.252	-	0.369
H-bond donor	-	-	-	0.167	0.109
H-bond acceptor	-	0.345	-	0.280	0.193

Table 6.5 Summary of statistical results of the derived CoMSIA models for the training set. (Abbreviations: STR, Steric; ES, Electrostatic; ACC, H-bond acceptor; DON, H-bond donor; HYD, Hydrophobic).
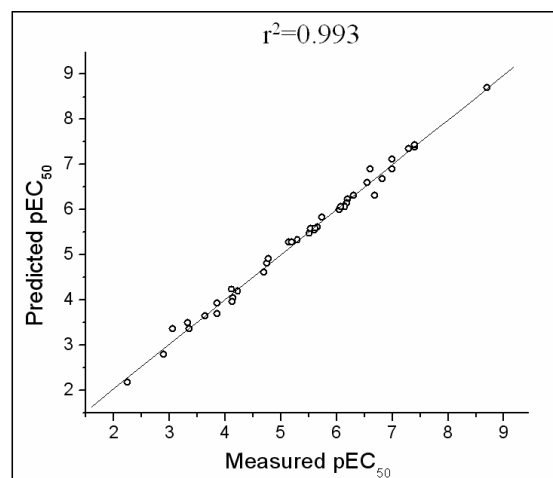
Figure 6.13 Plot of measured and CoMSIA-predicted binding affinities (given as pEC_{50}) of fullerene analogues in the training set at the HIV-1 PR.

shown in purple color. H-bond acceptor favored fields have been shown in orange color while its disfavored fields have been shown in white color. Derived 3D contour maps of CoMSIA models are investigated in the binding cavity of the HIV-1 PR. Contour plots confirmed the stability of the constructed models. For example, the estimated EC_{50} values of **23** and **36** are in nM and in mM ranges, respectively. This can

be explained by different topographical requirements for **23** and **36**, in contour maps (Figure 6.14). There are large yellow colored contours close to the flap regions of HIV-1 PR. Compound **36** shows the existence of sterically unfavorable fields (the areas in which bulky groups are predicted to decrease binding). A part of cyclohexadiene and dihydropyridine groups of **36** fit these unfavorable regions; (right on the Figure 6.14 (top)). However, sterically unfavorable yellow colored contour maps do not match with the subgroups of **23** (left on the Figure 6.14 (top)).

Compound No.	Measured pEC50	3D QSAR CoMSIA Predicted pEC50	Difference
1	7.0	6.90	0.10
2	5.3	5.33	-0.03
3	6.82	6.69	0.13
4	5.14	5.28	-0.14
5	6.31	6.32	-0.01
6	2.89	2.80	0.09
7	4.12	4.24	-0.12
8	3.64	3.65	-0.01
9	3.85	3.70	0.15
10	5.60	5.55	0.05
11	6.05	5.99	0.06
12	5.14	5.28	-0.14
13	5.66	5.62	0.04
14	5.20	5.28	-0.08
15	5.54	5.58	-0.04
17	3.86	3.93	-0.07
18	4.75	4.81	-0.06
19	4.14	4.05	0.09
21	6.60	6.90	-0.30
22	7.29	7.34	-0.05
23	8.70	8.70	0.00
24	6.15	6.07	0.08
25	3.33	3.49	-0.16
26	6.20	6.24	-0.04
28	6.18	6.15	0.03

29	6.68	6.31	0.37
30	4.70	4.62	0.12
31	4.77	4.92	-0.15
32	5.50	5.47	0.03
33	6.55	6.59	-0.04
35	4.23	4.19	0.04
36	2.25	2.18	0.07
37	3.36	3.36	0.00
38	5.73	5.83	-0.10
40	3.06	3.36	-0.30
41	7.40	7.38	0.02
42	7.40	7.43	-0.03
44	5.61	5.59	0.02
45	4.13	3.96	0.17
46	7.00	7.12	-0.12
47	6.15	6.07	0.08
49	6.08	6.07	0.01

Table 6.6 Measured and 3D QSAR/CoMSIA predicted binding affinities of compounds in training set.

Compound	Measured pEC50	3D QSAR CoMSIA Predicted pEC50	Difference
16	4.66	5.39	-0.73
27	5.82	5.31	0.51
34	4.04	3.67	0.37
39	4.61	5.60	-0.99
43	6.74	7.68	-0.94
48	5.22	5.47	-0.25

Table 6.7 Measured and 3D QSAR/CoMSIA predicted binding affinities of compounds in test set.

Furthermore, the subgroups of **23** fit perfectly with the sterically favorable areas which are shown with green colored contours (left on the Figure 6.14 (top)). Electrostatic contour maps (shown with blue and red colored contour maps) are mainly ob-

served in the catalytic region of binding cavity. Red colored contours which show electronegative favored fields are in very close neighborhood with the –COOH groups of **23**. H-bond donor and H-bond acceptor contour maps have been shown in Figure 6.14 (bottom) in the binding cavity of HIV-1 PR. H-bond acceptor and H-bond donor favored regions fit very well with the –COOH groups of **23** and disfavored regions are far from the subgroups of the **23** (left on the Figure 6.14 (bottom)). However, subgroups of **36** fit mainly with the purple colored contours which are disfavored for the H-bond donor interactions (right on the Figure 6.14 (bottom)).

Leapfrog *de novo* drug design software has been used to propose predicted novel fullerene HIV-1 PR inhibitors. Molecules **1, 3** (biologically measured potent molecules, Table 6.1); **23** and **42** (computationally designed monoadduct and bisadduct fullerene derivatives that have best predicted binding energy with HIV-1 PR, Table 6.4) were used as starting structures with allowing the modifications only for the subgroups of fullerene derivatives in Leapfrog simulations. In addition to the *de novo* drug design, 3D QSAR/CoMSIA contour maps were also used to design new monoadducts and bisadducts [60]fullerene. More than 100 compounds have been designed to vary in polarity and contain various groups exerting electrostatic and steric interactions at different topographical requirements and their binding energies with HIV-1 PR have been computed using molecular docking studies (representative structures and their computed binding energies were presented in Table 6.8). These are expected to differ in their mode of action as it is indeed observed with the active site of the receptor. The binding interactions of **1** (high potent biologically evaluated fullerene derivative, Table 6.1) and **23** (computationally designed fullerene derivative with predicted very high potency, Table 6.4) at the active site of the HIV-1 PR have been compared in Figure 6.15. Compound **1** forms two hydrogen bonds between hydrogen atom of –OH group of ligand and oxygen atoms of the –COO group of Asp25 catalytic amino acid residue at chain B, together with oxygen atom of –OH group of ligand and hydrogen atom of backbone –NH group of Ala28 at chain B (left on the Figure 6.15). On the other hand, **23** forms a hydrogen bond network between –COOH groups of ligand and catalytic amino acid residues Asp25, Gly27, Asp29, Asp30 at chain A together with Asp25 and Gly27 at chain B (right on the Figure 6.15). The van der Waals interactions of these ligands with non-polar HIV-1 PR surface have been observed mainly at the flap part of the cavity.

It must be noted that this is a computational study, thus the easy feasibility of synthesizing any of the designed derivatives has not been thoroughly examined.
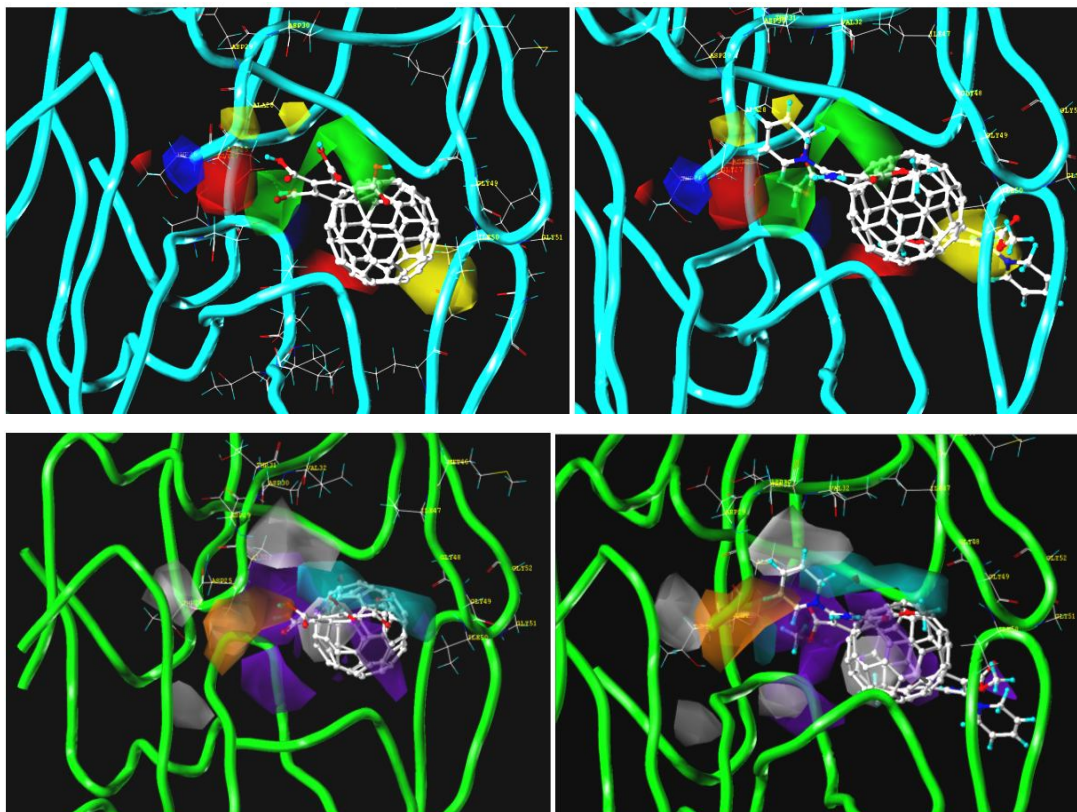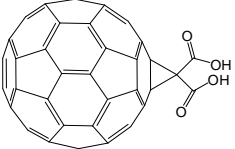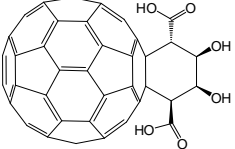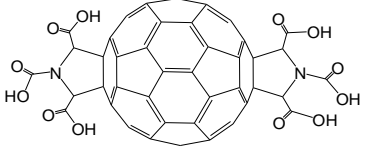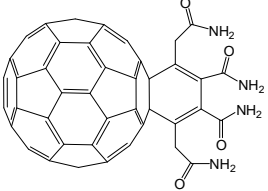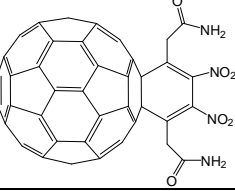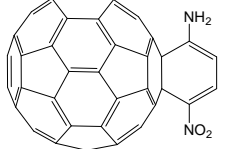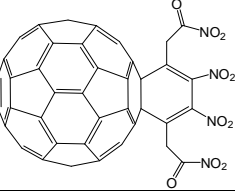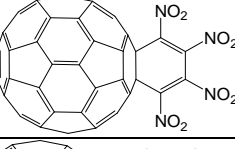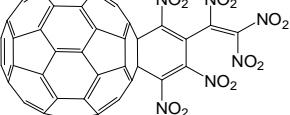

Figure 6.14 (top) CoMSIA steric/electrostatic contour maps of template compound **23** (template compound; has best binding affinity in training set, left on the figure) and compound **36** (has worst binding affinity in training set, right on the figure). **(bottom)** CoMSIA H-bond donor/H-bond acceptor contour maps of compounds **23** and **36** (on the left and right of the figure, correspondingly). The individual contributions from the H-bond donor and H-bond acceptor favored and disfavored levels are fixed at 80% and 20%, respectively. The contours for H-bond donor favored fields have been shown in cyan color while its disfavored fields have been shown in purple color. H-bond acceptor favored fields have been shown in orange color while its disfavored fields have been shown in white color.

No.	Compound	Calculated Binding Energy (kJ/mol)
D1		-38.38
D2		-35.56
D3		-36.57
D4		-41.59
D5		-43.74
D6		-41.09
D7		-41.93
D8		-43.03
D9		-44.88

D10		-43.09
D11		-46.25
D12		-53.20

Table 6.8 Computationally designed fullerene derivatives using 3D QSAR contour maps and *de novo* drug design and their binding energies (the top binding score from FlexX molecular docking at HIV-1 PR).

Figure 6.15 The binding interactions of **1** and **23** at the active site of the HIV-1.

Chapter 7. Summary and Conclusions

In this study, bioactive conformations of CBs and [60]fullerene derivatives have been investigated. The major structural characteristics of these molecules are their amphiphilic properties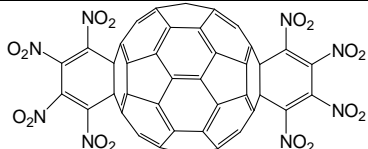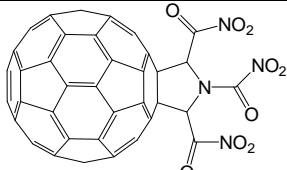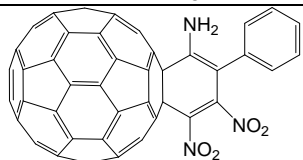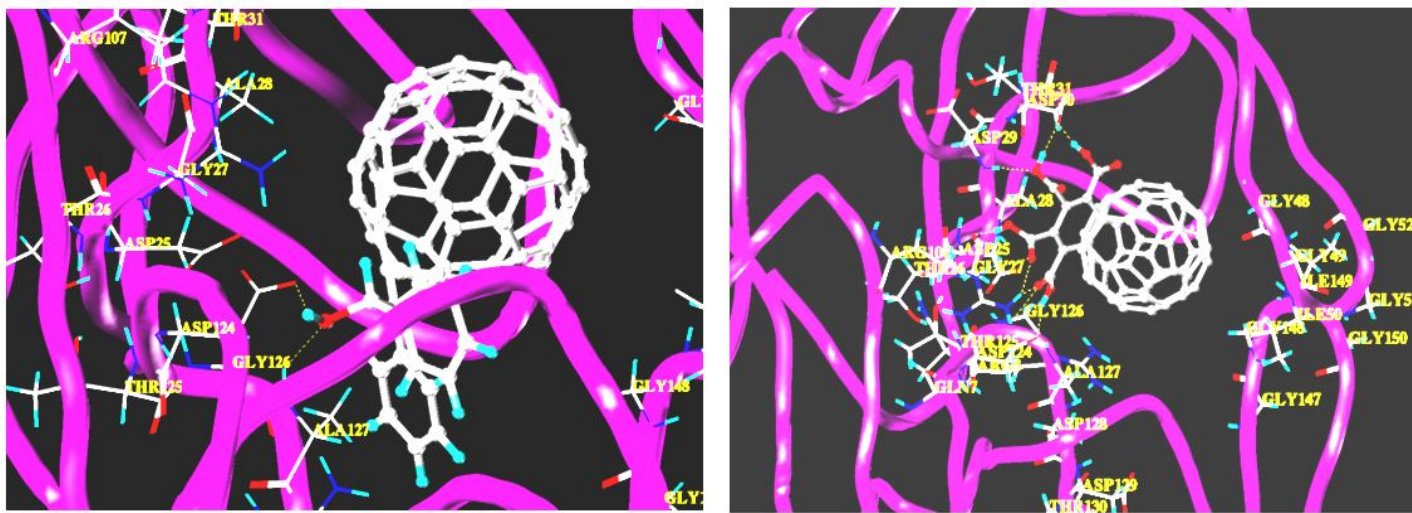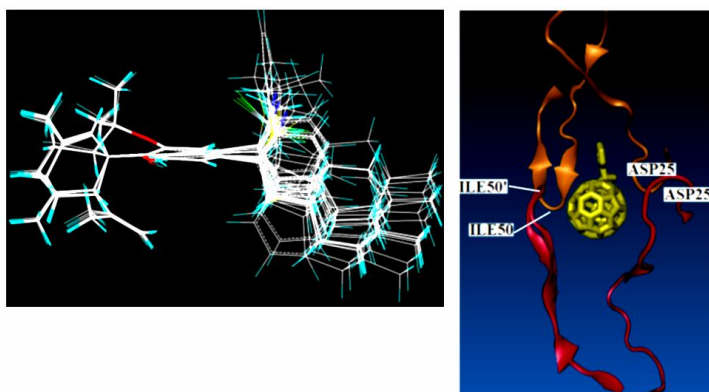 as well as the existence of flexible and rigid pharmacophoric segments. Their flexible segment constitutes a challenging field for conformational analysis exploring of putative bioactive conformations. In the first part, 3D QSAR, molecular docking, MC calculations, MD simulations and *de novo* drug design of CBs have been analyzed in order to propose novel and selective high affinity CB analogues for CB1 and CB2 receptors. Since conformational analysis of high active compound AMG3 was performed in different environments and derived conformations were used as template in the generation of 3D QSAR models, the connectivity between the used environments, thus complexity level of calculations, and obtained statistical results were discussed. In the second part, methodologies applied to CBs were performed to the monoadducts and bisadducts of [60]fullerene analogues in order to design higher inhibitory effect [60]fullerenes against HIV-1 PR enzyme as well as to analyze the connection between biological activity of fullerene derivatives and conformational changes of catalytic and flap regions of the enzyme.

7.1 CB Derivatives

Cannabis sativa (*marijuana*) has long been used for its psychotropic and pharmacological effects. In the 1960's, the correct stereochemistry of both Δ^9-THC and CBD were elucidated, and Δ^9-THC was identified as the principle psychoactive component of cannabis sativa. A wide range of pharmacological effects was described for Δ^9-THC (e.g., analgesic, neuroprotective, anti-inflammatory, antiemitic, bronchodilatory, anticonvulsant). The structure elucidation of Δ^9-THC allowed also the design of synthetic analogues (e.g., Δ^8-THC) starting from 1970's. Δ^8-THC has a very similar pharmacological profile as Δ^9-THC, however it is chemically more stable. Given psychotropic effects of CBs, many biological investigations employed brain and brain plasma membranes as study objects. Consensus data describing several key characteristics of CB action emerged that CB analogues elicit biological effects in a stereo and structurally selective manner. Their binding to brain plasma membrane is saturable, stereospecific, concordant with *in vitro* and *in vivo* bioresponses (e.g., analgesia) and nonrandom in select brain regions. Thus, these characteristics strongly implied that

CB pharmacology is receptor-mediated and search led to discovery of cloning of two GPCRs for CBs namely CB1 and CB2 receptors with a high sequence homology. However, hitherto no direct observation of a CB ligand bound to a CB receptor using X-ray crystallography has been reported. Thus, active sites of these receptors have been postulated from various *in silico* methods such as molecular docking using homology model receptor systems and 3D QSAR studies. Alignment process is a critical part in 3D QSAR studies and affects statistical results. In addition, if derived 3D QSAR PLS analysis will be used in the steps of *de novo* drug design, it may affect also the proposed novel molecules. Thus, in the present study, the conformational analysis of template compound AMG3 which used in 3D QSAR studies have been analyzed at with and without receptor environments.

MD simulations of AMG3 in lipid bilayer trajectory analysis showed that the dihedral angle defined between aromatic and dithiolane ring of the alkyl side chain of AMG3 shows more resistance to be transformed to another torsional angle because the heterocyclic part interacts with the ring A of the rigid segment and leads to restriction of the rotation. The τ_3-τ_6 dihedral angles are transformed to optimal dihedral angle value of *all trans* conformation. Thus, the wrapped conformations are dynamically less probable in lipid bilayer than the linear conformations. The second dihedral angle τ_2 at the alkyl chain shows very flexible character and it is the critical dihedral angle in lipid bilayer for producing different low energy conformations. Both of MM and QM geometry optimization calculations showed that, favored conformations of AMG3 derived from MD in lipid bilayer form a perpendicular plane angle between tricyclic ring and flexible segments. The obtained results suggest that synthetic analogues incorporating restraints (multiple bonds or rings) at different positions of alkyl chain may be of importance to be synthesized. Especially, such synthetic analogues are important if they restrict the favorable angles and avoid the perturbation effects around the favorable dihedral angle.

MD simulations have been also performed to the systems including AMG3 at the binding site of the CB1 and CB2 receptors merged with membrane bilayer in order to analyze the effect of amino acid residues at the binding cavities of CB receptors to the conformational properties of ligand. Although the CB1 and CB2 receptors exhibit a high sequence homology, there are certain behavior differences of AMG3 conformers

at the binding sites of receptors. One of the main differences between the MD simulations of ligand at the CB1 and CB2 receptors is the different behavior of the first dihedral angle τ_1 of the alkyl side chain of AMG3. In the CB1 receptor, there is a high propensity of τ_1 to establish a *gauche+* conformation, however in the CB2 receptor; it prefers to have a *trans* conformation. It is well-known that, different conformational rearrangements of third and sixth TMs of GPCR determine the activation of CBs. In CB2 receptor, alkyl side chain of AMG3 conformers align parallelly in the ligand recognition part of TM3, while in the CB1 receptor they align perpendicularly with the ligand recognition part of TM3. This observation may help to understand the selectivity of CB ligands for the CB1 and CB2 receptors.

Three 3D QSAR models were generated using favored derived template conformers from (i) free ligand, (ii) molecular docking and (iii) MD simulations of ligand at binding site of the receptor merged with lipid bilayer . Although there are some fine differences between the derived models, three general conclusions could be drawn from the characteristic contour maps: (i) The relative contributions of steric fields are larger than electrostatic fields; (ii) the orientation of the C3-alkyl side chain plays an important role in determining the biological activity; (iii) because of the structural differences of Δ^8-THC and CBD derivatives at the cyclic ring segment, these groups have different pharmacophoric requirements for their receptors in these regions. While sterically unfavorable areas are located on the methyl or propenyl groups of CBD analogues, these unfavorable regions are located at the vicinity of the tricyclic segment of Δ^8-THC analogues. Therefore, Δ^8-THC analogues have higher binding affinities than their corresponding CBD analogues. Comparison of generated QSAR models were shown that when the complexity level of calculations increased (mimicking more accurately the real environment), it positively affects the obtained statistical results. The optimal QSAR PLS analysis which is obtained from the third generation of 3D QSAR model was used in the *de novo* drug design studies and these simulations provided novel compounds with enhanced predicted binding affinities.

If the X-ray structure of a ligand-bound receptor is not available, homology models of the protein can be used to obtain the ligand binding cavities. The steroelectronic properties of these cavities are directly related to the performed molecular model coordinates. In the present study, a homology modeling study based on the β2-adrenergic

receptor for both CB1 and CB2 receptors was also performed and results confirmed the obtained binding pockets found in receptor models based on rhodopsin.

7.2 Fullerene Derivatives

Since the discovery of fullerene in 1985, derivative of fullerenes have been extensively investigated for biomedical applications (e.g., antibacterial, neuroprotective, antiviral and apoptosis). Antiviral activity of fullerenes has been supported by molecular modeling and *in vitro* studies, which provided that the fullerene can be accommodated inside the catalytic cavity present in the HIV-1 PR.

The binding affinity values of "first generation" fullerene inhibitors against HIV-1 PR were not significant (EC_{50} ~10^{-6} M). Thus, further structural investigation is required in order to propose new HIV-1 PR/fullerene complexes with better binding affinity. Novel monoadducts and bisadducts of [60]fullerene have been designed with a combination of 3D QSAR models and molecular docking studies. In order to use proper input coordinates of HIV-1 PR/fullerene derivative complex in the docking, MD simulations were employed for the ligand-free and the inhibitor-bound HIV-1 PR systems. MD simulations contributed substantially in the understanding of the structural changes at the catalytic and flap regions for these two different systems. MD simulations results showed structural differences between the unbound and bound systems of the binding cavity of HIV-1 PR. In the fullerene inhibitor-bound forms, the flaps are pulled in toward the bottom of active site (the closed form), while, the structure for the free HIV-1 PR adopts a semi-open conformation with flaps shifted away from the dual Asp25 catalytic site. MD simulations have also shown that flap, catalytic and termino-lateral regions of HIV-1 PR show more flexibility during the simulations.

Obtained 3D QSAR models gave high relative contributions of steric fields which confirm the importance of the van der Waals interactions with non-polar HIV-1 PR surface in the activity of fullerenes. The contour maps from constructed 3D QSAR/CoMSIA models together with *de novo* drug design studies assisted to propose novel fullerene derivatives with better predicted potency which can aid synthetic chemists to initiate the synthesis of novel fullerene derivatives as HIV-1 PR inhibitors.

Bibliography

1. Richton, A.B. Mathematech, *1*, 83, **1984**.

2. Leach, A.R. *Molecular Modelling: Principles and Applications*, 2nd edition, Pearson Educ. Ltd., England, **2001**.

3. Barton, D.H.R. *Experientia*, *6*, 316, **1950**.

4. Patrick, G.L. *An Introduction to Medicinal Chemistry*, 2nd edition, Oxford University Press Inc., New York, **2001**.

5. Reggio, P.H. *Curr. Pharm. Des. 9*, 1607, **2003**.

6. Khanolkar, A. D.; Palmer, S. L.; Makriyannis, A. *Chemistry and Physics of Lipids*, *108*, 37, **2000**.

7. Picone, R.P.; Fournier, D.J.; Makriyannis, A. *J. Peptide Res. 60*, 348, **2002.**

8. Shim, J.Y.; Welsh, W.J.; Howlett, A.C. *Biopolymers (Peptide Science), 71*, 169, **2003**.

9. Cramer, R. D. I.; Depriest, S.; Patterson, D.; Hecht, P. *The Developing Practice of Comparative Molecular Field Analysis*; Kubinyi, H., Ed.; ESCOM: Leiden, **1993**, 443.

10. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem. 37*, 4130, **1994**.

11. a) Sijbesma, R.; Srdanov, G.; Wudl, F.; Castoro, J. A.; W-,Q, Charles;. Friedman, S. H.; Decamp, D. L.; Kenyonl, G. L. *J. Am. Chem. Soc. 115*, 6510, **1993**. b) Friedman, H. S.; DeCamp, D. L.; Sijbesma, R. P.; Srdanov, G.; Wudl, F.; Kenyon, G. *J. Am. Chem. Soc. 115*, 6506, **1993**.

12. Benyamini, H.; Shulman-Peleg, A.; Wolfson, H.J.; Belgorodsky, B.; Fadeev, L.; Gozin, M. *Bioconjugate Chem.*, *17*, 378, **2006**.

13. Friedman, H. S.; Ganapathi, P. S.; Rubin, P. S.; Kenyon, G. L. *J. Med. Chem. 41*, 2424, **1998**.

14. Levine, I.N. *Quantum Chemistry*, 5th edition, Prentice Hall, **2000**.

15. Hinchliffe, A. *Computational Quantum Chemistry*, 1st edition, John Wiley & Sons Ltd. **1989**.

16. Cramer, C.J. Essentials of Computational Chemistry: Theories and Models, 1st edition, John Wiley & Sons Ltd. **2002**.

17. Szabo, A.; Ostlund, N.S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, 1st edition, Mc Graw Hill Publishing Co. **1982**.

18. Lipkowitz, K.B., Boyd, D.B. *Reviews in Computational Chemistry*, VCH Publishers, **1990**.

19. Jensen, F. *Introduction to Computational Chemistry*, John Wiley & Sons Ltd. **1999**.

20. Born, M. *Zeitschrift für Physik*, *1*, 45, **1920**.

21. Hinchliff, A. *Molecular Modelling for Beginners*. John Wiley & Sons Ltd. **2003**.

22. Ajay, A.; Murco, M.A. J. Med. Chem. 38, 4951, **1995**.

23. Bohm, H.J. J. *Comput. Aid. Mol. Des.* 8, 243, **1994**.

24. Thomas, G. *Medicinal Chemistry: An Introduction*, John Wiley & Sons Ltd. **2000**.

25. Cramer, R.D.; Patterson, D.E.; Bunce, J.D. *J. Am. Chem. Soc. 110*, 5959, **1988**.

26. Durdagi, S.; Kapou, A.; Kourouli, T.; Andreou, T.; Nikas, S. P.; Nahmias, V. R.; Papahatjis, D. P.; Papadopoulos, M. G.; Mavromoustakos, T. *J. Med. Chem. 50*, 2875, **2007**.

27. *Sybyl Molecular Mmodeling Software Package*, ver. 6.8, Tripos Inc., St Louis, **2001**.

28. Durdagi, S.; Mavromoustakos, T.; Kronakis, N.; Papadopoulos, M. G. *Bioorg. Med. Chem. 16*, 9957, **2008**.

29. Durdagi, S.; Mavromoustakos, T.; Papadopoulos, M. G. *Bioorg. Med. Chem. Lett. 18*, 6283, **2008**.

30. Durdagi, S.; Reis, H.; Papadopoulos, M. G.; Mavromoustakos, T. *Bioorg. Med. Chem. 16,* 7377, **2008**.

31. Durdagi, S.; Papadopoulos, M. G.; Papahatjis, D. P.; Mavromoustakos, T. *Bioorg. Med. Chem. Lett. 17*, 6754, **2007**.

32. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* 4, 187, **1983**.

33. Tuccinardi,T.; Ferrarini, P.L.; Manera, C.; Ortore, G.; Saccomanni, G.; Martinelli, A. *J. Med. Chem. 49*, 984, **2006**.

34. Salo, O. M. H.; Lahtela-Kakkonen, M.; Gynther, J.; Jarvinen, T.; Poso, A. *J. Med. Chem.* 47, 3048, **2004**.

35. Rutenber, E.; Fauman, E. B.; Keenan, R. J.; Fong, S.; Furth, P. S.; Ortiz de Montellano, P. R.; Meng, E.; Kuntz, I. D.; DeCamp, D. L.; Salto, R.; Ros, J. R.; Craik, C. S.; Stroud, R.M. *J. Biol. Chem. 268*, 15343, **1993**.

36. a) van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Ber-endsen, H. J. *J. Comput. Chem. 26*, 1701, **2005**. b) Lindhal, E.; Hess, B.; van der Spoel, D. *J. Mol. Model. 7*, 306, **2001**.

37. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J.R. *J. Chem. Phys. 81*, 3684, **1984**.

38. Gromacs user manual, version 3.2, **2004**.

39. Hess, B.; Bekker, H.; Berendsen, H.J.C.; Fraaije, J.G.E.M. *J. Comput. Chem.*, *18*, 1463, **1997**.

40. Humphrey, W., Dalke, A.; Schulten, K., *J. Molec. Graphics*, *14*, 33, **1996**.

41. Schuettelkopf, A. W.; van Aalten, D. M. F. *Acta Crystal. D60*, 1355, **2004**.

42. Patra, M.; Karttunen, M.; Hyvonen, M.; Falck, E.; Lindqvist, P.; Vattulainen, I. *Biophysical Journal, 84,* 3636, **2003**.

43. Patra, M.; Karttunen, M.; Hyvonen, M.; Falck, E.; Lindqvist, P. *J. Am. Chem. Soc. 108*, 4485, **2004**.

44. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys. 103*, 8577, **1995**.

45. http://www.microcal.com.

46. Sybyl LeapFrog Manual, Tripos Inc., **1999**.

47. Nair, C. P.; Sobiha, M. E. *J. Mol. Graph. Model. 26*, 117, **2007**.

48. Makhija, M. T.; Kasliwal, R. T.; Kulkarni, V. M.; Neamati, N. *Bioorg. Med. Chem. 12*, 2317, **2004**.

49. Kapou, A.; Benetis, N. P.; Durdagi, S.; Nikolaropoulos, S.; Mavromoustakos, T. *J. Chem. Info. Model. 48*, 2254, **2008**.

50. Tanrikulu, Y.; Schneider, G. *Nat. Rev. Drug. Dis.*, *7*, 667, **2008**.

51. Reddy, M.R.; Erion, M.D. *Curr. Pharm. Des., 11,* 283, **2005**.

52. Taft, C.A.; Da Silva, B.V.; Da Silva, C.H.T.P. *J. Pharm. Sci., 97*, 1089, **2008**.

53. a) Schneider, G.; Fechner, U. *Nat. Rev. Drug. Dis.*, *4*, 649, **2005**. b) Klebe, G. *Drug Discov. Today 11,* 580, **2006**. c) Lloyd, D. G.; Buenemann, C. L.; To-dorov, N. P.; Manallack, D. T.; Dean, P. M. *J. Med. Chem. 47*, 493, **2004**. d) Kitano, H. *Nature 420*, 206, **2002**. e) Caitlin, S. *Nature 422*, 341, **2003**. f) Bleicher, K. H.; Bohm, H. J.; Muller, K.; Alanine, A. I. *Nature Rev. Drug Dis-cov. 2,* 369, **2003**. g) Congreve, M.; Murray, C.W.; Blundell, T.I. *Drug Dis-cov. Today 10,* 894, **2005**.

54. Gaoni, Y.; Mechoulam, R. *J. Am. Chem. Soc. 86*, 1646, **1964**.

55. Devane, W. A.; Dysarz, F. A.; Johnson, M. R.; Melvin, L. S.; Howlett, A. C. *Mol. Pharmacol. 34*, 605, **1988**.

56. Gerard, C. M.; Mollereau, C.; Parmentier, M. *Nucleic Acid Research 18*, 7142, **1990**.

57. Matsuda, L. A.; Lolait, S. J.; Brownstein, M. J.; Young, A. C.; Bonner, T. I. *Nature 346*, 561, **1990**.

58. Gerard, C. M.; Mollereau, C.; Vassart, G.; Parmentier, M. *Biochemical Journal, 279,* 129, **1991**.

59. Munro, S.; Thomas, K. L.; Abu-Shar, M. *Nature 365*, 61, **1993**.

60. Padgett, L.W. *Life Sciences 77*, 1767, **2005**.

61. Sugiura, T.; Kishimoto, S.; Oka, S.; Gokoh, M. *Progress in Lipid Research 45*, 405, **2006**.

62. Herkenham, M.; Lynn, A. B.; Little, M. D.; Johnson, M. R.; Melvin, L. S.; De Costa, B. R.; Rice, K. C. *Proc. Natl. Acad. Sci.*, *87*, 1932, **1990**.

63. Glass, M.; Dragunow, M.; Faull, R. L. *Neuroscience*, *77*, 299, **1997**.

64. Straiker, A. J.; Maguire, G.; Makie, K.; Lindsey, J. *Invest Ophthalmol. Visual. Sci. 40*, 2442, **1999**.

65. Pertwee, R. G.; Fernando, S. R. *Br. J. Pharmacol. 118*, 2053, **1996**.

66. Klein, T. W.; Newton, C.; Friedman, H. *Immunol. Today 19*, 373, **1998**.

67. Thakur, G. A.; Duclos Jr., R. I.; Makriyannis, A. *Life Sciences 78*, 454, **2005**.

68. Fichera, M.; Cruciani, G.; Bianchi, A.; Musumarra, G. *J. Med. Chem.*, *43*, 2300, **2000**.

69. Galie`gue, S.; Mary, S.; Marchand, J.; Dussossoy, D.; Carriere, D.; Carayon, P.; Bouaboula, M.; Shire D.; Le Fur, G.; and Casellas, P. *Eur. J. Biochem. 232*, 54, **1995**.

70. Oz, M. *Pharmacology & Therapeutics*, *111*, 114, **2006**.

71. Parkkari, T.; Savinainen, J. R.; Raitio, K. H.; Saario, S. M.; Matilainen, L.; Sirvio, T.; Laitinen, J. T.; Nevalainen, T.; Niemi, R.; Jarvinen, T. *Bioorg. Med. Chem. 14*, 5252, **2006**.

72. Salo, O. M. H.; Savinainen, J. R.; Parkkari, T.; Nevalainen, T.; Lahtela-Kakkonen, M; Gynther, M; Laitinen, J. T.; Jarvinen, T.; Poso, A. *J. Med. Chem. 49*, 554, **2006**.

73. a) Lambert, D. M.; Fowler, C. J. *J. Med. Chem. 48*, 1, **2005**. b) Felder, C. C. *Drug Dis. Today: Ther. Str. 3*, 561, **2006**.

74. Anand, U.; Otto,W.; Sanchez-Herrera, D.; Facer, P.; Yiangou, Y.; Korchev, Y.; Birch,R.; Benham, C.; Bountra,C.; Chessell, I. *Pain*, *138*, 667, **2008**.

75. Devane, W. A.; Hanus, L.; Breuer, A.; Pertwee, R. G.; Stevenson, L. A.; Griffin, G.;Gibson, D.; Mandelbaum, A.; Etinger, A.; Mechoulam, R. *Science 258*, 1946, **1992**.

76. Lin, S.; Khanolkar, A. D.; Fan, P.; Goutopoulos, A.; Qin, C.; Papahadjis, D.; Makriyannis, A. *J. Med. Chem.* 41, 5353, **1998**.

77. Sugiura, T.; Kondo, S.; Sukugawa, A.; Nakane, S.; Shinoda, A.; Itoh, K.; Yamashita, A.; Waku, K. *Biochem. Biophys. Res. Commun.*, *215*, 89, **1995**.

78. Mechoulam, R.; Ben-Shabat, S.; Hanus, L.; Ligumsky, M.; Kaminski, N. E.; Schatz, A.   R.; Gopher, A.; Almog, S.; Martin, B. R.; Compton, D. R.; Pertwee, R. G.; Griffin, G.; Bayewitch, M.; Barg, J.; Vogel, Z. *Biochem. Pharmacol. 50*, 83, **1995**.

79. Stella, N.; Schweitzer, P.; Piomelli, D. *Nature 388*, 773, **1997**.

80. Mackie, K. *Annu. Rev. Pharmacol. Toxicol. 46*, 101, **2006**.

81. Felder, C. C.; Glass, M. *Annu. Rev. Pharmacol. Toxicol. 38*, 179, **1998**.

82. Palmer, S. L.; Thakur, G. A.; Makriyannis, A. *Chem. Phys. Lip. 121*, 3, **2002**.

83. Razdan, R. K. *Pharmacol. Rev. 38*, 75, **1986**.

84. Makriyannis, A.; Rapaka, R. S. *Life   Sciences 47*, 2173, **1990**.

85. Papahatjis, D. P.; Kourouli, T.; Abadji, V.; Goutopoulos, A.; Makriyannis, A. *J. Med. Chem.*, *41*, 1195, **1998**.

86. Xie, X-Q.; Pavlopoulos, S.; DiMeglio, C. M.; Makriyannis, A. *J. Med. Chem. 41*, 167, **1998**.

87. Melvin, L. S.; Johnson, M. R. *NIDA Res. Monogr.*, *79*, 31, **1987**.

88. Howlett, A. C.; Johnson, M. R.; Melvin, L. S.; Milne, G. M. *Mol. Pharmacol.*, *33*, 297, **1988**.

89. Mavromoustakos, T.; Theodoropoulou, E.; Zervou, M.; Kourouli, T.; Papahatjis, D. *J. Pharm. Biomed.* Anal. *18*, 947, **1999**.

90. Keimowitz, A. R.; Martin, B. R.; Razdan, R. K.; Crocker, P. J.; Mascarella, S. W.; Thomas, B. F. *J. Med. Chem.*, *43*, 59, **2000**.

91. Shire, D.; Calandra, B.; Delpech, M.; Dumont, X.; Kaghad, M.; Le Fur, G.; Caput, D.; Ferrara, P. *J. Biol. Chem.*, *271,* 6941, **1996**.

92. Reggio, P. H.; Panu, A. M.; Miles, S. *J. Med. Chem.*, *36*, 1761, **1993**.

93. Semus, S. F.; Martin, B. R. *Life Sci. 46,* 1781, **1990**.

94. Shim, J. Y.; Collantes, E. R.; Welsh, W. J.; Subramaniam, B.; Howlett, A. C.; Eissenstat, M. A.; Ward, S. J. *J. Med. Chem.*, *41,* 4521, **1998**.

95. Thomas, B. F.; Compton, D. R.; Martin, B. R.; Semus, S. F. *Mol. Pharmacol.*, *40,* 656, **1991**.

96. Thomas, B. F.; Adams, I. B.; Mascarella, S. W.; Martin, B. R.; Razdan, R. K. *J. Med. Chem. 39*, 471, **1996**.

97. Chen, J-Z.; Han, X-W.; Liu, Q.; Makriyannis, A.; Wang, J.; Xie, X-Q. *J. Med. Chem.*, *49*, 625, **2006**.

98. Doytchinova, I. A.; Flower, D. R. *J. Med. Chem.*, *44,* 3572, **2001**.

99. Folkers, G.; Merz, A.; Rognan D. *CoMFA: Scope and Limitations. In 3D QSAR in Drug Design*; Kubinyi, H., Ed.; ESCOM: Leiden, 583, **1993**.

100. Tong, W.; Collantes, E. R.; Welsh, W. J.; Berglund, B. A.; Howlett, A. C. *J. Med. Chem.*, 41, 4207, **1998**.

101. Kapou, A.; Benetis, N. P; Avlonitis, N.; Calogeropoulou, T., Koufaki, M.; Scoulica, E., Nikolaropoulos, S. S.; Mavromoustakos, T. *Bioorg. Med. Chem.*, *15*, 1252, **2007**.

102. Schmetzer, S.; Greenidge, P.; Kovar, K. A.; Schulze-Alexandru, M.; Folkers, G. *J. Comput. Aided Mol. Des.*, *11,* 278, **1997**.

103. Xie, A.; Sivaprakasam, P.; Doerksen, R. J. *Bioorg. Med. Chem. 14*, 7311, **2006**.

104. Papahatjis, D. P.; Nikas, S. P.; Kourouli, T.; Chari, R.; Xu, W.; Pertwee, R. G.; Makriyannis, A. *J. Med. Chem.*, *46*, 3221, **2003**.

105. Papahatjis, D. P.; Nahmias, V. R.; Andreou, T.; Fan, P.; Makriyannis A. *Bioorg. Med. Chem. Lett.*, *16*, 1616, **2006**.

106. Papahatjis, D. P.; Nikas, S. P.; Andreou, T.; Makriyannis, A. *Bioorg. Med. Chem. Lett.*, *12*, 3583, **2002**.

107. Gareau, Y.; Dufrense, C.; Gallant, M.; Rochette, C.; Sawyer, N.; Slipetz, D. M.; Tremblay, N.; Weech, P. K.; Metters, K. M.; Labelle, M. *Bioorg. Med. Chem. Lett.*, *6*, 189, **1996**.

108. Mavromoustakos, T.; Kapou, A.; Benetis, N. P.; Zervou, M. *Drug Des. Rev.- Online*, *1*, 235. **2004**.

109. Zoumpoulakis, P.; Mavromoustakos, T. *Drug Des. Rev.-Online*, *2*, 537, **2005**.

110. Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. *Science 289*, 739, **2000**.

111. Honoria, K. M.; da Silva, A. B. F.; *Theochem 578*, 111, **2002**.

112. Xie, X-Q.; Yang, D-P.; Melvin, L. S.; Makriyannis, A. *J. Med. Chem. 37*, 1418, **1994**.

113. Xie, X-Q.; Melvin, L. S.; Makriyannis, A. *J. Biol. Chem. 71* 169, **1996**.

114. Song, Z. H.; Slowey, C. A.; Hurst, D. P.; Reggio, P. H. *Mol. Pharmacol. 56* 834, **1999**.

115. Makriyannis, A. *Cannabinoid Receptor*, Chapter 3, Pertwee, R., Ed.; Academic Press Limited: London, **1995**.

116. Mavromoustakos,T.; Zervou, M.; Zoumpoulakis, P.; Kyrikou, I.; Benetis, N.P.; Polevaya, L.; Roumelioti, P.; Giatas, N.; Zoga, A.; Moutevelis Minakakis, P.; Kolocouris, A.; Vlahakos, D.; Golic Grdadolnik, S.; Matsoukas, J. *Curr. Top. Med. Chem. 4*, 385, **2004**.

117. Mavromoustakos, T.; Papahatjis, D.; Laggner, P. *Biochimica et Biophysica Acta 1512* 183, **2001**.

118. Mavromoustakos, T.; Daliani, I. *Biochimica et Biophysica Acta 1420* 252, **1990**.

119. Mavromoustakos, T. Yang, D-P. Makriyannis, A. *Biochimica et Biophysica Acta 1237* 183, **1995**.

120. Yang, D-P.; Mavromoustakos, T.; Beshah, K. Makriyannis, A. *Biochimica et Biophysica Acta 1103* 25, **1992**.

121. Becke, A. D. *J. Chem. Phys. 98* 5648, **1993**.

122. Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys. 62* 2921, **1975**.

123. Soteriadou, K.; Tzinia, A. K.; Panou-Pamonis, E.; Tsikaris, V.; Sakarellos-Daitsiotis, M.; Sakarellos, C.; Papapouluo, Y.; Matsas, R. *Biochem. J. 313* 455, **1996**.

124. Stewart, J. J. P. *J. Comp. Chem. 10*, 209, **1989**.

125. Durdagi, S.; Hofer, T. S.; Randolf, B. R.; Rode, B. M. *Chem. Phys. Lett.*, *406*, 20, **2005**.

126. Loev, B.; Bender, P. E.; Dawalo, F.; Macko, E.; Fowler, P. J. *J. Med. Chem. 16*, 1200, **1973**.

127. Hanson, M. A.; Cherezov, V.; Griffith, M. T.; Roth, C. B. Stevens R. C. *Structure 16*, 897, **2008**.

128. http://accelrys-ds-visualizer.software.informer.com.

129. Toth, G.; Borics, A. *J. Mol. Graph. Model. 24*, 465, **2006**.

130. Bosi, S.; Da Ros, T.; Spalluto, G.; Balzarini, J.; Prato, M. *Bioorg. Med. Chem. Lett.*, *13*, 4437, **2003**.

131. Hou, T.; Yu, R. *J. Med. Chem. 50*, 1177, **2007**.

132. Lebon, F.; Ledecq, M. *Curr. Med. Chem. 7*, 455, **2000**.

133. Nair, A. C.; Bonin, I.; Tossi, A.; Welsh, W. J.; Miertus, S. *J. Mol. Graph. Model. 21*, 171, **2002**.

134. Lee, V. S.; Nimmanpipug, P.; Araksakunwong, O.; Promsri, S.; Sompornpisut, P.; Hannongbua, S. *J. Mol. Graph. Model.*, *26*, 558, **2007**.

135. Schuster, D. I.; Wilson, S. R.; Schinazi, R. F. *Bioorg. Med. Chem. Lett. 6*, 1253, **1996**.

136. http://chemdb.niaid.nih.gov.

137. An, Y-Z.; Ellis, G. A.; Viado, A. L.; Rubin, Y. *J. Org. Chem.*, *60*, 6353, 1995.

138. Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *Proc. Nat. Am. Sci.*, *103*, 915, **2006**.

139. Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *J. Am. Chem. Soc.*, *128*, 2812, **2006**.

140. Zhu, Z.; Schuster, D. I.; Tuckerman, M. E. *Biochemistry*, *42*, 1326, **2003**.

141. Sengupta, D.; Verma, D.; Naik, P.K. *J. Biosci.*, *32*, 1316, **2007**.

142. Conn, P.J.; Sanders-Bush, E.; Hoffman, B.J.; Hartig, P.R. *Proc. Nat. Am. Sci. 83*, 4086, **1986**.

143. Bingel, C. *Chem. Ber. 126*,1957, **1993**.

Acknowledgement

I am sincerely grateful to my supervisors of my Ph.D. thesis, Prof. Thomas Mavro-moustakos, Dr. Manthos G. Papadopoulos and Prof. Hartmut Oschkinat for their great help, kind guidance, support and enthusiastic interest and for giving me the opportunity to get into the wonderful world of rational drug design. Actually, it was my dream to perform research activity in the drug design field, thus, during my study, I really felt that my dream became a reality. Without financial support this research work could not be very easy, so I am grateful to the European Union 6th Frame work Marie Curie Actions for providing me the generous financial support during the three years of my study.

I would like to thank Dr. Heribert Reis for his great guidance in performing MD simulations and the helpful discussions in theoretical aspects.

I want to thank my colloquies in computational chemistry office Dr. Oleksandr Loboda, Dr. Aggelos Avraomopoulos, Dr. Robert Zalesny for their collaboration, moral support, motivation and enjoyment.

I would like to thank Dr. Panagiotis Zoumpulakis, Dr. Katerina Koukoulitsa, Dr. Maria Zervou, Angeliki Politi, Costas Potamitis and Eleni Siapi for their kind collaboration and providing friendship environment while I was working in the Laboratory of Molecular Analysis.

I would like to thank all of Marie-Curie Fellows, Robert Feldman, Helene Guerrand, Alia Deleanu, George Szaloki, Robert Csonka, Kyraki M. Alexakou, and Malgorzata Nepelska for sharing together common courses and friendship.

I would like to express my deepest gratitude to my parents Ali and Sakine Durdagi and to the rest of my family for their love and encouragements.

I started my Ph.D. studies as single but I ended up married Pelin Durdagi. This flourished my life and made the difficult life of being graduate student easier. Her moral

and continues support encourages me to continue the diffucult life of research through love and dedication.

Publications

Publications within the framework of this thesis

1. **Durdagi, S**.; Papadopoulos, M.G.; Zoumpoulakis, P.G.; Koukoulitsa, C.; Mavromoustakos, T. "A Computational study on CB receptors and potent CB ligand: Homology modeling, docking and molecular dynamics analysis" *Mol. Diversity* **2009** (in press).

2. **Durdagi, S**.; Supuran, C. T.; Strom, A. T.; Doostdar, N.; Kumar, M.K.; Barron, A. R.; Mavromoustakos, T.; Papadopoulos, M.G. "*In Silico* Drug Screening Approach for the Design of Magic Bullets: A Successful Example with Anti-HIV Fullerene Derivatized Amino Acids" *J. Chem. Inf. Model. 49,* 1139-1143, **2009**.

3. **Durdagi, S**.; Mavromoustakos, T.; Kronakis, N.; Papadopoulos, M. G. "Computational Design of Novel Fullerene Derivatives as Potential HIV-1 PR Inhibitors: Analysis of Binding Interactions between Fullerene Inhibitors and HIV-1 PR Residues Using 3D QSAR, Molecular Docking and Molecular Dynamics Simulations" *Bioorg. Med. Chem. 16*, 9957-9974, **2008**.

4. **Durdagi, S**.; Reis, H.; Papadopoulos, M. G.; Mavromoustakos, T. "Comparative Molecular Dynamics Simulations of the Potent Synthetic Classical Cannabinoid Ligand AMG3 in solution and at the Binding Site of the CB1 and CB2 Receptors" *Bioorg. Med. Chem. 16*, 7377-7387, **2008**.

5. **Durdagi, S**.; Mavromoustakos, T.; Papadopoulos, M. G. "3D QSAR CoMFA/CoMSIA, Molecular Docking and Molecular Dynamics Studies of Fullerene-based HIV-1 PR Inhibitors" *Bioorg. Med. Chem. Lett*., *18*, 6283-6289, **2008**.

6. **Durdagi, S**.; Papadopoulos, M. G.; Papahatjis, D. P.; Mavromoustakos, T. "Combined 3D QSAR and Molecular Docking Studies to Reveal Novel Can-

nabinoid Ligands with Optimum Binding Activity" *Bioorg. Med. Chem. Lett.* *17*, 6754-6763, **2007**.

7. **Durdagi, S.**; Kapou, A.; Kourouli, T.; Andreou, T.; Nikas, S. P.; Nahmias, V. R.; Papahatjis, D. P.; Papadopoulos, M. G.; Mavromoustakos, T. "The Application of 3D-QSAR Studies for Novel Cannabinoid Ligands Substituted at the C1' Position of the Alkyl Side Chain on the Structural Requirements for Binding to Cannabinoid Receptors CB1 and CB2" *J. Med. Chem. 50*, 2875-2885, **2007**.

8. **Durdagi, S.;** Koukoulitsa, C.; Kapou, A.; Kourouli, T.; Andreou, T.; Nikas, S.P.; Nahmias, V.R.; Papahatjis, D.P.; Papadopoulos, M.G.; Mavromoustakos, T. "Testing the 3D QSAR/ComFA-CoMSIA Results of Flexible Bioactive Compounds with Molecular Docking Studies". *Drugs of the Future*, *32 (Suppl. A)*, 79, **2007**.

9. Mavromoustakos, T.; Zervou, M.; Zoumpoulakis, P.G.; Potamitis, C.; Katsiaris, V.; Politi, A.; Mantzourani, E.; **Durdagi, S**., Koukoulitsa, C. "Putative Bioactive Conformers of Small Molecules: A Concerted Approach Using NMR Spectroscopy and Computational Chemistry". *Drugs of the Future*, *32 (Suppl. A)*, 33, **2007**.

Chapters in books

1. **Medicinal Chemistry Research Progress**. Mavromoustakos, T.; Grdadolnik, S.G.; Zervou, M.; Zoumpoulakis, P.; Potamitis, C.; Politi, A.; Manzourani, E.; Platts, J. A.; Koukoulitsa, C.; Minakakis, P.; Kokotos, G.; Tselios, T.; Matsukas, J.; **Durdagi, S**.; Papadopoulos, M. G.; Papahatjis, D. P.; Spyranti, Z.S.; Dalkas, G.A.; Spyroulias, G.A,. "*Chapter-7, Putative Bioactive Conformers of Small Molecules: A Concerted Approach Using NMR Spectroscopy and Computational Chemistry*", Colombo, G. P. and Ricci, S. (Eds.), Nova Science Publisher, **2009.**

Publications from the same period of time concerning other topics

1. Politi, A.; **Durdagi, S**.; Moutevelis-Minakakis, P., Kokotos, G. "Molecular Docking and 3D QSAR CoMFA/CoMSIA studies on renin inhibitors" *Eur. J. Med. Chem.* **2009** (in press).

2. Potamitis, C.; Zervou, M.; Katsiaras, V.; Zoumpoulakis, P.; **Durdagi, S**.; Papadopoulos, M. G.; Hayes, J.; Grdadolnik, S. G.; Kyrikou, I.; Argyropoulos, D.; Vatougia, G.; Mavromoustakos, T. "Antihypertensive Drug Valsartan: Conformational Analysis, Dynamic NMR Spectroscopy, *in silico* Docking and Molecular Dynamics studies" *J. Chem. Inf. Model. 49*, 726-729, **2009**.

3. Mavromoustakos, T.; **Durdagi, S**.; Mouchlis, V.; Politi, A.; Papadopoulos, M. G.; Kokotos, G. "Strategies in Rational Drug Design" *Drug Dis. Today* **2009** (submitted).

4. Minakakis, P.M.; Vranas, J.; Georgiades, D.; **Durdagi, S**.; Papadopoulos, M.G.; Mavromoustakos, T. "Elucidation of the product of ammoniolysis of an aromatic compound based on a combination of organic spectroscopy and molecular modeling: An ideal educational experiment to show that simple products are not easy to structurally elucidate" *J. Chem. Edu.* **2009** (submitted).

4. Kapou, A.; Benetis, N. P.; **Durdagi, S**.; Nikolaropoulos, S.; Mavromoustakos, T. "3D QSAR/CoMFA and CoMSIA Studies on Antileukemic Steroidal Esters Coupled with Conformationally Flexible Nitrogen Mustards" *J. Chem. Inf. Model. 11*, 2254-2264, **2008**.

Chapters in books from the same period of time concerning other topics

1. **5th Hellenic Forums of Bioactive Peptides**. "*Peptide Mimetics and their Interdigitation with Lipid Bilayers*", Cordopatis, P. (Ed.) (in press) **2009**.

2. **5th Hellenic Forums of Bioactive Peptides**. "*Computational Predictions of Proton-NMR Chemical Shifts of Irbesartan using MD Simulations and ONIOM Method*", Cordopatis, P. (Ed.) (in press) **2009**.

Earlier Publications

1. Salzner, U.; Karalti, O.; **Durdagi, S.** "Does the Donor–Acceptor Concept Work for Designing Synthetic Metals? Theoretical Investigation of Copolymers between Quinoid Acceptors and Aromatic Donors" *J. Mol. Model.*, *12*, 687-701, **2006**.

2. **Durdagi, S.**; Hofer, T. S.; Randolf, B. R.; Rode, B. M. "Structural and Dynamical Properties of Bi^{3+} in Water" *Chem. Phys. Lett.*, *406*, 20-23, **2005**.

Earlier chapters in books

1. **Highlights in Computational Chemistry II**. "*Does the Donor–Acceptor Concept Work for Designing Synthetic Metals? Theoretical Investigation of Copolymers between Quinoid Acceptors and Aromatic Donors*", Clark, T. (Ed.) ISBN: 978-3-540-37592-0, published by Springer, pp 687-702, **2006**.

Zusammenfassung

Der Einfluß der Konformation eines bioaktiven Moleküls auf sein pharmakologischen Profil is seit langem bekannt. Nur die biologisch aktive Konformation eines Wirkstoffmoleküls ist in der Lage, eine Bindung mit der aktiven Stelle eines Rezeptors einzugehen. In dieser Arbeit wird die Konformationsanalyse bioaktiver Verbindungen in verschiedenen Umgebungen diskutiert. Zwei verschiedene Molekülkategorien wurden untersucht: Cannabinoide (CB) sowie [60]Fullerenderivate. Die bedeutsamem strukturellen Merkmale dieser Moleküle sind zum einen, daß sie amphiphil sind und zum zweiten, daß sie sowohl flexible als auch starre pharmakophorische Segmente besitzen. Insbesondere die flexiblen Teile stellen eine Herausforderung für die Konformationsanalyse möglicher bioaktiver Konformationen dar.

Im Falle der CB-Verbindungen wurde eine Serie neuer Derivate von Δ^8-tetrahydrocannabinol (Δ^8-THC) und Cannabidiol (CBD) durch 3-dimensionale Quantitative Struktur-Aktivitätsbeziehungsstudien (3D QSAR) untersucht, einmal mittels vergleichender Molekularfeld-Analyse (CoMFA), als auch mit Methoden, die auf vergeichenden molekularen Ähnlichkeitsindizes (CoMSIA) basieren. Die hochaktive Verbindung AMG3, (C-1'-dithiolan-Δ^8-THC) wurde als Template-Molekül aus dem benutzten Datensatz ausgewählt. Die Bestimmung der potentiel bioaktiven Konformation von AMG3 in Lösung erfolgte durch verschiedene molekulare Modellierungstechniken: Monte Carlo (MC), Molküldynamik (MD) sowie Gitterscananalysen. Das erhaltene Konformer wurde dann als Template weiter benutzt, und CB1 und CB2 Pharmakophor-Modelle wurden entwickelt. Verfügbare Homologiemodelle von CB1 und CB2, die auf Rhodopsin basieren, ermöglichten die Konformationsanalyse von AMG3 an der Bindungsstelle der Rezeptoren. Die erhaltenenen energetisch begünstigten Konformere von AMG3 an der Bindungsstelle wurden mit den entsprechenden Konformationen in Lösung verglichen. Die stereoelektronischen Eigenschaften der Bindungskavitäten eines Rezeptormodells stehen in direktem Zusammenhang mit den benutzten molekularen Koordinaten des Modells. In der vorliegenden Arbeit wurde auch eine auf dem β2-Adrenorezeptor basierende Homologie-Modellstudie für die CB1 und CB2-Rezeptoren durchgeführt, und mit den Ergebnissen des mit dem auf dem Rhodopsin-Rezeptor basierten Homologiemodells verglichen. Ähnliche

Bindungsstellen in den CB1 und CB2-Rezeptoren wurden sowohl von den auf Rhodopsin basierten Modellen als auch von auf β2-Adrenorezeptor-basierten Modellen erzeugt. Die 3D QSAR Modelle wurden regeneriert mithilfe von potentiellen Konformeren von AMG3 an den Bindungsstellen der CB1- und CB2-Rezeptoren. Die von den 3D QSAR/CoMFA bzw: CoMSIA Pharmakophormodellen berechneten relativen Beiträge der sterischen und elektrostatischen Felder zeigten, daß die Bioaktivität der Verbindungen hauptsächlich durch sterische Effekte bestimmt wird, obwohl elektrostatische Effekte auch eine Rolle spielen. Ein Vergleich entsprechender QSAR Modelle zeigte, daß die erhaltenen statistischen Resultate positiv beeinflußt wurden, wenn die Komplexität der Rechnungen im Sinne einer realistischeren Modellierung des umgebenden Mediums erhöht wurde. Die optimale QSAR Analyse mit partieller minimierter quadratischer Abweichung (PLS) wurde in Arbeiten zur *de novo* Wirkstoff-Entwicklung benutzt, und haben zur Entwicklung neuer Verbindungen mit verbesserten vorhergesagten Bindungsaktivitäten geführt.

Im Fall der Fullerenderivate wurde eine Serie von experimentell bekannten als auch von theoretisch entwickelten Mono- und Bisaddukten von [60]-Fullerenderivaten ausgewählt und in Bezug auf die Bindungswechselwirkungen zwischen Fullerenbasierten Inhibitoren und Immunodefizienzvirus Typ I Endopeptidase (HIV-1 PR) mithilfe von Dockingsstudien analysiert. MD-Simulationen des freien als auch des Inhibitor-gebundenen HIV-1 PR Systems ergänzten die genannten Studien und lieferten geeignete Startstrukturen für die Dockingsimulationen von HIV-1 PR. Die erhaltenen Ergebnisse zeigen eine unterschiedliche Orienterung der sogennannten β-Haarnadel Laschen zwischen den beiden Systemem. In der Form mit angebundenem Fullereninhibitor werden die Laschen in Richtung des Bodens des aktiven Bereichs hin gezogen (sogenannte geschlossene Form), während die freie Form von HIV-1 PR eine halboffene Konformation bevorzugt. Die Strukturanalyse der katalytischen Segemente als auch der flexiblen Laschenregionen im Verlauf der Simulation von HIV-1 PR unterstützt das Verständnis sowohl der strukturellen Präferenzen dieser Regionen, als auch der von den Fullerenverbindungen eingenommenen Orientierungen innerhalb der aktiven Kavität des Enzyms. Die Fullerenverbindung aus der Datenbank, die sich als die aktivste erwies, wurde anschließend als Template ausgewählt zur Erstellung von 3D QSAR-Modellen. Die damit erhaltenen Konturoberflächen sowie die Ergebnisse der PLS-Analyse wurden für *de novo* Wirkstoffentwickungsstudien benutzt, mit dem Ziel, neue Fullerenderivate mit

höherem Bindungsaktivitäten vorzuschlagen. Solche Moleküle können für den medizinischen Chemiker zur Synthese neuer HIV-1 PR Inhibitoren mit höhere Bioaktivität von Interesse sein, und damit auf der Suche nach dringend benötigten neuen Anti-HIV Wirkstoffen von Bedeutung sein.

Appendix

List of Important Abbreviations

CB: Cannabinoid

GPCR: G protein coupled receptors

Δ^8-THC: Δ^8-tetrahydrocannabinol

CBD: Cannabidiol

3D QSAR: Three-dimensional quantitative structure-activity relationships

CoMFA: Comparative molecular field analysis

CoMSIA: Comparative molecular similarity analysis

TM: Transmembrane

MD: Molecular dynamics

MC: Monte Carlo

PLS: Partial least squares

PCR: Principal components regression

2D-NOESY: Nuclear Overhauser effect spectroscopy

2D-ROESY: Rotating-frame Overhauser effect spectroscopy

2D-COSY: Correlated spectroscopy

r^2: the squared correlation coefficient

r_{cv}^2: Cross-validated r^2

TSS: Total sum of squares

ESS: Explained sum of squares

RSS: Residual sum of squares

MLR: Multiple linear regression

(i) The complete lists of residues consisting the binding site of the CB1 and CB2 receptors:

CB1: Ile119, Leu122, Ser123, Leu126, Ala162, Asp163, Leu165, Gly166, Ser167, Ile167, Phe170, Val171, Tyr172, Ser173, Phe174, Ile175, Asp176, Phe177, His178, Val179, Ser185, Arg186, Asn187, Val188, Phe189, Leu190, Phe191, Lys192, Leu193, Gly194, Gly195, Val196, Thr197, Ala198, Ser199, Phe200, Thr201, Ala202, Ser203, Val204, Gly205, Ser206, Ile243, Ala244, Val246, Ile247, Ala248, Val249, Leu250, Pro251, Leu252, Tyr275, Thr283, Leu287, Val351, Leu352, Ile353, Ile354, Cys355, Trp356, Gly357, Pro358, Leu359, Leu360, Ala361, Ile362, Met363, Val364, Val367, Val378, Phe379, Ala380, Phe381, Cys382, Ser383, Met384, Leu385, Cys386, Leu387, Leu388, Asn389, Ser390, Thr391 and Val392.

CB2: Phe106, Leu107, Leu108, Ile110, Gly111, Ser112, Val113, Thr114, Met115, Thr116, Phe117, Thr118, Ala119, Gly122, Leu160, Ser161, Ala162, Leu163, Val164, Ser165, Tyr166, Leu167, Pro168, Leu169, Asp189, Tyr190, Leu191, Leu192, Ser193, Trp194, Leu196, Phe197, Ile198, Leu201, Leu254, Cys257, Trp258, Phe259, Pro260, Val261, Leu262, Ala263, Leu264, Met265, Ala266, His267, Ser268, Leu269.

(ii) Complete lists of residues used at the binding site of HIV-1 PR receptor:

Chain A: Gln7, Arg8, Pro9, Leu23, Leu24, Asp25, Thr26, Gly27, Ala28, Asp29, Asp30, Thr31, Val32, Leu33, Met46, Ile47, Gly48, Gly49, Ile50, Gly51, Phe53, Ile54, Lys55, Val56, Pro79, Thr80, Pro81, Val82, Asn83, Ile84, Ile85.

Chain B: Arg8, Ala22, Asp25, Thr26, Gly27, Ala28, Asp29, Asp30, Val32, Met46, Ile47, Gly48, Gly49, Ile50, Gly51, Gly52, Phe53, Ile54, Thr80, Pro81, Val82, Ile84, Ile85.
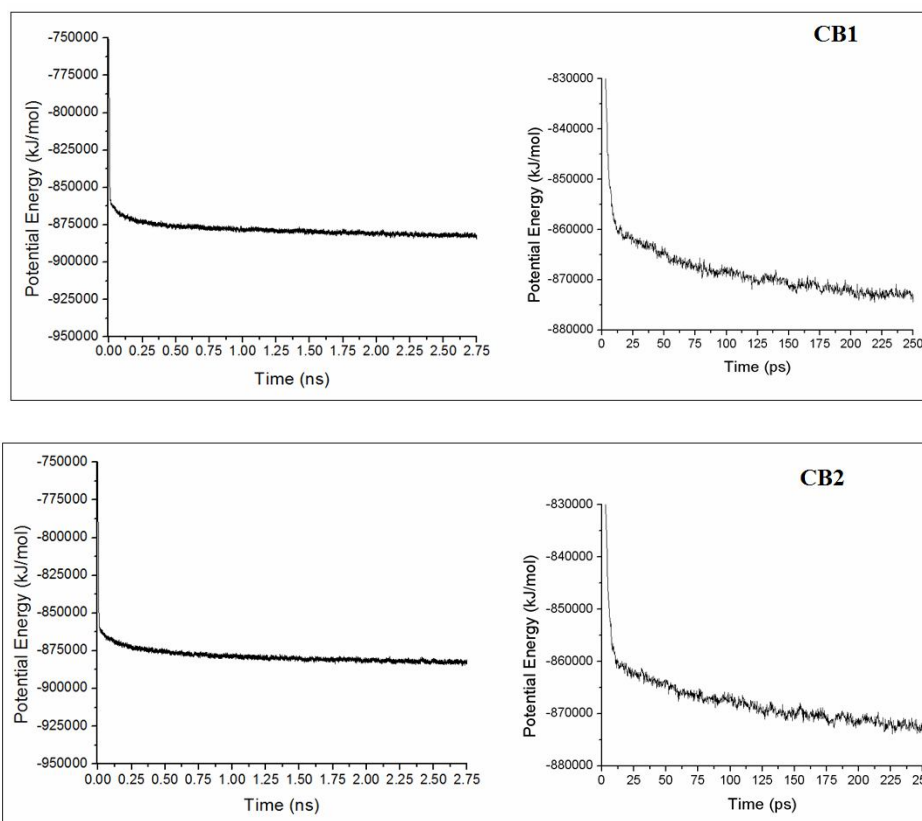
Figure A1. (top) Potential energy versus time plots throughout the MD simulations (include equilibration and simulation parts) for the membrane bound CB1 receptor systems (corresponding plot for the equi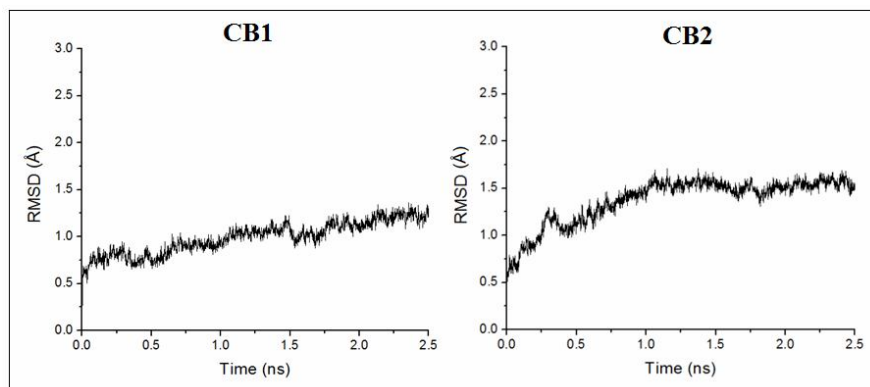libration has been shown on the right for clarity). **(bottom)** Potential energy versus time plot throughout the MD simulations (include equilibration and simulation parts) for membrane bound CB2 receptor systems (corresponding plot for the equilibration has been shown on the right for clarity).

Figure A2. Receptor backbone RMSD versus time plots throughout the MD simulations for membrane bound CB1 (left) and CB2 receptor (right) systems.
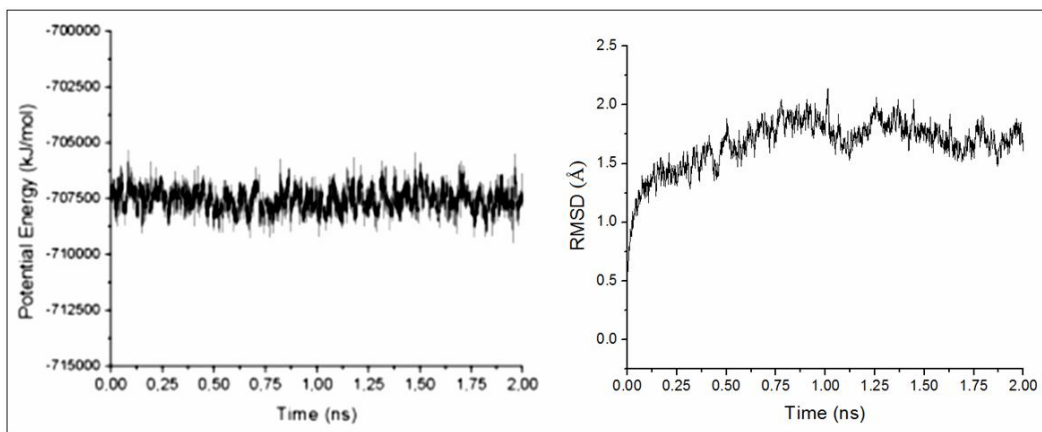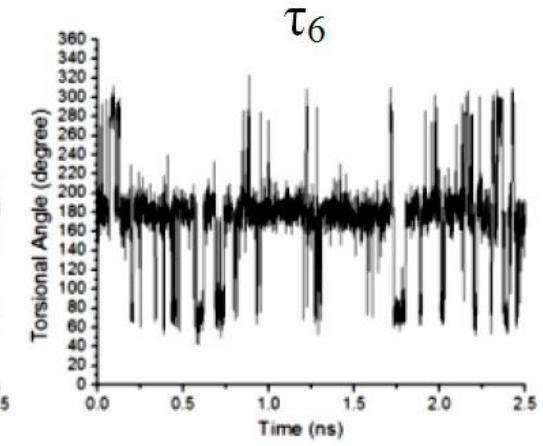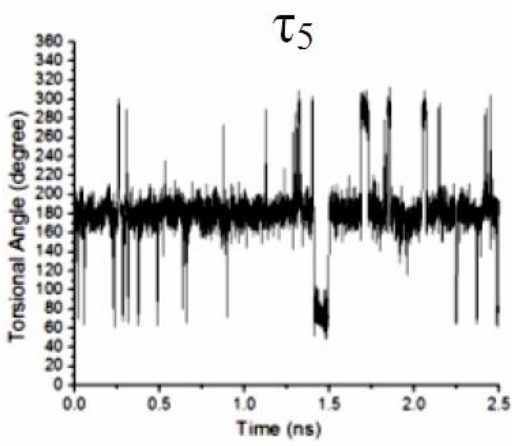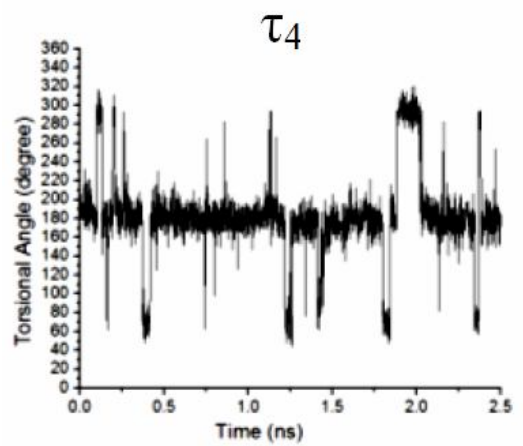
Figure A3. Potential energy versus time (left) and receptor backbone RMSD versus time (right) plots throughout the MD simulations of fullerene/HIV-1 PR systems.
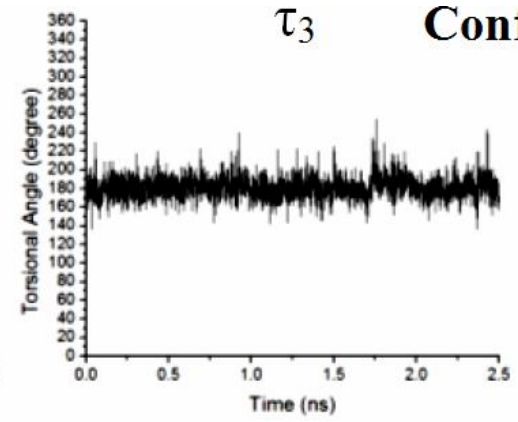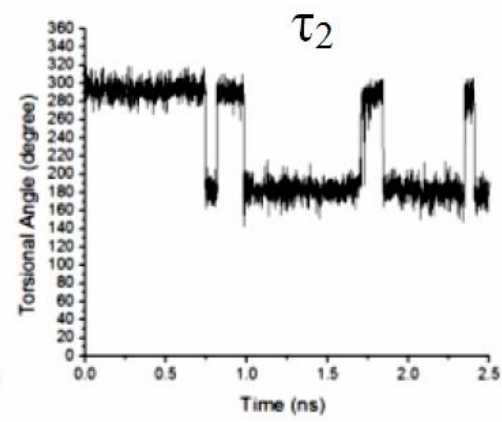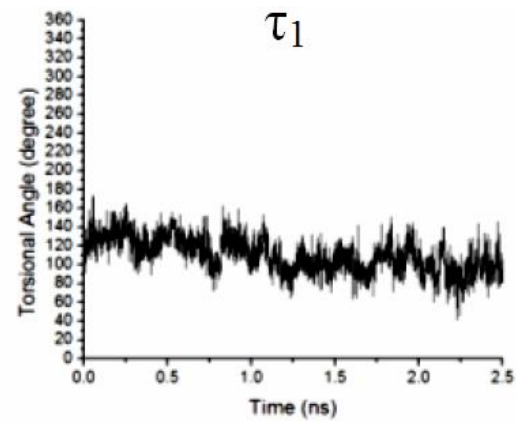
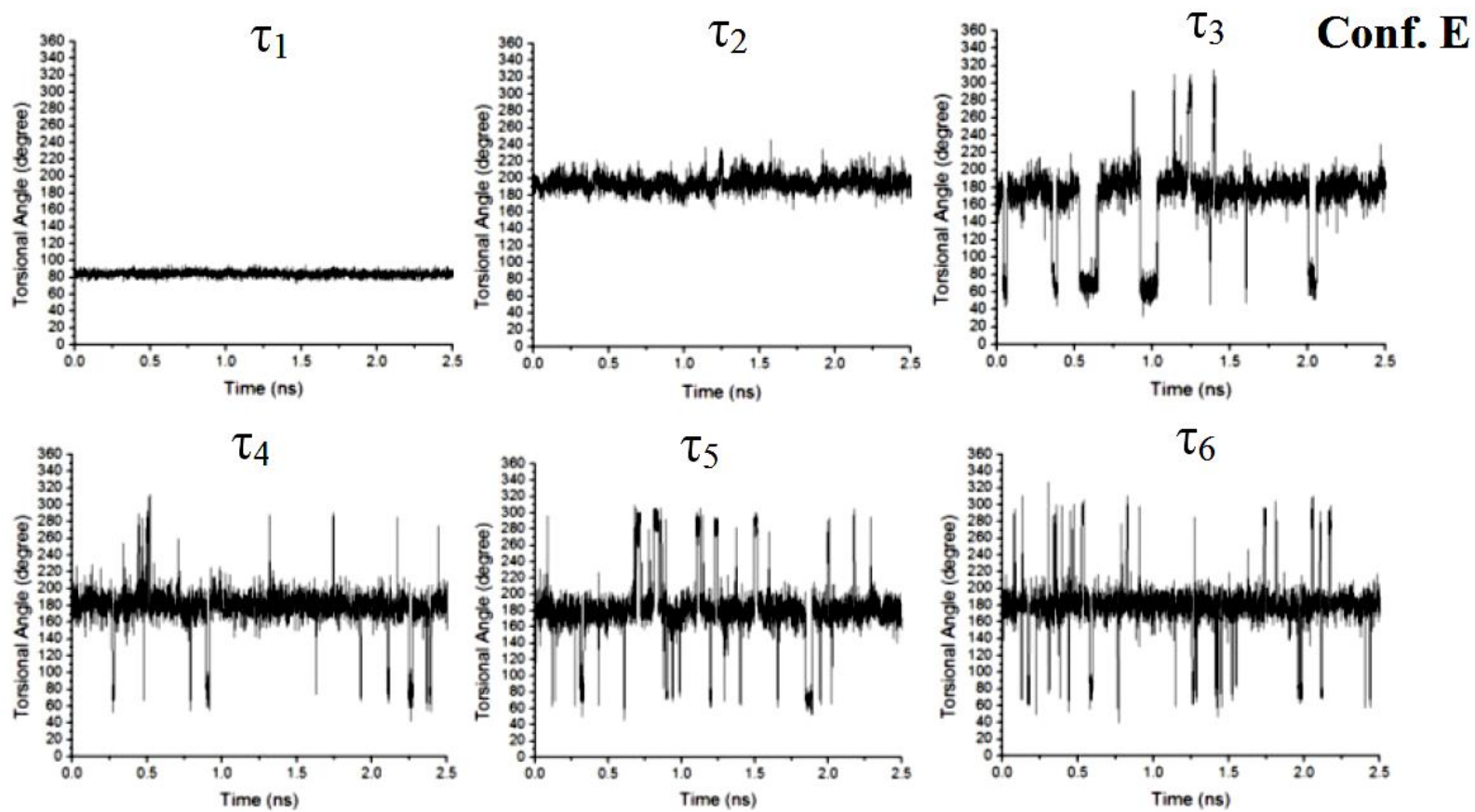**Figure A4.** Torsional angle screening throughout in lipid bilayer simulations for wrapped conformers **D** and **E**.